philosophia naturalis

Journal for the Philosophy of Nature

KLOSTERMANN Band/Volume 46/2009 Heft/Issue I

philosophia naturalis

Journal for the Philosophy of Nature

46 / 2009 / 1

KLOSTERMANN

Inhalt

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*

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

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

ISSN 0031-8027

Brigitte Falkenburg, Andreas Hüttemann,
Manfred Stöckler

Physik, Philosophie und die Einheit der Wissenschaften

Nachruf auf Erhard Scheibe

Erhard Scheibe ist am 7. Januar 2010 nach langer Krankheit im Alter von 82 Jahren gestorben. Er ist einer der bedeutendsten Wissenschaftsphilosophen in Deutschland und gehörte von 1989 bis 2002 dem Herausgebergremium dieser Zeitschrift an. Im Zentrum seines Lebenswerks standen die philosophischen Probleme der Relativitäts- und Quantentheorie und die Suche nach der Einheit unseres Wissens von der Natur. Erhard Scheibe wurde am 24. September 1927 in Berlin geboren – kurz nachdem Niels Bohr in Como den Vortrag gehalten hatte, an dem sich die Bohr-Einstein-Debatte über die Quantenmechanik entzündet hat. Nach dem Zweiten Weltkrieg studierte Erhard Scheibe in Göttingen Mathematik, Physik und Philosophie, wo er zum Kreis um Carl Friedrich von Weizsäcker gehörte und 1955 in Mathematik promovierte. Mit von Weizsäcker ging er als Assistent nach Hamburg und habilitierte sich dort 1963 mit einer philosophischen Studie zur Quantenmechanik. Die Schrift erschien 1964 unter dem Titel *Die kontingenten Aussagen in der Physik*.

Im selben Jahr wurde er als Professor für Philosophie nach Göttingen berufen. Dort befasste er sich vorwiegend mit der formalen Struktur physikalischer Theorien und veröffentlichte 1973 *The Logical Analysis of Quantum Mechanics*. Das Buch beginnt mit einer glasklaren Darstellung der „Komplementaritätsphilosophie" Bohrs und endet mit einer formalen Analyse des Gedankenexperiments von Einstein, Podolski und Rosen, dem Höhepunkt der Bohr-Einstein-Debatte. Seine Aufsätze und Vorträge umspannen Themen von Platon bis Leibniz und Kant, von Grundlagenfragen der Mathematik bis zur Rolle der Mathematik in der Physik, von der Struktur der Raumzeit bis zu den konzeptuellen Brüchen, die mit der Quantenrevolution in die Physik kamen. Erhard Scheibe wurde nun als Philosoph der exakten Wissenschaften bekannt,

der die Wissenschaftsphilosophie im deutschsprachigen Raum prägte. Er wurde in deutsche und internationale Akademien der Wissenschaft aufgenommen. 1983 folgte er dem Ruf nach Heidelberg auf den damals neu eingerichteten Lehrstuhl für *Philosophie unter besonderer Berücksichtigung der Philosophie der formalen Wissenschaften und der Naturwissenschaften*, den er bis zu seiner Emeritierung im Jahr 1992 innehatte.

In den Heidelberger Jahren wandte sich Erhard Scheibe verstärkt dem geschichtlichen Hintergrund der Physik zu. In Forschung und Lehre verfolgte er nun den Weg, den sich die Physik des 20.Jahrhunderts zwischen Empirismus und Rationalismus bahnte. Aus einem Kolloquium zu Ehren von Erhard Scheibe ist 1995 ein Buch mit dem sein Werk gut charakterisierenden Titel *Physik, Philosophie und die Einheit der Wissenschaften* hervorgegangen, das auch ein Verzeichnis seiner bis zu diesem Zeitpunkt erschienenen Schriften enthält.

Bezeichnend für sein Engagement als akademischer Lehrer, doch auch für sich verschlechternde universitäre Forschungsbedingungen war, dass er sein zweibändiges Hauptwerk *Die Reduktion physikalischer Theorien*, eine detaillierte Untersuchung zur Einheit der Physik (1997, 1999), erst nach der Emeritierung abschließen konnte. Erhard Scheibe geht dabei im Detail der Frage nach, was es heißt, dass eine Theorie der Physik auf eine andere reduziert wird, genauer „was jemand, der sich als Reduktionist fühlt, *in concreto* machen müsse, wenn er seinem Gefühl allgemeine Geltung verschaffen wollte und wenn er bereit ist, sich dabei gewissen heute in der Wissenschaftstheorie üblichen, aber nicht zu unbescheidenen Standards zu unterwerfen“ (Bd. 1, S. 8). Die neue Antwort auf diese Frage geht nicht von einem einheitlichen Schema der Theorienerklärung aus, sondern rekonstruiert aus der Praxis der Physik ein Spektrum unterschiedlicher elementarer Reduktionsarten, die in der Regel in Kombination zum Einsatz kommen. Nach einer mengentheoretisch orientierten Explikation des Begriffs einer physikalischen Theorie werden im Wechselspiel von Fallstudien und formaler Explikation verschiedene Spielarten der exakten Reduktion (Äquivalenz, Einbettung, Verfeinerung, Erweiterung, Vereinigung) sowie die lange vernachlässigten approximativen und partiellen Reduktionen analysiert. Der zweite Band untersucht mit diesem Werkzeug Inkommensurabilitäten und Grenzfallreduktionen an komplexen Reduktionsfällen aus der Thermodynamik, der Relativitätstheorie und der Quantenmechanik. Auch wer sich nicht im Detail auf Hilbert/Schmidt-Operatoren und affine Zusammenhänge einlassen will,

wird von der Problemexposition fasziniert sein, in der Scheibe den Fortschrittsgedanken im Selbstverständnis der Physiker den entsprechenden Theorien der Wissenschaftsphilosophie gegenüber stellt.

Im Jahre 2001 erschien der Band *„Between Rationalism and Empiricism: Selected Papers in the Philosophy of Physics*, der die wichtigsten Aufsätze aus den Göttinger und Heidelberger Jahren versammelt. Sein letztes, zunehmendem Kräfteverfall abgerungenes Buch *Die Philosophie der Physiker* erschien 2006. Es stellt die philosophischen Überzeugungen der bedeutendsten Physiker des 20. Jahrhunderts für einen größeren Leserkreis dar, und es hatte einen so großen Erfolg, dass schon 2007 eine zweite Auflage notwendig wurde.

Diese Daten zeigen aber nur einen Teil des Einflusses, mit dem Scheibe die deutsche Wissenschaftsphilosophie geprägt hat. Wer ihn als Doktorand oder Mitarbeiter erlebt hat, wurde noch durch ganz andere Dinge beeindruckt. Dank Erhard Scheibe fanden zahlreiche Studierende der Mathematik und der Naturwissenschaften ihren Weg in das philosophische Seminar, um an wissenschaftsphilosophischen Veranstaltungen oder solchen zu den Schriften philosophischer Klassiker wie Locke oder Kant teilzunehmen. In seinen Lehrveranstaltungen faszinierte er durch die seltene Gabe, immenses Wissen und große Ernsthaftigkeit bei der Diskussion philosophischer Fragen mit einem feinsinnigen Humor zu verbinden.

Wie nur ganz wenigen ist es Erhard Scheibe gelungen, Logik, Physik und Philosophie in jeder Disziplin auf höchstem Niveau zu verknüpfen. Philosophie der Natur war für ihn ohne diese Verbindung nicht denkbar. Seine Integrität, seine nur an der Sache orientierte und von jeder Eitelkeit freie Persönlichkeit und sein unbestechlicher Blick für argumentative Schwächen und inhaltliche Unschärfen werden uns sehr fehlen. Erhard Scheibes Schriften aber bleiben, seine Gedankentiefe und abwägende Argumentation, seine Unabhängigkeit von Moden und seine Abneigung gegen jede Oberflächlichkeit machen sie immer wieder zu einer lohnenden Lektüre.

Sven Walter

Wie frei sind wir eigentlich – empirisch?*

Zusammenfassung

Es gehört zu den elementaren Grunderfahrungen des menschlichen Daseins, dass wir uns in unserem Entscheiden und Handeln zumindest zeitweise als *frei* erleben. Diese unsere Selbstwahrnehmung wird von den Naturwissenschaften zunehmend zur Selbst*täuschung* degradiert. Freiheit, so wird dort immer wieder betont, ist eine Illusion, denn unsere Selbstwahrnehmung als allein aus der rationalen Abwägung von Gründen heraus entscheidende und handelnde Autoren unseres eigenen Tuns ist mit naturwissenschaftlichen Überlegungen prinzipiell nicht zu vereinbaren. Demgegenüber betont die Philosophie unentwegt, die Naturwissenschaften arbeiteten mit zu starken Freiheitskonzeptionen, und der naturwissenschaftliche Angriff auf die Freiheit liefe leer, sobald Freiheit in einem kompatibilistischen Sinn verstanden werde. Beide Seiten haben Unrecht. Einerseits reichen die üblicherweise diskutierten empirischen Befunde zum endgültigen Nachweis unserer Freiheit keineswegs aus, andererseits gibt es in der Tat empirische Befunde, die auch eine kompatibilistisch verstandene Freiheit einschränken, wenngleich nicht vollständig widerlegen.

Abstract

We arguably all experience ourselves as someone who is able, at least sometimes, to decide and act freely. Natural sciences have a tendency to denigrate this self-experience to a self-deception. Freedom, they maintain, is an illusion, because our self experience as freely and rationally deliberating authors of our own deeds is in principle incompatible with a scientific approach to the world. In contrast, philosophers tend to stress that the natural sciences are relying on too strong a notion of freedom and that the scientific attack on free will can be avoided by adopting a compatibilist notion of free will. Both parties are wrong, or so I argue. On the one hand, the empirical results that are typically discussed in the „free will is an illusion" literature are inconclusive. On the other hand, however, there are indeed some empirical results that limit our freedom (although they do not render us unfree), even if that freedom is understood in a compatibilist fashion.

1. Freiheit: Philosophisches oder empirisches Problem?

Es gehört zu den elementaren Grunderfahrungen des menschlichen Daseins, dass wir uns in unserem Entscheiden und Handeln zumindest zeitweise als *frei* erleben. Eng mit unserer Freiheit verknüpft ist unsere *Verantwortlichkeit*. Es wäre schlicht unfair, uns für Entscheidungen und Handlungen verantwortlich zu machen, die wir überhaupt nicht unterlassen konnten – „*Ultra posse nemo obligatur*". An unserer Freiheit hängt zudem auch unsere *Schuldfähigkeit*. §20 StGB regelt, dass schuldhaft nur der handelt, dem eine Entscheidung gegen die Tat und damit ihre Unterlassung möglich war, und der Bundesgerichtshof hat entschieden, dass der „innere Grund des Schuldvorwurfs [darin] liegt …, daß der Mensch auf freie, verantwortliche, sittliche Selbstbestimmung angelegt und deshalb befähigt ist, sich für das Recht und gegen das Unrecht zu entscheiden" (BGHSt 2, 200). Mit der Schuldfähigkeit steht und fällt schließlich auch unser *retributives Strafrecht*: Gemäß dem Grundsatz „*Nulla poena sine culpa*", der in Deutschland den Rang eines Verfassungsrechtsatzes hat (BVerfGE 20, 323), kann es ohne Schuld keine Strafe geben. An unserer Freiheit hängt also so einiges.[1]

Bekanntermaßen jedoch wurde unsere Selbstwahrnehmung als frei Entscheidende und Handelnde des Öfteren zur Selbst*täuschung* degradiert. Unter anderem wollte nicht jedem einleuchten, wie sich Freiheit mit dem Vorauswissen eines allwissenden Gottes, der Prädestinationslehre oder dem Determinismus vereinbaren lässt. Freiheit wurde damit zum Problem, und zwar zunächst zu einem philosophischen Problem, das außerakademisch nicht weiter von Interesse war.

In den letzten Jahren hat sich die Lage in zweierlei Hinsicht verschärft. Zum einen ist Freiheit nicht länger nur ein philosophisches Problem. Ob Entscheidungen oder Handlungen frei sind, hängt ganz offenbar mit davon ab, wie sie zustande kamen. Daher sieht sich die empirische Wissenschaft in dem Maß, in dem sie die Antezedenzien unseres Entscheidens und Handelns immer detaillierter beschreiben kann, ebenfalls berufen, zur Freiheit Stellung zu nehmen. Zum anderen wird das Problem der Freiheit im Zuge seiner Verwissenschaftlichung dank populärwissenschaftlicher TV-Sendungen und Magazine ins Interesse einer breiten Öffentlichkeit gerückt.

Das Bild, das dabei nach draußen kolportiert wird, ist *freiheitsskeptisch*. Unserer *erlebten* Freiheit, so wird behauptet, entspricht keine

tatsächliche Freiheit. *Freiheit ist eine Illusion*. „Die Idee eines freien menschlichen Willens ist mit wissenschaftlichen Überlegungen prinzipiell nicht zu vereinbaren", so der Psychologe Wolfgang Prinz (2004, 22), denn, so die ähnlich betrübliche Einschätzung des Hirnforschers Wolf Singer (2003, 59), „aus Sicht der Naturwissenschaft ergibt sich die mit der Selbstwahrnehmung unvereinbare Schlussfolgerung, dass der ‚Wille' nicht frei sein kann".

Das Problem liegt weniger im Angriff auf die Freiheit selbst als vielmehr darin, dass unsere Unfreiheit zur *wissenschaftlichen Tatsache* erhoben wird. Unser Freiheitserleben ist nicht deshalb illusorisch, weil sich Philosophen irgendwelche spitzfindigen Argumente ausgedacht haben, sondern weil es wissenschaftlich und für jeden nachvollziehbar so festgestellt wurde: „Die Hirnforscher sind wahrhaftig nicht die ersten, die uns die Idee der Freiheit austreiben wollen. Aber sie haben als erste das Vorzeichen der Spekulation gegen das Vorzeichen der Exaktheit ausgetauscht" (Geyer, 2004, 12). Diese vermeintliche Wissenschaftlichkeit, letztlich die einzige Illusion in der ganzen Debatte, lässt neuerdings Rechtswissenschaftler befürchten, dass „die gesamte Rechtsordnung auf dem Prüfstand steht" (Lampe *et al*., 2008, 16) und führt dazu, dass plötzlich eine „Entmoralisierung des Rechts" (Grün *et al*., 2008) propagiert wird. Was genau die Bedingungen unserer Freiheit sind und ob (und wenn ja, wie) sie sich empirisch widerlegen oder bestätigen lassen, steht scheinbar überhaupt nicht mehr zur Debatte. Die Frage scheint bloß noch zu sein, wie die neurowissenschaftliche Revolution von Moral und Recht im Detail auszusehen hat. Ich halte diese Entwicklung für höchst bedenklich, weil interessierten Laien und Fachkollegen, nicht selten in spektakulär überschriebenen Feuilletonartikeln, vorgegaukelt wird, unsere Unfreiheit sei wissenschaftlich unwiderruflich bewiesen.

Weil das Problem weniger die wissenschaftlichen Befunde selbst als ihre vorschnelle philosophische Interpretation ist, ist es Aufgabe der Philosophie, deutlich zu sagen, was genau aus den gesicherten empirischen Erkenntnissen für welche Freiheitskonzeption folgt. Das heißt nicht, dass die Philosophie die alleinige Deutungshoheit beanspruchen und sich empirische Befunde so zurechtbiegen kann, wie es ihr gerade passt. Benötigt wird eine in beide Richtungen unvoreingenommene Analyse, die einerseits der gesamten Bandbreite philosophischer Theoriebildung Rechnung trägt und nicht einfach voraussetzt, frei seien nur Entscheidungen und Handlungen mit einem „völlig immateriell[en]" (Roth, 2001,

436) Ursprung, die aber andererseits die Möglichkeit einer empirischen Einschränkung der Freiheit auch nicht grundsätzlich ausschließt.

Die Philosophie liegt in meinen Augen falsch, wenn sie den Eindruck vermittelt, empirische Befunde unterminierten nur libertarische und/oder dualistische Freiheitskonzeptionen, wonach wir um der Freiheit Willen „in einem ansonsten deterministisch verfassten Bild von der Welt lokale Löcher des Indeterminismus … akzeptieren" müssen (Prinz, 1996, 92), man hätte aber empirisch nichts zu befürchten, wenn man seine philosophischen Hausaufgaben gemacht und einen solchen Freiheitsbegriff durch einen vernünftigen – sprich: kompatibilistischen – ersetzt hat. Ebenso falsch ist der von der Gegenseite vermittelte Eindruck, unsere Unfreiheit sei wissenschaftlich eine ausgemachte Sache. Die Philosophie hat Recht mit ihrer Einschätzung, dass die üblichen Befunde als Beleg unserer Unfreiheit unzureichend sind. Entsprechend Unrecht hat hier die empirische Wissenschaft. Die Philosophie hat jedoch Unrecht, wenn sie suggeriert, kompatibilistische Freiheit könne empirisch *prinzipiell* nicht unterminiert werden. Entsprechend richtig liegen hier die Empiriker, wenn sie ihr grundsätzliches Recht einfordern, auch etwas Substanzielles zur Freiheitsdebatte beitragen zu können. Es gibt in meinen Augen tatsächlich empirische Befunde, die eine Einschränkung unserer Freiheit nahe legen – nur sind es eben nicht die in der von der Öffentlichkeit wahrgenommenen Debatte üblicherweise diskutierten, und sie rechtfertigen auch nicht die Behauptung, unser Freiheitserleben sei immer und vollständig illusionär. Für diesen zweiten Teil meiner Analyse der Freiheitsdebatte habe ich in Walter (in Begutachtung) ausführlich argumentiert. In der vorliegenden Arbeit möchte ich primär (aber nicht nur; vgl. Abschnitt 6 und 7) zeigen, warum die üblicherweise diskutierten empirischen Befunde der Freiheit unserer Entscheidungen und Handlungen nichts anhaben können.

2. Determinismus

Ein Thema, das mit einer *empirischen* Widerlegung der Freiheit auf den ersten Blick wenig zu tun hat, von empirischen Freiheitsskeptikern aber dennoch immer wieder ins Feld geführt wird, ist der *Determinismus*. In einer deterministischen Welt gibt es keine ontologisch offenen zukünftigen Weltverläufe, damit keine echten Alternativen und damit scheinbar

keine Freiheit.[2] Unser Entscheiden und Handeln ist in einer solchen Welt die unausweichliche Konsequenz von Geschehnissen, die wir nicht kontrollieren können, weil sie vor unserer Geburt stattfanden, und ist damit scheinbar unfrei.

Grundsätzlich lässt sich gegen diesen inkompatibilistischen Angriff auf die Freiheit einwenden, dass kompatibilistische Ansätze ja gerade die Vereinbarkeit von Freiheit und Determinismus behaupten: Der Determinismus kann der Freiheit schon deshalb nicht im Wege stehen, weil eine freie Handlung nicht eine solche ist, die überhaupt nicht bedingt ist, sondern eine solche, die „auf ganz bestimmte Weise bedingt ist: durch unser Denken und Urteilen" (Bieri, 2001, 80).[3]

Im Folgenden geht es mir jedoch um etwas anderes, nämlich darum, dass der Determinismus für *empirische* Freiheitsskeptiker gänzlich irrelevant ist. Ein allgemeiner Laplace'scher Determinismus ist eine Behauptung über den Charakter fundamentaler Naturgesetze – er besagt, dass die Naturgesetze zusammen mit den Anfangsbedingungen den Weltverlauf eindeutig beschreiben.[4] Damit hätte, wenn überhaupt, die Physik und nicht die Neurowissenschaft oder die Psychologie den Nachweis für den Determinismus zu erbringen. Die Physik kann aber auch wenig tun. Der allgemeine Determinismus behauptet, dass jede mögliche Welt, die zu irgendeinem Zeitpunkt exakt mit der aktualen Welt übereinstimmt, dies zu jedem Zeitpunkt tut, und das ist keine empirisch überprüfbare Hypothese, sondern ein metaphysisches Postulat.

Laut Prinz ist Freiheit mit wissenschaftlichen Überlegungen prinzipiell nicht zu vereinbaren:

> Wissenschaft geht davon aus, daß alles, was geschieht, seine Ursachen hat und daß man diese Ursachen finden kann. Für mich ist unverständlich, daß jemand, der empirische Wissenschaft betreibt, glauben kann, daß freies, also nichtdeterminiertes Handeln denkbar ist. (Prinz, 2004, 22)

Abgesehen davon, dass Prinz die Möglichkeit eines determinierten und zugleich freien Handelns unterschlägt, setzt er wie viele andere den Determinismus mit der Behauptung gleich, alles, was geschehe, habe eine Ursache. Das ist falsch. In einer deterministischen Welt kann es unverursachte Ereignisse geben (z.B. die Anfangsbedingungen) und in einer indeterministischen Welt kann alles, was geschieht, eine Ursache haben (dann nämlich, wenn es probabilistische Ursachen gibt). Davon abgesehen ist überhaupt nichts daran auszusetzen, dass es für Naturwis-

senschaftler wie Prinz auf der Hand liegt, dass alles, was geschieht, eine Ursache hat, und dass sie den Indeterminismus für unverständlich halten. Man sollte sich aber der Tatsache bewusst bleiben, dass es sich hierbei um eine *Annahme* handelt, und nicht um etwas, was sich wissenschaftlich nachweisen ließe.

Anstelle eines allgemeinen Determinismus wird oftmals ein so genannter „bereichsspezifischer" Determinismus ins Spiel gebracht, wonach Entscheidungen und Handlungen durch neuronale, psychologische oder genetische Faktoren, wie z.B. unsere biologische Ausstattung, Erziehung, soziale Einbettung und Krankheiten determiniert werden. Singer (2004, 52) etwa hat einen *neuronalen Determinismus* im Sinn, wenn er erklärt: „der [zur Entscheidung führende; S.W.] Abwägungsprozess selbst beruht natürlich … auf neuronalen Prozessen und folgt somit … deterministischen Naturgesetzen".[5] Dasselbe gilt für Gerhard Roths (2001, 447) Behauptung, es könne „keinen vernünftigen Zweifel daran geben, daß es auch bei den hochstufigen Prozessen in unserem Gehirn, die für die Steuerung unseres Verhaltens zuständig sind, deterministisch zugeht". Andere Autoren vertreten einen *psychologischen Determinismus*. John Bargh und Tanya Chartrand (1999, 462) z.B. argumentieren dafür, dass „most of a person's everyday life is determined not by their conscious intentions and deliberate choices but by mental processes that … operate outside of conscious awareness". Ganz ähnlich äußern sich Bargh und Melissa Ferguson (2000, 925): „For every psychological effect … there exists a set of causes, or antecedent conditions, that uniquely lead to that effect".

Singer und Roth führen keine Belege für ihren neuronalen Determinismus an. Aus gutem Grund – es gibt nämlich keine. Ein bereichspezifischer Determinismus erfordert, dass in der entsprechenden Disziplin strikte, ausnahmslose Gesetze formuliert werden, und solche Gesetze kennen weder die Neurowissenschaft noch die Psychologie oder die Genetik. Bargh und Kollegen führen zur Stützung ihres psychologischen Determinismus eine Vielzahl von sozialpsychologischen Studien an, die zeigen, dass vermeintlich selbst initiierte Handlungen durch unbewusste Faktoren beeinflusst werden. In einer Studie von Bargh *et al.* (1996) z.B. mussten Versuchspersonen verbale Aufgaben lösen, mit denen angeblich ihre sprachlichen Fähigkeiten getestet wurden. In einer Gruppe suggerierten die Aufgaben das Charaktermerkmal der *Unhöflichkeit*, in einer anderen das der *Höflichkeit* und eine dritte Kontrollgruppe erhielt semantisch

neutrale Aufgaben. Anschließend wurden die Versuchspersonen in eine Situation verwickelt, in der sie um ihrer Aufgabe nachzukommen ein vermeintlich persönliches Gespräch zwischen dem Versuchsleiter und einem Mitarbeiter hätten unterbrechen müssen. 67 % der „unhöflichen" Gruppe unterbrachen die Unterhaltung, verglichen mit 38 % in der neutralen Kontrollgruppe und 16 % in der „höflichen" Gruppe. Studien wie diese zeigen, so Bargh und Ferguson (2000, 925), dass wir psychologisch determiniert sind: „scientists have accumulated evidence of determinism by their many demonstrations of mental and behavioral processes that can proceed without the intervention of conscious deliberation and choice". Ganz sicher sollten derartige Befunde unsere Vorstellungen von Freiheit maßgeblich beeinflussen (Walter 2010; vgl. auch Abschnitt 7), aber sie als Beleg für den Determinismus anzuführen ist Unsinn. Die „unhöflichen" Probanden waren ja gerade *nicht* determiniert, die Unterhaltung zu unterbrechen, denn 33 % davon taten es nicht.

Das grundsätzliche Problem ist: Empirische Daten rechtfertigen immer nur statistische Korrelationen, keine deterministischen Zusammenhänge, und deshalb sind empirische Studien als Beleg für die Behauptung, jedes psychologische Phänomen hätte „a set of causes, or antecedent conditions, that uniquely lead to that effect", untauglich. Ein Determinismus gleich welcher *couleur* ist also empirisch nicht zu bestätigen.[6]

3. Libet

Benjamin Libet (1985; Libet *et al.*, 1983) instruierte Probanden, sich innerhalb eines vorgegebenen Zeitraums frei zu entscheiden, eine einfache Bewegung auszuführen: „[S]ubjects performed a simple flick or flexion of the wrist at any time they felt the urge or wish to do so" (Libet, 2002, 552). Die Probanden sollten sich den Zeitpunkt merken, zu dem ihnen ihre Entscheidung, die Bewegung *jetzt* auszuführen, bewusst wurde. Dieser wurde mit dem mittels EMG erfassten tatsächlichen Beginn der Bewegung und mit dem mittels EEG erfassten Einsetzen des so genannten „Bereitschaftspotenzials" (BP) verglichen, das im Gehirn für die Bewegungsvorbereitung zuständig ist. Im Schnitt wurden sich die Probanden 200 ms vor dem Beginn der Bewegung der entsprechenden Entscheidung bewusst. Allerdings begann der Aufbau des BP im Schnitt schon 550 ms vor der Bewegung, also 350 ms *vor* dem Bewusstwerden

der Entscheidung. Libet (2002, 555) schloss daraus: „The initiation of the freely voluntary act appears to begin in the brain unconsciously, well before the person consciously knows he wants to act!", und lange Zeit war Libets Experiment *der* Grund für empirisch motivierte Freiheitsskepsis. Ich möchte an dieser Stelle nur auf eine Reihe von kritischen Punkten und möglichen Missverständnissen hinweisen, ohne damit eine umfassende Diskussion zu beanspruchen.[7]

(1.) Mit dem Determinismus hat das BP nichts zu tun. Nach dem Bewusstwerden der Entscheidung bleiben noch zwischen 100 und 150ms, in denen die Handlung unterdrückt werden kann (maximal also 500ms *nach* Einsetzen des BP) – Libets berühmtes *Veto*. Es ist daher auf groteske Art richtig, wenn Prinz (2004, 22) bemerkt: „um festzustellen, daß wir determiniert sind, bräuchten wir die Libet-Experimente nicht". Natürlich nicht. Dafür sind sie ganz einfach ungeeignet.[8]

(2.) Libets Rede vom „In-Gang-Setzen" („initiation") einer Handlung ist irreführend. Er kann nicht meinen, dass das BP eine neue Kausalkette anstößt, an deren Ende die Handlung steht, denn Kausalketten beginnen nicht einfach so irgendwo (was im Übrigen aus der laut Prinz unumgänglichen Voraussetzung folgt, dass alles, was geschieht, eine Ursache hat). Kausalketten fangen ebenso wenig vor einer Handlung an, wie sie danach aufhören, sondern sie bestanden schon immer und werden immer bestehen. Kausalketten mögen sich kreuzen, und wir mögen sie beeinflussen oder auch nicht, aber sie anstoßen, und damit eine Handlung initiieren, können weder wir noch bewusste Entscheidungen oder das unbewusste BP. Bewusste Entscheidungen sind also in der Tat nicht der *Ursprung* von Handlungen. Nicht aber etwa, weil ihnen das unbewusste BP vorausgeht, sondern weil die Rede von einem „Ursprung" hier keinen Sinn macht – das BP ist ebenso wenig Ursprung, denn ihm geht ja auch wieder eine Ursache voraus. Statt nach dem Ursprung einer Handlung sollte man also besser nach ihren *Ursachen* fragen.

(3.) Libets Experiment zeigt, dass Entscheidungen und Handlungen von unbewussten neuronalen Prozessen verursacht werden. Für die Freiheit ist das unproblematisch. Wodurch sonst sollten Entscheidungen und Handlungen gesteuert werden? Wir sind keine immateriellen Cartesischen *res cogitantes*, die auf mirakulöse Weise ihren materiellen Körper in Bewegung versetzen müssen, sondern komplexe physikalische Systeme, deren Steuerungsmechanismen gar nicht anders als physikalisch, in unserem Fall neuronal, realisiert sein können (Pauen, 2008, 53). Die

bloße Tatsache einer solchen neuronalen Realisierung rechtfertigt nicht die Degradierung unseres Freiheitserlebens zur Illusion: Nashörner, Hibiskusblüten und Seealgen werden nicht zur Illusion, nur weil biologische Eigenschaften mikrophysikalisch realisiert sind. Übersieht man diesen Punkt, läuft man wie hier Roth (2004, 73) Gefahr, das Gehirn zum Entscheidungsträger zu machen:

> Der Neurobiologe wird darauf hinweisen, daß der bewußte Willensakt gar nicht der Verursacher der genannten Bewegung sein könne, weil diese Bewegung bereits vorher durch neuronale Prozesse festgelegt, d. h. kausal verursacht sei. … Entsprechend müsse in der Tat die korrekte Formulierung lauten: „Nicht mein bewußter Willensakt, sondern mein Gehirn hat entschieden!"

Die Alternative „Willensakt oder Gehirn?" ist jedoch sinnlos. Entscheidungen werden schon rein sprachlich weder von Willensakten noch von Gehirnen getroffen, sondern von Personen.[9] Darüber hinaus läuft man Gefahr zu glauben, die Aufdeckung neuronaler Ursachen impliziere, dass Entscheidungen nichts zur Handlungsgenese beitragen könnten. Das ist falsch. Wie Ansgar Beckermann zu Recht bemerkt: „Es gibt keine Konkurrenz zwischen mir und meinem Gehirn. Eine Handlung kann sehr wohl meine Handlung sein, auch wenn sie auf Prozesse in meinem Hirn zurückgeht. Entscheidend ist allein, ob diese Hirnprozesse einen angemessenen internen Steuerungsmechanismus realisieren" (Beckermann, 2008, 91). *Eine* Ursache für John F. Kennedys Tod war, dass Lee Harvey Oswald mit einer geladenen Waffe auf ihn gezielt und abgedrückt hat; daraus folgt nicht, dass der Einschlag des Projektils in Kennedys Kopf nicht *auch eine* Ursache seines Todes war. Gleichermaßen gilt: Daraus, dass das BP *eine* Ursache ist, folgt nicht, dass die später auftretende bewusste Entscheidung nicht *auch eine* ist. Andernfalls ließe sich mit gleicher Berechtigung das BP als Ursache ausschließen – ihm geht ja seinerseits auch wieder eine Ursache voraus.

Libets Experiment bestätigt also weder den Determinismus noch zeigt es, dass Entscheidungen keine Ursachen von Handlungen sein können, und dass Entscheidungen nicht der Ursprung von Handlungen sind, stimmt zwar, ist aber uninteressant. Wo also liegt das Problem? Zwei Aspekte von Libets Experiment werden immer wieder als problematisch empfunden: die neuronalen Ursachen sind *unbewusst* und sie treten *vor* den bewussten Entscheidungen auf.

(4.) Freie Entscheidungen und Handlungen müssen natürlich auf bewussten Prozessen beruhen. Wer frei entscheidet, der muss Optionen

bewusst abgewogen haben, und wer frei handelt, der muss sich seiner handlungsleitenden Motive und Ziele bewusst sein. Das bedeutet jedoch nicht, dass freie Entscheidungen und Handlungen *ausschließlich* auf bewussten Deliberationsprozessen zu beruhen haben. Einerseits steuern wir Entscheidungen und Handlungen mittels der neuronalen Maschinerie in unserem Gehirn, und viele der dort ablaufenden Prozesse sind unbewusst, andererseits werden Entscheidungen und Handlungen nicht schon dadurch unfrei, dass sie von unbewussten Gefühlen und Stimmungen beeinflusst werden. „Die Annahme von Willensfreiheit ist die Annahme, dass der Wille *nicht ausschließlich* und *nicht vollständig* durch anonyme Kausalprozesse, sondern auch und oft ausschlaggebend durch selbstgewählte Gründe bestimmt ist" (Mohr, 2008, 79). Libets Experiment zeigt also, dass unbewusste Faktoren eine Rolle spielen, aber das ist unproblematisch, so lange *auch* ein bewusster deliberativer Abwägungsprozess beteiligt ist.

Aber zeigt Libets Experiment nicht gerade, dass ein deliberativer Abwägungsprozess an der Hervorbringung der Bewegung gar nicht beteiligt sein kann, weil er schlicht *zu spät* kommt?

(5.) Das Unbehagliche an Libets Experiment ist, dass die Bewegungsvorbereitung im Gehirn beginnt, *bevor* sich der Handelnde seines Bewegungsimpulses bewusst wird. Das zeigt jedoch nicht, dass die Entscheidung zur Bewegung nichts beiträgt. Ginge es um den *Ursprung* der Bewegung, wäre die zeitliche Abfolge wichtig – wäre das BP der Ursprung der Bewegung, dann könnte nichts, was danach auftritt, die Ursprungsrolle beanspruchen. Aber um den Ursprung geht es wie gesehen nicht, sondern um Ursachen, und dafür ist die zeitliche Abfolge irrelevant – frühere Ursachen schließen zeitlich spätere nicht aus. Nichtsdestotrotz bleibt das ungute Gefühl, dass, wie Roth (2003, 523) es plakativ ausdrückt, der *„Willensakt … in der Tat auf[tritt], nachdem das Gehirn bereits entschieden hat, welche Bewegung es ausführen wird"*. Aber selbst wenn der bewusste Drang („urge") erst auftritt, nachdem der Aufbau des BP bereits begonnen hat, wurden schon weit vorher relevante bewusste Entscheidungen getroffen. Die Probanden beschlossen z.B. bewusst, an dem Experiment teilzunehmen und der Anweisung zu folgen, innerhalb eines vorgegebenen Zeitrahmens eine Handbewegung auszuführen, sobald sie den Drang spürten, es zu tun. Vielleicht tritt dieser Drang erst nach dem BP auf, aber die bewusste Deliberation geht dem BP voraus – was die Probanden zeitlich nach dem BP lokalisieren ist

eventuell nur noch der Zeitpunkt, zu dem sie die Bewegung „freigeben", für die sie sich vorher entschieden haben.

Man könnte entgegnen, dies zeige lediglich, dass man zwischen der grundsätzlichen Entscheidung, am Experiment teilzunehmen, und dem konkreten Impuls, *jetzt* die Hand zu bewegen, unterscheiden müsse, und dass die Handbewegung nach wie vor ausschließlich auf unbewusste Prozesse zurückzuführen und damit unfrei sei. Wenn dem so ist, dann ist es für die Freiheit nicht tragisch. Libets „spontaner Drang" und die entsprechende Handbewegung sind denkbar schlechte Beispiele für jene Art von Entscheidung und Handlung, an deren Freiheit uns liegt. Dasselbe gilt für Folgestudien wie etwa von Haggard und Eimer (1999), in der Probanden entscheiden konnten, ob sie den linken oder den rechten von zwei Knöpfen drückten. Typische Alltagsentscheidungen und -handlungen sehen anders aus. Üblicherweise haben wir mehr als zwei Optionen: Wir überlegen uns, an welcher Universität wir studieren sollen, wo wir nächstes Jahr unseren Urlaub verbringen oder bei welcher Zeitschrift wir einen Aufsatz einreichen. Üblicherweise haben wir an unseren Entscheidungen auch ein persönliches Interesse. Der Drang, die Hand *jetzt* zu bewegen, oder die Entscheidung, den linken oder den rechten Knopf zu drücken, sind für uns hingegen mit keinem intrinsischen Interesse verbunden – das ist gerade der Grund, warum sie, ganz unüblich, ohne bewusste Deliberation zustande kommen können.[10] Selbst wenn solche Bewegungen also ausschließlich unbewusst zustande kommen, bedeutet dies noch lange nicht, dass jene Handlungen, auf die es uns ankommt, nicht maßgeblich auch von bewussten Faktoren beeinflusst sind, und komplexe Entscheidungen und Handlungen der Art, wie sie im Alltag eine Rolle spielen, sind experimentell (bislang) nicht zu kontrollieren (Pockett, 2006) und damit (bislang) nicht durch Libet-artige Untersuchungen als unfrei zu erweisen.

4. Hirnstimulation und Kontrollillusionen

So genannte „Kontrollillusionen" haben Libet inzwischen als Lieblingsbeispiel der empirischen Freiheitsskeptiker abgelöst. So schreibt z. B. Roth (2006, 10): „Man kann Versuchspersonen unterschwellig (z. B. über maskierte Reize) durch experimentelle Tricks, Hypnose oder Hirnstimulation zu Handlungen veranlassen, von denen sie später behaupten, sie

hätten sie *gewollt*", und an anderer Stelle nahezu wortgleich: „Schließlich sind zahlreiche Befunde bekannt, bei denen Versuchspersonen aufgrund von Hypnose und Patienten aufgrund von Hirnstimulationen Bewegungen ausführen, die sie als von ihnen gewollt empfinden" (Merkel und Roth, 2008, 61).

Für die Frage nach der Freiheit im Allgemeinen sind solche Befunde jedoch kaum einschlägig – dass wir dazu gebracht werden können, etwas zu tun, das wir im Nachhinein als von uns gewollt empfinden, impliziert nicht, dass jene Handlungen, die wir tatsächlich gewollt haben, unfrei sind. Bestenfalls zeigt es, dass unsere introspektiven Kausalurteile darüber, welche Wirkungen auf unser Tun zurückgehen und welche auf das Tun anderer, fallibel sind. Aber erstens ist die Fallibilität sowohl introspektiver als auch kausaler Urteile nichts Neues, und zweitens handelt es sich um Täuschungen, die in Extremsituationen auftreten. Niemand bestreitet z. B., dass unter Hypnose ausgeführte Handlungen unfrei sind, aber was soll daraus für die Freiheit im Allgemeinen folgen? Für maskierte Reize, experimentelle Tricks und Hirnstimulationen gilt dasselbe.

Ganz unabhängig davon muss man fragen, wie gut Roths Behauptung, man könne Versuchspersonen „zu Handlungen veranlassen, von denen sie später behaupten, sie hätten sie gewollt", empirisch abgesichert ist. In Bezug auf Hirnstimulationen z. B. findet sich in Roth (2003, 515) folgende Passage:

> Bei einer Reihe von Patienten führte jedoch die Stimulation eines Cortexareals am Fuß der Zentralfurche im Übergang zur Sylvischen Furche zuverlässig zum Willen bzw. Bedürfnis, die linke bzw. rechte Hand oder den linken oder den rechten Fuß zu bewegen (Penfield und Rasmussen, 1950). … Ähnlich konnte mithilfe der Transkranialen Magnetstimulation (TMS) der Neurologe Brasil-Neto Fingerbewegungen auslösen, die die Versuchsperson als „gewollt" beschrieb …

Zweifellos sind viele Fälle dokumentiert, in denen die Stimulation von Kortexarealen zu Kitzelempfindungen, Spasmen oder auch Bewegungen führte (Halgren *et al*., 1993), aber üblicherweise beschrieben die Patienten diese Phänomene als aufgezwungen, nicht als gewollt. Darüber hinaus kann man in diesen Fällen nicht von *Handlungen* in einem substanziellen Sinn sprechen – es mag sich ein Arm oder ein Bein heben, aber noch niemand hat aufgrund einer Hirnstimulation den Entschluss gefasst, sich am MIT zu bewerben und diesen anschließend als seinen eigenen ausgegeben. Wie Roth selbst sagt, führte die Stimulation bei Penfield und

Rasmussen (1950) noch nicht einmal zu einer Bewegung, sondern zum „Willen bzw. Bedürfnis" eine Bewegung auszuführen. Wie dies belegen soll, dass Personen zu Handlungen veranlasst werden können, die sie hinterher als gewollt empfinden, bleibt Roths Geheimnis.[11]

In der von Roth ebenfalls zitierten Studie von Brasil-Neto et al. (1992) wurden Probanden aufgefordert, nach eigenem Belieben den linken oder den rechten Zeigefinger zu bewegen. Mittels TMS konnte die relative Häufigkeit, mit der der linke bzw. der rechte Finger bewegt wurde, minimal beeinflusst werden, obwohl die Probanden im Nachhinein bekundeten, frei entschieden zu haben. Wie man daraus ableiten kann, Brasil-Neto hätte „Fingerbewegungen auslösen [können], die die Versuchsperson als ‚gewollt' beschrieb", bleibt wiederum Roths Geheimnis. Ausgelöst wurde durch die TMS gar nichts, zumindest keine Fingerbewegungen, sondern es wurde die relative Häufigkeit der Wahl beeinflusst. Es ging auch nicht darum, dass die Fingerbewegung als „gewollt" beschrieben wurde (was im Übrigen völlig korrekt gewesen wäre, denn die Fingerbewegung war ja gewollt), sondern darum, dass die Probanden den Einfluss der TMS nicht bewusst wahrnahmen. Und schließlich gilt hier dasselbe wie für Libet und die Studien zu Hirnstimulationen: Vielleicht war die Kontrolle der Probanden so weit herabgesetzt, dass man ihre Entscheidung, den linken oder den rechten Finger zu bewegen, nicht mehr uneingeschränkt als „frei" bezeichnen kann (vgl. Abschnitt 7), aber über die Freiheit „echter" Alltagshandlungen und -entscheidungen, an denen wir ein persönliches Interesse haben und an deren Freiheit uns daher tatsächlich gelegen ist, sagt das wenig aus.

Dasselbe gilt im Wesentlichen auch für sonstige experimentelle Tricks und maskierte Reize. Üblicherweise kann von der *Auslösung* einer *Handlung* oder Entscheidung, die im Nachhinein *als „gewollt" beschrieben wird*, nicht die Rede sein. In der berühmten *I-Spy* Studie von Daniel Wegner (2002; Wegner und Wheatley, 1999) z.B., auf die sich Roth und andere empirische Freiheitsskeptiker immer wieder beziehen, konnten je zwei Personen mittels einer Vorrichtung gemeinsam einen Cursor auf dem Bildschirm bewegen, auf dem verschiedene Objekte zu sehen waren. Sie sollten mit dem Cursor nach etwa 30 bis 40 Sekunden auf eines dieser Objekte zeigen und auf einer Skala von 0 bis 100 angeben, wie groß ihr Anteil an der Entscheidung war, den Cursor dort zum Halten zu bringen. Eine der beiden Personen war ein Mitarbeiter, der ab und zu auf Anweisung den Cursor über einem bestimmten Objekt platzierte.

Wurde der anderen Person durch den Experimentator jedoch über Kopfhörer die Bezeichnung des Objekts, auf dem der Mitarbeiter kurz darauf den Cursor zum Halten brachte, mitgeteilt, berichtete sie hinterher fälschlich, einen nicht unerheblichen Anteil an der Entscheidung gehabt zu haben.

Zu den methodologischen Schwierigkeiten mit Wegners Studie und ihrer philosophischen Interpretation ließe sich viel sagen. Ich beschränke mich an dieser Stelle auf ein paar Punkte, die im gegenwärtigen Zusammenhang besonders wichtig sind. Mit Roths Behauptung, wir könnten Personen zu Handlungen veranlassen, von denen sie später behaupten, sie hätten sie gewollt, hat Wegners Studie nichts zu tun – die Entscheidung bzw. Handlung war ja gerade *nicht* die des Probanden, sondern die des Mitarbeiters. Wegners eigene Einschätzung, seine Studie zeige, es sei möglich „to lead people to feel that they have performed a willful action when in fact they have done nothing" (2002, 74) ist angemessener, aber immer noch irreführend. Die Probanden bemaßen lediglich ihren Anteil an einer gemeinsamen Entscheidung zu hoch, sie glaubten nicht, sie alleine getroffen zu haben (die höchsten Zuschreibungswerte lagen bei ca. 62 %). Unsere Freiheit wäre bedroht, wenn wir Entscheidungen träfen und Handlungen ausführten, die nicht unserer Kontrolle unterliegen – aber das kann man nicht dadurch zeigen, dass man in Situationen, in denen überhaupt nichts entschieden oder getan wird, mittels experimenteller Tricks ein Kontrollgefühl induziert.

Dasselbe gilt für Studien mit maskierten Reizen. Linser und Goschke (2007) konnten in Probanden das Gefühl induzieren, durch Drükken der linken bzw. rechten Taste die Farbe eines auf dem Bildschirm auftauchenden Kreises zu kontrollieren, indem sie ihnen kurz vor dem Drücken der Taste subliminal die Farbe darboten, die anschließend tatsächlich auftauchte. Auch hier gilt: Diese Studie hat weder mit Roths Behauptung, man könne Personen zu Handlungen veranlassen, die sie später als gewollt empfinden,[12] noch mit der Freiheitsproblematik etwas zu tun. Die Entscheidung, die linke oder rechte Taste zu drücken, wurde nicht beeinflusst, sondern von den Probanden frei getroffen. Beeinflusst wurde lediglich das subjektiv empfundene Gefühl, Kontrolle über die Farbe des später auftauchenden Kreises zu haben.

Ließe sich empirisch zeigen, dass eine Person, die eine alltägliche Handlungsentscheidung in dem festen Glauben trifft, sie unterläge ihrer Kontrolle, über diese Kontrolle nachweislich nicht verfügt, dann wäre

es in der Tat fraglich, ob wir sie in dieser konkreten Situation uneingeschränkt als „frei" bezeichnen sollten. Die bislang betrachteten Studien reichen dafür allerdings bei weitem nicht aus. Erstens haben sie mit Entscheidungen und Handlungen überhaupt nichts zu tun, schon gar nicht mit alltäglichen. Zweitens zeigt die Tatsache, dass man die Illusion einer Kontrolle hervorrufen kann, obwohl überhaupt keine eigene Entscheidung oder Handlung involviert war, nicht, dass das Gefühl der Kontrolle über die eigenen Entscheidungen und Handlungen illusorisch ist. Drittens wäre selbst dann, wenn sich unser Kontrollgefühl über unsere eigenen Entscheidungen und Handlungen experimentell in die Irre führen ließe, immer noch nicht gezeigt, dass dies im Alltag immer so ist.

5. Wissenschaft und Freiheit

Eine empirische Widerlegung der Freiheit ist bislang nicht in Sicht. Die üblicherweise angeführten Überlegungen zu einem allgemeinen oder bereichsspezifischen Determinismus, zu Libets Experiment (und vergleichbaren Studien) sowie zu Hirnstimulationen und Kontrollillusionen können die radikale These der Illusion Freiheit nicht stützen. *Es stimmt ganz einfach nicht*, dass „[n]eurowissenschaftliche Forschung … inzwischen unmissverständlich klar gemacht [hat], dass die Vorstellung eines bewusst erlebten freien Willens als Auslöser einer Handlung und damit als deren Motiv nicht in Rechnung gestellt werden kann" (Grün *et al.*, 2008, 7). *Es stimmt ganz einfach nicht*, dass die „Ergebnisse empirischer Forschung dazu zwingen, den Gedanken an eine Autonomie des Willens aufzugeben" (Wuketits, 2007, 38). *Es stimmt ganz einfach nicht*, dass die Naturwissenschaft die Geisteswissenschaft ein weiteres Mal „ihrer Bodenlosigkeit überführt" hat (Grün, 2008, 29). Wer dies in aller Öffentlichkeit so sagt, der lenkt die öffentliche Meinungsbildung in eine Richtung, die wissenschaftlich nicht abzusichern ist, und verstößt damit gegen die Grundsätze wissenschaftlicher Redlichkeit.

Das soll nicht heißen, dass sich die Philosophie ihren Freiheitsbegriff unwidersprochen so zurechtzimmern kann, dass er gegen empirische Einwände grundsätzlich immun ist. Ich behaupte lediglich, dass die in der von der Öffentlichkeit wahrgenommenen Freiheitsdebatte immer wieder angeführten vermeintlichen Belege unserer Unfreiheit unzureichend sind und dass viele Behauptungen in dieser Debatte bei Licht

betrachtet blanker Unsinn sind. Nichtsdestotrotz glaube ich, dass alltägliche Entscheidungen und Handlungen nicht so frei sind, wie es uns die Philosophie und unsere Selbstwahrnehmung weismachen möchten, und dass dies empirisch nachweisbar ist. Bevor ich abschließend kurz erläutere, wo eine empirische Einschränkung der Freiheit meines Erachtens ansetzen sollte, zunächst ein paar Bemerkungen zu zwei Ideen, die in gegenwärtigen Freiheitstheorien einen wichtigen Platz einnehmen: *Kontrolle* und *normative Einbettung*.[13]

6. Freiheit als Kontrolle und normative Einbettung

Einer gegenwärtig weit verbreiteten Variante des Kompatibilismus zufolge besteht Freiheit im „zustimmende[n] Handlungsvollzug, der sich auf abgewogene Präferenzen zurückführen lässt" (Wuchterl, 2007, 46). Frei ist der, der sein Entscheiden und Handeln *kontrolliert*. „Kontrolle" impliziert hier kein substanzielles Anders-Handeln-Können, sondern bezieht sich auf den Besitz rationaler Fähigkeiten. So argumentiert z. B. Beckermann (2008, 114) dafür, dass eine Person dann frei ist, „wenn sie zwei Fähigkeiten besitzt – die Fähigkeit, vor dem Handeln innezuhalten und zu überlegen, und die Fähigkeit, dem Ergebnis dieser Überlegung gemäß zu entscheiden und zu handeln". Michael Pauen vertritt eine ähnliche Position: „Wenn ein Mensch aufgrund der ihm zuschreibbaren Motive handelt, dann handelt er selbstbestimmt und damit frei" (Pauen und Roth, 2008, 176). Im angelsächsischen Sprachraum macht Jay Wallace „general powers of reflective self-control" (Wallace, 1994, 157) ebenso zum Maßstab für Freiheit wie Fischer und Ravizza (1998), die eine Handlung als frei einstufen, wenn sie aus für Gründe empfänglichen und dem Handelnden eigenen Mechanismen herrührt.

Neben Kontrolle spielt die *normative Einbettung* der Entscheidung oder Handlung in das Präferenz- und Werteprofil des Handelnden eine wichtige Rolle. Frei ist der, der nicht nur aus der Abwägung von Gründen heraus handelt, sondern sich mit diesen Gründen identifiziert, sie als die seinen akzeptiert und reflektierend anerkennt, dass die Entscheidung oder Handlung vor dem Hintergrund seines Präferenz- und Werteprofils betrachtet richtig ist: „Unser Wille ist frei, wenn er sich unserem Urteil darüber fügt, was zu wollen richtig ist", so Bieri (2005, 125). Wallace z. B. bezeichnet Kontrolle explizit als das Vermögen „to grasp and apply

moral reasons" und die Fähigkeit „to control or regulate one's behavior by the light of such reasons" (Wallace, 1994, 157; Hervorhebung S. W.). Ebenso deutlich tritt die Idee normativer Einbettung bei Harry Frankfurt zu Tage, dem zufolge ein frei Entscheidender und Handelnder bereit sein muss „to endorse or repudiate the motives from which he acts … to guide his conduct *in accordance with what he really cares about*" (Frankfurt, 1993, 23; Hervorhebung S. W.). Frei ist also, wer es schafft, sein Entscheiden und Handeln durch jene Wünsche erster Stufe leiten zu lassen, von denen er auf zweiter Stufe möchte, dass sie wirksam werden (Frankfurt, 1971). Auch Gerald Dworkin (1988, 61) betont, ein Entscheidender oder Handelnder sei autonom, „if he identifies with his desires, goals, and values, and such identification is not influenced in ways which make the process of identification alien to the agent". Susan Wolf schließlich knüpft Freiheit an die Fähigkeit, das Gute und Wahre zu erkennen und zu verwirklichen – frei ist, wer aus den richtigen Gründen das rational und moralisch Richtige tut: „responsibility depends on the ability to act in accordance with the True and the Good. If one is psychologically determined to do the right thing for the right reasons, this is compatible with having the requisite ability" (1990, 79).

Kontrolle und normative Einbettung spielen darüber hinaus auch in libertarischen Freiheitskonzeptionen eine Rolle. Timothy O'Connor (2000, 61) z.B. behauptet: „Exerting active power is intrinsically a direct exercise of control over one's behavior" und: „I … am unable to conceive an agent's directly controlling his own activity without any awareness of what is motivating him" (2000, 88). Und Laura Ekstrom zufolge ist ein Akteur nur dann frei, wenn „the agent's act results from a preference – that is, a desire formed by *a process of critical evaluation with respect to one's conception of the good*" (2000, 108; Hervorhebung S. W.).

Eine empirische Widerlegung oder Einschränkung der Freiheit muss also zeigen, dass alltäglichen Entscheidungen und Handlungen normaler gesunder Erwachsener die eben skizzierte Art von Kontrolle und normativer Einbettung fehlt, zumindest im vollen Umfang. Man kann dies tun und zwar durch genau jene sozialpsychologischen Studien, die Bargh als Evidenz für einen psychologischen Determinismus zu verkaufen versucht (vgl. Abschnitt 2). Diese Studien zeigen, dass die Steuerung vermeintlich selbst initiierter Entscheidungen und Handlungen unbewusst auf eine Art und Weise beeinflusst werden kann, dass Kontrolle und normative Einbettung nachhaltig eingeschränkt sind.

7. Sozialpsychologische Befunde zur unbewussten Handlungssteuerung

Dutton und Aron (1974) ließen männliche Passanten von einer jungen Frau ansprechen, die sie bat, einen kurzen Fragebogen auszufüllen. Am Ende notierte sie ihre Telefonnummer und forderte die Angesprochenen auf, sie anzurufen, wenn sie sich treffen und Näheres über die Umfrage erfahren wollten. Einige wurden auf einer 450 Fuß langen und 230 Fuß hohen, schmalen und schwankenden Hängebrücke angesprochen, einige auf einer breiten, soliden und nur 10 Fuß hohen Holzbrücke. Während von den 16 auf der Holzbrücke Angesprochenen nur 2 zurückriefen, griffen von den 18 auf der Hängebrücke Angesprochenen 9 zum Telefon.[14]

Darley und Latane (1968) ließen Versuchspersonen mittels Gegensprechanlage entweder mit einem oder mit mehreren anderen kommunizieren. Eine Versuchsperson hörte im Verlauf des Experiments, wie der bzw. einer der andere(n) einen epileptischen Anfall hatte, nach Luft schnappte und um Hilfe rief. War das vermeintliche Opfer der einzige Partner, kamen ihm 100 % der Versuchspersonen zu Hilfe, waren jedoch weitere Partner zugeschaltet, nur 60 %. In einer anderen Studie von Latane und Darley (1970) halfen 70 % der Versuchsteilnehmer einer Mitarbeiterin, die scheinbar schwer zu Fall kam, wenn sie die einzigen Anwesenden waren, aber nur 12%, wenn noch jemand dabei war, der untätig blieb.

In einer Studie von Van Baaren et al. (2003) erhielten Bedienungen in einem niederländischen Restaurant öfter und mehr Trinkgeld, wenn sie die Bestellungen wörtlich wiederholten und nicht nur notierten, ansonsten aber gleich freundlich waren und ihre Aufmerksamkeit auf andere Weise demonstrierten. 81 % der Gäste, deren Bestellungen wörtlich wiederholt worden waren, gaben ein Trinkgeld (durchschnittlich 2,97 Niederländische Gulden), aber nur 61 % der Kontrollgruppe (durchschnittlich 1,76 Gulden).

Todorov et al. (2005) zeigten Versuchspersonen paarweise Schwarzweißfotografien republikanischer und demokratischer Kandidaten der US-amerikanischen Wahlen für den Senat. In 71,6 % der Fälle war der Kandidat, der aufgrund seines Gesichts als kompetenter eingeschätzt wurde, auch derjenige, der gewählt worden war. In einer Studie von Antonakis und Dalgas (2009) konnten 681 Schweizer Kinder im Alter

zwischen 5 und 13 Jahren anhand von Fotos die Gewinner der Wahlkreise der französischen Parlamentswahlen mit einer Wahrscheinlichkeit von 0,72 korrekt vorhersagen, indem sie angaben, welchen von beiden Kandidaten sie lieber als Kapitän eines Schiffes hätten.

„Der paradigmatische Fall einer aus freiem Willen begangenen Handlung", so fasst Frank Hofmann (2008, 166) die in Abschnitt 6 skizzierte Grundthese moderner kompatibilistischer Freiheitstheorien treffend zusammen, ist eine Handlung, „die aus reflektierter und abwägender normativer Urteilsbildung über die Gründe und das Gute im Handeln hervorgeht". Die oben referierten Studien sollten erhebliche Zweifel daran aufkommen lassen, dass laut unserer Selbstwahrnehmung paradigmatische Fälle freien Entscheidens und Handelns diese Bedingungen immer uneingeschränkt erfüllen. Gemäß Pauen (2008, 49) bedeutet die Tatsache, dass sich eine Person „nur in einem sehr eingeschränkten Sinne über die eigenen Präferenzen sowie die Konsequenzen bestimmter Entscheidungen für die übrigen Ziele im Klaren ist …, dass hier nur in Ansätzen von einer Fähigkeit zu Selbstbestimmung gesprochen werden kann", und genau das ist meines Erachtens oftmals der Fall.

Wenn die Wahrscheinlichkeit, mit der ich eine junge Frau zurückrufe, davon abhängt, ob ich sie auf einer Hänge- oder einer Holzbrücke getroffen habe, mir dieser Einfluss aber während des Abwägungsprozesses nicht bewusst ist, dann beeinflusst das meine Fähigkeit reflektierter und abwägender normativer Urteilsbildung. Wenn ich wochenlang Wahlprogramme studiere, meine Entscheidung aber letztlich zu einem nicht unerheblichen Teil von unbewussten und irrelevanten Faktoren beeinflusst wird, dann unterliegt mein Tun nicht uneingeschränkt meinem Denken und Urteilen, und alles Innehalten und Überlegen ist zwecklos. *Kontrolle sieht anders aus*.

Wenn die Wahrscheinlichkeit, mit der ich einen Hilfebedürftigen unterstütze, systematisch durch die Zahl der anderen Anwesenden beeinflusst wird, dann füge ich mich mit meinem Handeln nicht meinem „Urteil darüber, was zu wollen richtig ist", wie Bieri es formuliert, denn dieses Urteil besagt, ich sollte unabhängig von der Anzahl der sonstigen Beteiligten helfen. Wenn sich herausstellt, dass ich unter anderem deshalb ein Trinkgeld gegeben habe, weil die Bedienung meine Bestellung wörtlich wiederholt hat, dann handle ich nicht in Wolfs Sinne aus Einsicht in „the True and the Good" und auch nicht in Frankfurts Sinne „in accordance with what I really care about". Wie Pauen (2008, 50) es ausdrückt: „wenn

ich mir nicht darüber im Klaren bin, dass eine bestimmte Handlung im Widerspruch zu einer meiner Präferenzen steht, dann leidet darunter … meine Fähigkeit, selbstbestimmt zu handeln". *Normative Einbettung sieht anders aus*.

Insofern Kontrolle und normative Einbettung für Freiheit maßgebend sind, schränken die oben genannten Studien die Freiheit unserer Entscheidungen und Handlungen also empirisch ein. Hierzu ein paar abschließende Bemerkungen.

(1.) Die Einschränkung betrifft jeden Freiheitsbegriff, der Freiheit an Kontrolle und normativer Einbettung festmacht, unabhängig davon, ob er libertarisch oder kompatibilistisch ist. Der Kompatibilismus ist also gegenüber empirischen Befunden keineswegs immun.

(2.) Die oben angeführten Studien zeigen, *pace* Bargh (vgl. Abschnitt 2), nicht, dass Entscheidungen und Handlungen durch unbewusste Faktoren *determiniert* werden. Die aufgedeckten Abhängigkeiten sind immer nur probabilistisch (wenn auch natürlich statistisch signifikant) und zeigen somit keine unbewusste *Determination* unserer Entscheidungen und Handlungen, sondern nur eine unbewusste *Beeinflussung*.

(3.) Dass diese unbewusste Beeinflussung hier als Einschränkung der Freiheit interpretiert wird, widerspricht nicht der in Abschnitt 3 gemachten Behauptung, dass die Tatsache, dass an der Hervorbringung von Entscheidungen und Handlungen unbewusste Faktoren beteiligt sind, die Freiheit so lange nicht bedroht, wie diesen Entscheidungen und Handlungen *auch* ein bewusster Deliberationsprozess vorausgeht. So lange ein bewusster Deliberationsprozess den Ausgang der Entscheidung oder Handlung maßgeblich mitbestimmt, mag die Existenz unbewusster Faktoren grundsätzlich tolerierbar sein. Wie in Abschnitt 3 erwähnt, kann man der Meinung sein, jede Theorie von Freiheit müsse anerkennen, dass z. B. unbewusste Stimmungen, die unseren affektiven Hintergrund strukturieren, einen Einfluss auf unser Entscheiden oder Handeln ausüben können, ohne dass dadurch deren Freiheit unterminiert wird. Auch wenn ich dieser liberalen Einstellung gegenüber unbewussten Faktoren durchaus etwas abgewinnen kann, gibt es Grenzen. Der bewusste Deliberationsprozess muss immer noch eine nachhaltige, nicht unerhebliche Rolle spielen. Wird diese Grenze überschritten, und in den oben zitierten Studien scheint genau das der Fall zu sein, sollten wir, auch wenn wir noch so liberal gesinnt sind, nicht mehr von Freiheit sprechen.[15]

(4.) Es ist keineswegs so, dass die Studien der empirischen Sozialpsy-

chologie nur bestätigen, was sich jeder, der über etwas Lebenserfahrung verfügt, sowieso längst hätte klar machen sollen – dass nämlich unser Entscheiden und Handeln oft von anderen Motiven abhängt als wir glauben. Diese Studien gehen über den Erkenntnisgewinn, den wir aus dem „gesunden Menschenverstand" ziehen können, in zweierlei Hinsicht hinaus. Erstens zeigen sie, dass die unbewussten Einflüsse auf unser Entscheiden und Handeln wesentlich häufiger, wesentlich stärker und von gänzlich anderer Art sind als es für den reflektierenden „Ottonormalverbraucher" im Lehnstuhl den Anschein haben mag. Zweitens, und das ist der in meinen Augen wichtigere Punkt, lassen sie uns erkennen, dass die unbewussten Faktoren in vielen Fällen dergestalt sind, dass wir ihren Einfluss auf unser Entscheiden und Handeln missbilligten, machten wir ihn uns bewusst – und das ist etwas, das wir uns mit seinen freiheitsunterminierenden Konsequenzen ohne diese Studien nicht vollends klar machen würden.

(5.) Im Gegensatz zu den Studien zu durch maskierte Reize ausgelösten Kontrollillusionen ist die unbewusste Beeinflussung hier nicht an aufwendige Laborsituationen und alltagsfremde Entscheidungen und Handlungen gebunden, an denen der Entscheidende und Handelnde kein persönliches Interesse hat. Es geht vielmehr um Wahlentscheidungen, um (unterlassene) Hilfeleistung und darum, ob man eine Person attraktiv genug findet, um sich mit ihr zu verabreden, und die Einflussfaktoren sind durchgängig solche, die auch in Alltagssituationen auftreten können, keine subliminal dargebotenen Reize.

(6.) Wir sind nicht wirklich *unfrei*, unsere Freiheit ist bloß *eingeschränkt*. Im Gegensatz zur Schwangerschaft ist Freiheit keine Alles-oder-nichts-Angelegenheit – man kann auch ein bisschen frei sein, oder mehr oder weniger, denn Freiheit ist primär eine Eigenschaft konkreter Entscheidungen und Handlungen. Jemanden losgelöst von einer konkreten Entscheidungs- oder Handlungssituation generell „frei" zu nennen, macht keinen Sinn. Freiheit ist zweifach graduell: Entscheidungen und Handlungen können erstens mehr oder weniger frei sein, abhängig davon, wie sehr ihre Steuerung der Kontrolle der entscheidenden und handelnden Person unterliegt und wie stark ihre normative Einbettung ist. Die entscheidende und handelnde Person kann zweitens in einem mittelbaren Sinn mehr oder weniger frei sein, abhängig davon, wie viele ihrer Entscheidungen und Handlungen in welchem Maß frei sind.

(7.) In der Debatte zwischen der Philosophie und der empirischen

Wissenschaft begehen also beide Seiten einen entscheidenden Fehler: Man kann empirisch nicht zeigen, dass wir unfrei sind, man kann aber philosophisch auch nicht zeigen, dass wir frei sind. Nicht etwa, weil die jeweilige Argumentation der Sache nach fehlerhaft ist, sondern ganz einfach deshalb, weil „Frei oder nicht?" die völlig falsche Frage ist. Wer sich um die Freiheit sorgt, der muss vielmehr fragen: Wie frei sind wir eigentlich? Und wie oft?

Die oben angeführten empirischen Befunde zeigen, dass unsere Freiheit doppelt eingeschränkt ist. Wir sind erstens *seltener* und zweitens *weniger* frei als es unsere Selbstwahrnehmung suggeriert. Herauszufinden wie stark das Ausmaß dieser Einschränkung ist, ist Aufgabe der empirischen Sozialpsychologie. Und *nur* der empirischen Sozialpsychologie – den Determinismus, Libet, Hirnstimulationen und Kontrollillusionen sollten wir endlich getrost und endgültig zu den Akten legen.

Anmerkungen

* Frühere Fassungen dieser Arbeit wurden im November 2007 am Forum für Philosophie in Frankfurt, im Oktober 2008 auf einer Konferenz zum Thema *Brauchen wir eine neue Ethik? Herausforderungen der Moral durch die Neurowissenschaft* in Saarbrücken und im Januar 2009 an der Universität Hannover vorgetragen. Ich danke allen Kollegen und Zuhörern, die mir bei diesen und anderen Gelegenheiten durch Fragen oder Anmerkungen geholfen haben, besser zu verstehen, was ich eigentlich sagen möchte.

1 John Fisher und Mark Ravizza argumentieren bekanntlich dafür, dass wir verantwortlich sein können, ohne frei zu sein, weil Verantwortlichkeit anders als Freiheit kein Anders-Handeln-Können erfordert (Fischer, 1994; Fisher und Ravizza, 1998; Ravizza, 1994). Achim Lohmar (2005) verteidigt eine ähnliche Position. Auch Schuld und Strafe kann man von der Freiheit abzukoppeln versuchen, z. B. indem man die Kant'sche Retributionsidee durch ein Bentham'sches Maßregelrecht ersetzt, das Strafe durch die Sicherheitsinteressen des Staates und seiner Bürger legitimiert, so dass Strafe letztlich einzig der Vermeidung künftiger Normverletzungen und der Aufrechterhaltung des Rechts- und Sozialsystems dient. So z. B. Reinhard Merkel:
„Die Straftat verursacht einen Riss in der normativen Welt. Die Strafe kann die Welt nicht wirklich heilen …, aber sie kann den Fortbestand der normativen Welt sichern … Deshalb darf das Recht für die Kosten der unvermeidlichen Reparatur den „bezahlen" lassen, der den Riss erzeugt hat. Das ist auch dann nicht unfair, wenn der Täter möglicherweise nichts für seine Tat konnte [weil er aufgrund fehlender Freiheit die Norm gar nicht einhalten konnte; S. W.] …" (2008, 367–368)

2 Laut Bettina Waldes (2006) *epistemischem Libertarismus* ist die (mit dem Determinismus verträgliche) Tatsache, dass wir in gewöhnlichen Entscheidungs- und Handlungssituationen über *epistemisch* offene Möglichkeiten verfügen, d.h. schlicht nicht wissen, welcher zukünftige Weltverlauf ontologisch festgelegt ist, Grundlage unserer Freiheit.

3 Handlungsalternativen reduzieren sich in diesem Zusammenhang üblicherweise darauf, dass wir anders hätten handeln können, hätten wir uns – *per impossibile* – anders entschieden. Diese Art von Anders-Handeln-Können ändert jedoch nichts daran, dass der Kompatibilist uns für Normverletzungen verantwortlich macht, die *sensu strictu* unvermeidbar waren. Es bleibt scheinbar bloß Merkels Trost, dass dies insofern nicht unfair ist, als es den Fortbestand der normativen Welt sichert (vgl. Fußnote 1).

4 *Be*schreiben! Gesetze haben keine normative Kraft, d.h. sie können der Welt nicht *vor*schreiben, wie sie zu verlaufen hat. Es ist also Unsinn zu sagen, der Determinismus impliziere, dass die Naturgesetze zusammen mit den Anfangsbedingungen den zukünftigen Weltverlauf *festlegen*.

5 Dieses „somit" hat natürlich nur dann seine Berechtigung, wenn die dem Abwägungsprozess zugrunde liegenden neuronalen Prozesse ihrerseits deterministischen Gesetzen unterliegen.

6 Man könnte natürlich argumentieren, die „unhöflichen" Probanden seien eben durch andere Faktoren determiniert gewesen, das Gespräch doch nicht zu unterbrechen. Ohne Belege für diese Behauptung zeigt sie jedoch bloß einmal mehr, dass der Determinismus vorausgesetzt wird, ohne selbst beweisbar zu sein.

7 Insbesondere werde ich annehmen, dass das BP tatsächlich zuverlässig vor dem Bewusstwerden der Entscheidung auftritt – eine Annahme, die man mit guten Gründen hinterfragen kann (Rösler, 2008; Trevena und Miller, 2002).

8 Hätte Prinz ein wenig gründlicher nachgelesen, wäre ihm aufgefallen, dass Libet selbst klar gesehen hat, dass der Determinismus keine empirische Behauptung ist: „The assumption that a deterministic nature of the physically observable world (to the extent that may be true) can account for conscious functions and event [sic!], is a speculative *belief*, not a scientifically proven position" (2002, 562).

9 Auch sonst ist die zitierte Passage reichlich wirr: Sollen die neuronalen Prozesse die Bewegung nun festgelegen oder verursachen? ("Kausale Verursachung" suggeriert im Übrigen, dass es auch non-kausale Verursachung gibt, was ebenfalls Unsinn ist). Gegen Letzteres ist nichts einzuwenden, aber den Ausdruck „festlegen" sollte man vermeiden – mit dem Determinismus hat Libets Experiment wie gesehen nichts zu tun.

10 Libet forderte die Probanden ausdrücklich auf, den Zeitpunkt der Bewegung nicht im Voraus zu planen. Welche Vorstellung von Freiheit erfordert es, so ist man geneigt zu fragen, auf diesen Punkt so großen Wert zu legen? Ist meine Entscheidung, heute Abend einen Freund anzurufen unfrei, nur weil ich mir im Vorfeld vornehme, ihn um genau neun Uhr anzurufen?

11 Selbst Roths Rede vom Willen bzw. Bedürfnis ist stärker als das, was Penfield und Rasmussen (1950) tatsächlich sagen, denn bei ihnen heißt es nur

„*she felt as though* she wanted to move her left hand" (Hervorhebung S. W.). Diesen Punkt verdanke ich Löffler (2007), der eine viel ausführlichere und exzellente Kritik der Unstimmigkeiten in Roths Ausführungen liefert.

12 Roth selbst führt die Studie von Linser und Goschke nicht an.

13 Die beiden folgenden Abschnitte fassen im Wesentlichen die Diskussion in Walter (in Begutachtung) zusammen.

14 Dutton und Aron führen dies darauf zurück, dass die auf der Hängebrücke Angesprochenen ihre körperliche Erregung fälschlich der Interviewerin statt dem ungewöhnlichen Ort zuschrieben und unbewusst als sexuelle Anziehung interpretierten.

15 Wer nicht gewillt ist, in jeder Art von unbewusster Beeinflussung bereits eine Einschränkung unserer Freiheit zu sehen, der sieht sich natürlich mit der berechtigten Frage konfrontiert, wo genau diese Grenze zu ziehen ist, d. h. welches Ausmaß die unbewusste Beeinflussung annehmen muss, um in einem nachhaltigen Sinn freiheitsuntergrabend zu sein. Eine präzise Antwort auf diese Frage erforderte in meinen Augen u. a. eine ausgearbeitete Theorie von Kontrolle, die es derzeit nicht gibt. Fest scheint mir jedoch zu stehen, dass Freiheit in dem Moment unterminiert wird, wo die normative Einbettung in das Präferenz- und Werteprofil des Entscheidenden und Handelnden nicht mehr uneingeschränkt gegeben ist.

Literatur

Antonakis, John; Dalgas, Olaf, 2009: Predicting elections. In: *Science* 323, S. 1183.

Bargh, John; Chartrand, Tanya, 1999: On the unbearable automaticity of being. In: *American Psychologist* 54, S. 462–479.

Bargh, John; Chen, Mark; Burrows, Lara, 1996: Automaticity of social behavior. In: *Journal of Personality and Social Psychology* 71, S. 230–244.

Bargh, John; Ferguson, Melissa, 2000: Beyond behaviorism. In: *Psychological Bulletin* 126, S. 925–945.

Beckermann, Ansgar, 2008: *Gehirn, Ich, Freiheit.* Paderborn: mentis.

Bieri, Peter, 2001: *Das Handwerk der Freiheit*. München: Hanser.

Bieri, Peter, 2005: Unser Wille ist frei. In: *Der Spiegel* 2, 10.01.2005, S. 124–125.

Brasil-Neto, Joaquim; Pascual-Leone, Alvaro; Valls-Solé, Josep; Cohen, Leonardo; Hallett, Mark, 1992: Focal transcranial magnetic stimulation and response bias in a forced-choice task. In: *Journal of Neurology* 55, S. 964–966.

Darley, John; Latane, Bibb, 1968: Bystander intervention in emergencies: Diffusion of responsibility. In: *Journal of Personality and Social Psychology* 8, S. 377–383.

Dutton, Donald; Aron, Arthur, 1974: Some evidence for heightened sexual attraction under conditions of high anxiety. In: *Journal of Personality and Social Psychology* 30, S. 510–517.

Dworkin, Gerald, 1988: *The theory and practice of autonomy*. Cambridge: Cambridge University Press.

Ekstrom, Laura, 2000: *Free will*. Boulder, CO: Westview Press.

Fischer, John, 1994: *The metaphysics of free will*. Oxford: Blackwell.

Fischer, John; Ravizza, Mark, 1998: *Responsibility and control*. Cambridge: Cambridge University Press.

Frankfurt, Harry, 1971: Freedom of the will and the concept of a person. In: *Journal of Philosophy* 68, S. 5–20.

Frankfurt, Harry, 1993: On the necessity of individuals. In: Noam, Gil; Wren, Thomas (Hrsg.): *The moral self*. Cambridge, MA: MIT Press, S. 16–27.

Geyer, Christian, 2004: Vorwort. In: Geyer, Christian (Hrsg.): *Hirnforschung und Willensfreiheit*. Frankfurt: Suhrkamp, S. 9–19.

Grün, Klaus-Jürgen, 2008: Glaubensfragen. In: Grün, Klaus-Jürgen; Friedman, Michel; Roth, Gerhard (Hrsg.): *Entmoralisierung des Rechts.* Göttingen: Vandenhoeck und Ruprecht, S. 11–53.

Grün, Klaus-Jürgen; Friedman, Michel; Roth, Gerhard, 2008: Vorwort. In: Grün, Klaus-Jürgen; Friedman, Michel; Roth, Gerhard (Hrsg.): *Entmoralisierung des Rechts*. Göttingen: Vandenhoeck und Ruprecht, S. 7–9.

Haggard, Patrick; Eimer, Martin, 1999: On the relation between brain potential and the awareness of voluntary movements. In: *Experimental Brain Research* 126, S. 128–133.

Halgren, Eric; Chauvel, Patrick, 1993: Experiential phenomena evoked by human brain electrical stimulation. In: Devinsky, Orrin; Barić, Aleksandar; Dogali, Michael (Hrsg.): *Electric and magnetic stimulation of the brain and spinal cord*. New York: Raven Press, S. 123–140.

Hofmann, Frank, 2008: Willensfreiheit und der Preis für den Kompatibilismus. In: Spät, Patrick (Hrsg.): *Zur Zukunft der Philosophie des Geistes*. Paderborn: mentis, S. 163–187.

Lampe, Ernst-Joachim; Pauen, Michael; Roth, Gerhard, 2008: Einlei-

tung. In: Lampe, Ernst-Joachim; Pauen, Michael; Roth, Gerhard (Hrsg.): *Willensfreiheit und rechtliche Ordnung*. Frankfurt: Suhrkamp, S. 9–37.

Latane, Bibb; Darley, John, 1970: *The unresponsive bystander*. New York, NY: Appleton-Century-Croft.

Libet, Benjamin, 1985: Unconscious cerebral initiative and the role of unconscious will in voluntary action. In: *Behavioral and Brain Sciences* 8, S. 529–566.

Libet, Benjamin, 2002: Do we have free will? In: Kane, Robert (Hrsg.): *The Oxford handbook of free will*. Oxford: Oxford University Press, S. 551–564.

Libet, Benjamin; Gleason, Curtis; Wright, Elwood; Pearl, Dennis, 1983: Time of conscious intention to act in relation to onset of cerebral activities (readiness-potential). In: *Brain* 106, S. 623–642.

Linser, Katrin; Goschke, Thomas, 2007: Unconscious modulation of the conscious experience of voluntary control. In: *Cognition* 104, S. 459–475.

Löffler, Winfried, 2007: What naturalists always knew about freedom. In: Gasser, Georg (Hrsg.): *How successful is naturalism*. Frankfurt: ontos, S. 283–300.

Lohmar, Achim, 2005: *Moralische Verantwortlichkeit ohne Willensfreiheit*. Frankfurt: Klostermann.

Merkel, Reinhard, 2008: Handlungsfreiheit, Willensfreiheit und strafrechtliche Schuld. In: Lampe, Ernst-Joachim; Pauen, Michael; Roth, Gerhard (Hrsg.): *Willensfreiheit und rechtliche Ordnung*. Frankfurt: Suhrkamp, S. 332–370.

Mohr, Georg, 2008: Welche Freiheit braucht das Strafrecht? In: Lampe, Ernst-Joachim; Pauen, Michael; Roth, Gerhard (Hrsg.): *Willensfreiheit und rechtliche Ordnung*. Frankfurt: Suhrkamp, S. 72–96.

O'Connor, Timothy, 2000: *Persons and causes*. Oxford: Oxford University Press.

Pauen, Michael, 2008: Freiheit, Schuld und Strafe. In: Lampe, Ernst-Joachim; Pauen, Michael; Roth, Gerhard (Hrsg.): *Willensfreiheit und rechtliche Ordnung*. Frankfurt: Suhrkamp, S. 41–71.

Pauen, Michael; Roth, Gerhard, 2008: *Freiheit, Schuld und Verantwortung*. Frankfurt: Suhrkamp.

Penfield, Wilder; Rasmussen, Theodore, 1950: *The cerebral cortex of man*. New York: Macmillan.

Pockett, Susan, 2006: The neuroscience of movement. In: Pockett, Susan; Banks, William; Gallagher, Shaun (Hrsg.): *Does consciousness cause behavior?* Cambridge, MA: MIT Press, S. 9–24.

Prinz, Wolfgang, 1996: Freiheit oder Wissenschaft? In: von Cranach, Mario; Foppa, Klaus (Hrsg.): *Freiheit des Entscheidens und Handelns, ein Problem der nomologischen Psychologie*. Heidelberg: Arsanger, S. 86–103.

Prinz, Wolfgang, 2004: Der Mensch ist nicht frei. In: Geyer, Christian (Hrsg.): *Hirnforschung und Willensfreiheit*. Frankfurt: Suhrkamp, S. 20–26.

Rösler, Frank, 2008: Was verraten die Libet-Experimente über den „freien Willen"? In: Lampe, Ernst-Joachim; Pauen, Michael; Roth, Gerhard (Hrsg.): *Willensfreiheit und rechtliche Ordnung*. Frankfurt: Suhrkamp, S. 140–164.

Roth, Gerhard, 2001: *Fühlen, Denken, Handeln*, 1. Aufl. Frankfurt: Suhrkamp.

Roth, Gerhard, 2003: *Fühlen, Denken, Handeln*, 2. vollst. überarb. Aufl. Frankfurt: Suhrkamp.

Roth, Gerhard, 2004: Worüber dürfen Hirnforscher reden – und in welcher Weise? In: Geyer, Christian (Hrsg.): *Hirnforschung und Willensfreiheit.* Frankfurt: Suhrkamp, S. 66–85.

Roth, Gerhard, 2006: Willensfreiheit und Schuldfähigkeit aus Sicht der Hirnforschung. In: Roth, Gerhard; Grün, Klaus-Jürgen (Hrsg.): *Das Gehirn und seine Freiheit*. Göttingen: Vandenhoek und Ruprecht, S. 9–27.

Merkel, Grischa; Roth, Gerhard, 2008: Freiheitsgefühl, Schuld und Strafe. In: Grün, Klaus-Jürgen; Friedman, Michel; Roth, Gerhard (Hrsg.): *Entmoralisierung des Rechts*. Göttingen: Vandenhoek und Ruprecht, S. 54–95.

Ravizza, Mark, 1994: Semi-compatibilism and the transfer of non-reponsibility. In: *Philosophical Studies* 75, S. 61–94.

Singer, Wolf, 2003: *Ein neues Menschenbild?* Frankfurt: Suhrkamp.

Singer, Wolf, 2004: Verschaltungen legen uns fest: Wir sollten aufhören, von Freiheit zu sprechen. In: Geyer, Christian (Hrsg.): *Hirnforschung und Willensfreiheit.* Frankfurt: Suhrkamp, S. 30–65.

Todorov, Alexander; Mandisodza, Anesu; Goren, Amir; Hall, Crystal, 2005: Inferences of competence from faces predict election outcomes. In: *Science* 308, S. 1623–1626.

Trevena, Judy; Miller, Jeff, 2002: Cortical movement preparation before and after a conscious decision to move. In: *Consciousness and Cognition* 11, S. 162–190.

Van Baaren, Rick; Holland, Rob; Steenaert, Bregje; van Knippenberg, Ad, 2003: Mimicry for money. In: *Journal of Experimental Social Psychology* 39, S. 393–398.

Walde, Bettina, 2006: *Willensfreiheit und Hirnforschung*. Paderborn: mentis.

Wallace, Jay, 1994: *Responsibility and the moral sentiments*. Cambridge, MA: Harvard University Press.

Walter, Sven, in Begutachtung: Freiheit und Kontrolle: Plädoyer für einen moderaten skeptischen Kompatibilismus.

Wegner, Daniel, 2002: *The illusion of conscious will*. Cambridge, MA: MIT Press.

Wegner, Daniel; Wheatley, Thalia, 1999: Apparent mental causation. In: *American Psychologist* 54, S. 480–492.

Wolf, Susan, 1990: *Freedom within reason*. Oxford: Oxford University Press.

Wuchterl, Kurt, 2007: *Die Sonderstellung des Menschen*. Hamburg: merus.

Maria E. Kronfeldner

Meme, Meme, Meme: Darwins Erben und die Kultur[1]

Zusammenfassung

Charles Darwin und seine Erben wendeten die Theorie der Evolution biologischer Arten auch auf Kultur an. Kultur evolviere wie die Natur auf Darwinistische Weise. Die sog. Memtheorie, vertreten von verschiedenen Autoren auf der Basis des Darwinistischen Genselektionismus, ist eine Spielart einer solchen analogen Anwendung. Dieser Artikel kritisiert drei zentrale Aussagen der Memtheorie: (i) dass es Einheiten der Kultur – Meme – gibt, die analog zu Genen zu verstehen sind, (ii) dass Meme, in Analogie zu Genen, Replikatoren sind, und (iii) dass Meme als Einheiten der kulturellen Selektion auf die gleiche Art wie Gene 'egoistisch' sein können. Nach einer Einführung in die Memtheorie in Teil 1, werden diese drei Thesen in Teil 2 als entweder falsch oder trivial entlarvt. Dieser kritische Teil soll vor allem zeigen, dass die Memtheorie keine 'gefährliche Idee' ist, die das bisher in den Geistes-, Kultur- und Sozialwissenschaften tradierte Verständnis von Geist und Kultur herausfordern kann. Im Gegenteil, im besten Fall re-formuliert die Memtheorie lediglich Bekanntes in evolutionärer Sprache und ist in diesem Sinne trivial. In Teil 3 wird die Perspektive gewechselt: Nicht mehr der Gehalt, sondern die Funktion der Memtheorie, v. a. im Kontext interdisziplinärer Verständigung, soll betrachtet werden. Denn trotz der Kritik der drei Kernthesen kann die Memtheorie eine kommunikative und somit produktive Rolle zwischen den 'zwei Kulturen' der Wissenschaften spielen.

Abstract

Charles Darwin and his heirs applied the theory of evolution not only to biological species but also to culture: as with nature, culture evolves in a Darwinian manner. The so-called meme theory (or memetics), defended by various authors on the basis of gene selectionism, is one such analogical application of Darwinism. This article criticizes three central claims of meme theory: (i) that there are memes, i. e. units of culture analogous to genes; (ii) that memes are, in analogy to genes, replicators; (iii) that memes, as units of cultural selection, are as 'egoistic' as genes are claimed to be by gene selectionists for the biological case. After an introduction to meme theory in part 1, the three theses are

debunked as either wrong or trivial. Part 2 will illustrate that meme theory is not a 'dangerous idea' that is able to challenge the received view of mind and culture in the humanities and social sciences. On the contrary, meme theory at best re-formulates in evolutionary language what we already know and is in this sense trivial. In part 3, the perspective of the analysis changes: not the content, but the function of memetics is at issue, i. e. the function of this theory for interdisciplinary communication. Despite the critique of the three central claims, memetics can have a communicative and therefore productive role to play between the 'two cultures' of science.

1. Die Memtheorie

Der Mensch als der erste Freigelassene der Schöpfung

Dass nicht alles angeboren ist, davon können wir ausgehen. Der Mensch ist in diesem Sinne der erste „Freigelassene der Schöpfung", wie Herder (1784, 146) sagte. Er transzendiert seine Natur durch Kultur, die er dank seines Geistes hervorbringt. Doch selbst wenn wir von dieser Freiheit des Menschen ausgehen, scheint die Darwinistische Evolutionstheorie den Menschen noch nicht aus ihren Fängen entlassen zu haben. Nach Daniel C. Dennett ist die These, dass alles angeboren und in diesem Sinne determiniert ist, bloßer „minimal Darwinism" (Dennett, 2000, ix). Radikaler sei die These, dass Kultur selbst ein Darwinistischer Prozess ist. Kultur wird im Rahmen einer solchen These nicht auf Natur reduziert, sondern es wird vielmehr die Ontologie und die Logik der Veränderung in Form von Analogien von Natur auf Kultur übertragen.

Derlei analoge Übertragungen gehören seit langem zur Geschichte des Darwinismus. Bereits William James (1880) hat den Darwinismus auf geistige und kulturelle Veränderung übertragen. Er tat dies in Auseinandersetzung mit den psychologischen und soziologischen Theorien Herbert Spencers, die er als zu Lamarckistisch ansah. Auch Donald T. Campbells evolutionäre Erkenntnistheorie beinhaltete eine Theorie der kulturellen Evolution (Campbell 1965). Eine neue populationsgenetische Ausrichtung fanden derlei Bestrebungen in den Koevolutionstheorien von Cavalli-Sforza & Feldman (1981), Boyd & Richerson (1985) und Durham (1991). Gegenstand dieser populationsgenetischen Ansätze ist nicht nur die Distribution von kulturellen Einheiten, sondern auch die phylogenetische Interaktion zwischen Kultur und Natur, z. B. wie die Verbreitung bestimmter landwirtschaftlicher Praktiken die Verbreitung

bestimmter Gene beeinflusste. Obwohl anthropologisch höchst interessant, hat diese Theorie in der Philosophie nicht viel Aufmerksamkeit auf sich gezogen. Für Kontroversen hat hingegen in den letzten dreißig Jahren eine andere Spielart der analogen Übertragung gesorgt, nämlich die sog. Memtheorie. Um diesen Ansatz soll es im Folgenden gehen.[2]

Die drei Thesen der Memtheorie

Richard Dawkins hat die Memtheorie in seinem Buch *The Selfish Gene* (1976) eingeführt. Die Natur hat Gene, die Kultur Meme. Beides sind egoistische Replikatoren. Der Ausdruck 'Mem' leitet sich von 'memory' bzw. 'mimesis' ab. Richard Semon, deutscher Biologe und Schüler von Ernst Haeckel, gebrauchte in seiner monistischen Theorie des Zellgedächtnisses einen ähnlichen Terminus. Er sprach von 'Mnemen' (Semon 1904). Semon ist aber *kein* Vorläufer der Memtheorie, da er eine analoge Übertragung im direkt entgegengesetzten Sinne intendierte, weswegen er auch häufig als Lamarckist eingestuft wurde: Wie der Geist haben auch Zellen ein Gedächtnis, so seine These. Ob Lamarckist oder nicht, sei dahin gestellt. Fest steht, dass bei Semon keine Spur ist von einem genselektionistischen Darwinismus, wie er durch Dawkins seit den 1970ern populär geworden ist. Da dieser Genselektionismus, den ich im Verlauf der Argumentation erläutern werde, das Fundament der Memtheorie bildet, sollte man vorsichtig sein, frühere Ansätze mit der Memtheorie zu vergleichen, egal ob James', Semons oder Boyds und Richersons Ansatz. Die Memtheorie ist spezifisch, weil sie eine spezifische Version des Darwinismus voraussetzt.

Die Memtheorie wird trotz einiger Todesmeldungen immer noch kontrovers debattiert.[3] Einerseits wird davon gesprochen, dass diese Modetheorie der Meme einer wissenschaftlichen Betrachtung unwürdig ist; andererseits wird Dawkins' Memtheorie als Entdeckung der DNA der Kultur interpretiert und zur Erklärung für kulturelle Transmissions- bzw. Diffusionsprozesse, zur Erklärung von Irrationalität oder gar zur Erklärung des menschlichen Geistes herangezogen. Hauptvertreter sind neben Dawkins: David Hull (1982, 2000), Daniel Dennett (1991, 1995, 2001), Susan Blackmore (1999, 2000, 2010) und Robert Aunger (2002), obgleich jeder dieser Autoren unterschiedliche Thesen mit dem Begriff der Meme verbindet.

Da die Details dieser Unterschiede den Rahmen eines Aufsatzes sprengen würden, werde ich mich auf drei spezifische, gehäuft auftretende

oder provokante Thesen im Kern der Memtheorie konzentrieren: (i) dass es analog zu Genen so etwas wie Meme gibt, (ii) dass beide Replikatoren sind, und (iii) dass Gene und Meme 'egoistisch' sind.

(i) Meme

Anfang des 20. Jahrhundert waren Gene nur hypothetische Bausteine der Vererbung. Erst später wurden sie mit DNA-Molekülen identifiziert. Damit konnten zwar längst nicht alle ontologischen Fragen der belebten Natur geklärt werden, aber eine ontologische Frage wurde damit eindeutig geklärt: die materielle Natur der Gene. Dass genetische Einheiten aus DNA-Sequenzen bestehen, ist heutzutage unstrittig.[4]

Wie sieht nun die Ontologie der Kultur aus? Die Memetik scheint hier ein spezifisches Angebot zu machen: Kultur besteht aus Memen. Doch selbst Memetiker sind sich bis heute uneinig darüber, was Meme genau sind. Ohne auf die Details des oft unklaren bis widersprüchlichen Gestrüpps von Äußerungen der Memetiker einzugehen, können folgende Punkte festgehalten werden.[5] Erstens: meist werden Meme als ideelle Einheiten betrachtet. Beispiele sind Ideen, Theorien, Werte, Überzeugungen, kognitiv repräsentierte Verhaltensmuster und Herstellungsregeln für Artefakte (z. B. ein Kuchenrezept, eine Partitur oder die Idee des Rades). Dieser Begriff von Memen als ideelle Einheiten findet sich sowohl bei Dawkins, Dennett, Blackmore, als auch bei Hull.[6] Gehirnzustände sind dann, zweitens, eine bloße *physikalische Realisierung* der Meme, analog zur DNA bei Genen. Dennett, Hull und Blackmore gehen zudem davon aus, dass Meme auch andere physikalische Realisierungen haben können. Meme existieren somit auch in Artefakten, wie Büchern und dergleichen. Das Rad-mem kann in einem Gehirn genauso wie in einem Rad realisiert sein. Gene hingegen sind *nur* in DNA physikalisch realisiert. Wie Gene haben Meme, drittens, aber auch phänotypische Eigenschaften. Meistens sind dies Artefakte, manchmal ist es aber auch der menschliche Geist, der als Phänotyp des Memotyps gilt. Unterschiede zwischen den Autoren gibt es v.a. in Bezug auf die Frage, ob Gehirnzustände als materielles Substrat Exklusivität beanspruchen können und in Bezug auf die Frage, was als Phänotyp eines Memes anzusehen ist.

(ii) Replikation

Die Memtheorie (oft auch als *Memetik* bezeichnet) macht aber nicht einfach ontologische Anleihen bei der *Genetik*. Sie leitet sich vielmehr aus

dem Genselektionismus, einer ganz bestimmten Version des Darwinismus ab. Der Genselektionismus begreift Gene als egoistische Replikatoren der biologischen Evolution und somit als die eigentlichen Einheiten der Selektion. Analoges gilt dann in Hinblick auf kulturelle Veränderung für die Meme.

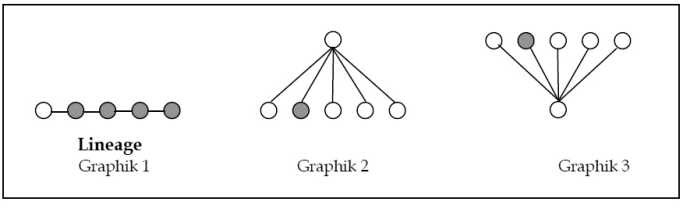
Genselektionismus widerspricht dem Individualselektionismus. In Darwins *Origin of Species* (1859) standen Individuen bzw. Organismen im Zentrum der Betrachtung. Sie waren die *Einheiten der Selektion*: Es sind Typen von Individuen, die dank adaptiver Eigenschaften in einer Population an Zahl zunehmen, also eine höhere Fitness aufweisen. Natürliche Selektion selektiert Individuen und sie befördert damit auch den Nutzen dieser Individuen. Adaptive Eigenschaften, die das Überleben und die Fortpflanzung dieser Individuen fördern, evolvieren. Kurzum, Selektion selektiert Individuen (selection *of*) und sie tut dies wegen der adaptiven Eigenschaften dieser Individuen (selection *for*).[7]

Nach Dawkins' Genselektionismus können hingegen nur Gene Einheiten der Selektion sein, denn nur Gene überleben in Form von echten Kopien über mehrere Generationen. Da die Gene die Einheiten des 'Überlebens' sind, fördert die Selektion somit notwendigerweise den Nutzen dieser Gene und nicht (bzw. nur kontingenterweise) den Nutzen der sich nicht replizierenden Individuen. Darüber hinaus sind nach Dawkins Organismen, und somit die adaptiven Eigenschaften dieser Organismen, im Wesentlichen bloße *Effekte* von Genen. Aus all dem folgt, dass jene Gene evolvieren, die 'egoistisch' sind, d.h. Eigenschaften aufweisen, die ihre Fitness, ihre Überlebenschancen, erhöhen, selbst wenn dies auf Kosten der einzelnen Organismen gehen sollte.[8]

Diese genselektionistische Konzeption der Evolution setzt einen ganz bestimmten Replikationsbegriff voraus. Ich werde hier nur auf die Aspekte des Replikationsbegriffs eingehen, die für die Beurteilung der drei Thesen der Memtheorie zentral sind.[9] Die erste und wichtigste Bedingung, die eine Entität erfüllen muss, um als Replikator zu gelten, ist eine *Ähnlichkeitsbedingung*. Replikatoren müssen Kopien von sich herstellen können und in dieser Form über lange Zeit hinweg existieren. Ein Original und eine Kopie müssen dabei hinreichend ähnlich sein, um als Original bzw. Kopie zu zählen.[10] Nach Dawkins erfüllen Gene diese Ähnlichkeitsbedingung, Organismen tun dies nicht. Der Grund, warum Organismen nicht als Replikator gezählt werden können, liegt nach Dawkins in der Unmöglichkeit der Vererbung erwor-

bener Eigenschaften und nicht in der sexuellen Fortpflanzung mancher Arten.[11]

Eine zweite Bedingung, die ich *Lineagebedingung* nennen möchte, betrifft nicht das Ähnlichkeitsverhältnis, sondern das kausale Verhältnis zwischen Original und Kopien. Sie besagt, dass ein Gen *von einem anderen* kopiert sein muss, um als Replikator zu gelten. Die Graphik 1 der Abb. 1 stellt einen solchen Vorgang bildlich dar.



Wenn die Kopien alle von *einem* Original stammen, wie in der Graphik 2, oder wenn eine einzelne Kopie aus mehreren Originalen 'zusammenkopiert' wird, wie in der Graphik 3, dann handelt es sich nicht um einen Replikationsprozess im engen Sinn. Ob es sich um Replikation im engen Sinne handelt, ist jedoch – so Dawkins – zentral für die Frage, ob ein System das Potenzial zur Evolution hat: nur dann, wenn eine Entität eine Lineage bildet, d. h. wenn eine Kopie die nächste hervorbringt, kann sich eine 'Mutation' in die nächste Generation von Kopie 'fortpflanzen'. Deswegen kann sich die Mutation von 'weiß' zu 'grau' in der Abb. 1. auch nur im Falle der Lineage (Graphik 1) 'fortpflanzen'.[12]

(iii) Egoismus

Aufregung erzeugte v. a. eine dritte These Dawkins': dass Replikatoren, Gene wie Meme, 'egoistisch' seien. Diese These zerfällt in zwei Behauptungen: In der biologischen wie kulturellen Evolution gehe es, erstens, allein um die Fitness und somit um das Überleben der Replikatoren. Es sei also das metaphorisch zugewiesene 'Eigeninteresse am Überleben' der Gene und Meme, das in allen Fällen durch die Selektion (natürlicher wie kultureller) belohnt wird. Hinzu kommt, zweitens, die These, dass der Geist lediglich ein phänotypischer Effekt von Memen ist, so wie der Körper im Genselektionismus als bloßer phänotypischer Effekt von Genen verstanden wird. In Dawkins' Sprache: Körper und Geist sind 'Überlebensmaschinen', von Genen und Memen, ontogenetisch 'gebaut' und phylogenetisch evolviert, um das Überleben der egoistischen Repli-

katoren zu befördern.[13] Was wir denken, *nützt nicht uns*, sondern den Memen, und eigentlich, auf einer anderen Ebene der Betrachtung, *sind* wir, Subjekte des Denkens, nichts anderes als ein Bündel *aus* Memen. Eigentlich gibt es uns gar nicht, sondern nur Gene und Meme bzw. es gibt uns nicht in der Form, wie wir bisher dachten. Das Subjekt des Denkens ist lediglich eine *Illusion*, die sich diese Replikatoren machen, um ihr Überleben und ihre Reproduktion vorantreiben zu können. Diese Zuspitzung wurde v. a. von Daniel C. Dennett[14] und Susan Blackmore[15] verteidigt.

Die postulierte Radikalität und die Architektur einer kritischen Beurteilung

Die angebliche Radikalität dieses 'Memselektionismus' liegt in dem scheinbar enthaltenen Angriff auf unser traditionelles Verständnis vom Menschen als Schöpfer der Kultur. Die drei Thesen zusammengenommen scheinen zu einer Art 'Umkehrung der Werte' zu führen. Die Memtheorie stellt das Verhältnis zwischen Menschen und den von ihnen hervorgebrachten kulturellen Einheiten auf den Kopf. Was Geschaffenes war (kulturelle Einheiten), wird zum Akteur (egoistische Replikatoren) und der Schöpfer (das Subjekt) wird zum Geschaffenen der vormaligen Geschöpfe (Überlebensmaschine).

Im Folgenden möchte ich zeigen, dass sich diese Radikalität auflöst, wenn wir die drei zentralen Thesen der Memtheorie kritisch prüfen. Die drei Thesen waren: dass es besagte Meme gibt, dass sie wie Gene Replikatoren sind, und dass sie im genselektionistischen Sinne egoistisch sind. Ich werde zeigen, dass die drei Behauptungen entweder falsch oder trivial sind. Aussagen werden als *falsch* beurteilt, wenn Ähnlichkeiten postuliert wurden, wo keine sind; sie werden als *trivial* eingestuft, wenn lediglich das allseits Bekannte in neuer, evolutionistischer Sprache re-formuliert wird. Dabei unterscheide ich zwischen explanatorischer und heuristischer Trivialität: Als *explanatorisch trivial* beurteile ich Aussagen, entweder wenn sie *keine* Erklärung beinhalten, z. B. weil sie *tautologisch* sind, oder wenn sie keine *adäquate* Erklärung auf dem Niveau vergleichbarer Ansätze anbieten, d. h. wenn sie nicht auf einer bestimmten Erklärungsebene als Alternative zu vergleichbaren Ansätzen auftreten. *Heuristisch trivial* sind Aussagen hingegen, wenn sie auf der Ebene vergleichbarer Ansätze angeboten werden, und somit einen gewissen explanatorischen Gehalt für sich reklamieren können, aber im Vergleich zu diesen alterna-

tiven Ansätzen keine *neuen* Beschreibungen oder Erklärungen beisteuern. Heuristische Trivialität ist eine begrenztere Art der Trivialität.

Die Frage nach der Trivialität ist wichtig. Wenn der Fokus auf der Frage liegt, ob die behaupteten Ähnlichkeiten faktisch bestehen, wird das Erklärungspotential und der Neuigkeitsgehalt der Memtheorie oft gar nicht in Frage gestellt. Eine solche einseitige Behandlung ignoriert aber, dass eine triviale Theorie fast so unbrauchbar ist wie eine falsche und überdies irreführend sein kann, z.B. wenn ein *wissenschaftlicher Imperialismus* im Spiel ist, im Rahmen dessen sich die Naturwissenschaften eine Dominanzstellung bzw. Deutungsmacht gegenüber den Geistes- und Sozialwissenschaften sichern, und sei es, indem die Naturwissenschaften den letzteren großzügig Theoreme zur analogen Übertragung 'überlassen'. Erst durch eine erweiterte Architektur der Kritik, gleichzeitig Falschheit und Trivialität prüfend, kann eine differenzierte Beurteilung gewährleistet werden, eine Beurteilung, die sowohl Gehalt als auch Funktion der Memtheorie betrachtet. Während sich Teil 2 auf den Gehalt konzentriert, bietet Teil 3 eine zusammenfassende Einbettung der Memtheorie anhand einer detaillierten Betrachtung der *Funktion*, welche die Memtheorie in der disziplinären Wissenschaftslandschaft übernehmen kann.

2. Eine Kritik der drei Kernthesen der Memtheorie

(i) Gibt es Meme?

Das oben angesprochene postulierte Verhältnis zwischen Memen und Gehirnzuständen führt zu einem Problem, das ich *Problem der materiellen Identifikation* nennen möchte. Selbst wenn es Meme gibt, dann sind sie schwer zu identifizieren. In der derzeitigen Philosophie des Geistes, die sich mit dem Verhältnis zwischen Gehirn und Geist beschäftigt, findet sich nur selten ein Konsens, aber das Folgende kann als solcher gezählt werden. Wenn wir neuronale Entsprechungen für mentale Einheiten (z.B. für den Gedanken „Der Apfel ist rot") annehmen, dann wird das Gehirn einer Person in einem ganz bestimmten Zustand sein, wenn die entsprechende Person diesen Gedanken hat. Trotzdem gilt, dass, wenn dieselbe Person denselben Gedanken irgendwann später noch einmal denkt oder eine andere Person denselben Gedanken denkt, der Gehirnzustand so gut wie sicher nicht der gleiche sein wird. Zwei Subjekte können zwar denselben Gedanken 'teilen', oder als ein Subjekt ein

und denselben Gedanken zweimal haben, aber der damit korrelierende Gehirnzustand verändert sich ständig. Es gibt keine *kontextunabhängige* Entsprechung zwischen Gehirnmustern und mentalen Gehalten. Gedanken sind in diesem Sinne auf der neuronalen Ebene multipel realisierbar. Da jedes Mem für den menschlichen Geist ein solcher mentaler Gehalt ist, gilt: Ein und dasselbe Mem kann neuronal unterschiedlichst 'codiert' sein. Das macht die Identifikation von Memen problematisch: Da es keinen universellen *Gehirncode* von Memen gibt, der es uns erlaubt, eindeutig festzustellen, ob ein bestimmtes Mem vorliegt oder nicht, bleibt nur das Verhalten als *Indikator* für bestimmte Meme. Man muss aus den beobachtbaren Einheiten (Verhalten, Artefakte) mittels Interpretation auf die ideellen Einheiten schließen.

Die Identifikation von Memen unterscheidet sich somit stark von der materiellen Identifikation von Genen. Bei Genen kann man eindeutig herausfinden, ob ein bestimmtes Gen vorhanden ist, da Gene ein eindeutiges materielles Substrat haben: Man überprüft schlicht, ob die entsprechende DNA-Sequenz vorliegt oder nicht. Wenn es einen Unterschied in einem Gen gibt, dann gibt es auch einen Unterschied in der DNA und *vice versa*, d.h. wenn es einen Unterschied in der DNA gibt, dann gibt es einen Unterschied im Gen. Nicht so bei den Memen. In Bezug auf das Problem der materiellen Identifikation schneiden Gene und Meme also unterschiedlich ab.[16]

Gegner wie Befürworter der Memtheorie stimmen in der Regel zu, dass Meme aufgrund der multiplen Realisierbarkeit mentaler Gehalte keine vergleichbar eindeutige und direkte materielle Identifikation wie Gene erlauben. Die beiden zu nennenden negativen Konsequenzen für den Status der Memtheorie werden hingegen von den Befürwortern überwiegend ignoriert.[17]

Erstens ist die Analogie zwischen Memen und Genen in Bezug auf die Möglichkeiten der materiellen Identifikation eindeutig *falsch*: Gene erlauben eine eindeutige materielle Identifikation, Meme tun das nicht. Dies gilt, selbst wenn bis in die Mitte des 20. Jahrhunderts die materielle Identifikation der Gene ebenso prekär war. Anfang des 20. Jahrhunderts hoffte man zumindest, dass man das materielle Fundament der Gene finden wird und diese Hoffnung wurde nicht enttäuscht. In Bezug auf Meme hingegen hofft m.W. niemand, dass man den Code der 'Mem-DNA' finden wird.

Zweitens kann der interpretative Schluss von Verhalten auf Meme

mehr oder weniger prekär sein, ein Problem, mit dem sich die Geistes- und Sozialwissenschaften seit geraumer Zeit auseinanderzusetzen haben. Dies bedeutet aber auch, dass mit dem Problem der materiellen Identifikation jegliche Hoffnung obsolet wird, dass die Memetik eines Tages eine Wissenschaft wie die Genetik werden könne. Wenn nur die Möglichkeit der *Identifikation durch Interpretation* gegeben ist, muss die Memetik sich einreihen in den Reigen der interpretativen Sozial- und Geisteswissenschaften.[18] Dies ist an sich nicht schlimmer, als es immer schon für die betroffenen Geistes- und Sozialwissenschaften war, führt die Memetik aber in die heuristische Trivialität. Lassen Sie mich diesen Vorwurf etwas detaillierter ausführen: Auf das Problem der materiellen Identifikation und die damit verbundene Disanalogie kann ein Memetiker antworten, dass Analogien keine Ähnlichkeiten in jeder Hinsicht behaupten müssen. Solange die behaupteten Ähnlichkeiten zu interessanten Einsichten führen, ist die Analogie trotzdem gerechtfertigt. Aber führt die These, dass es Meme gibt, zu solchen interessanten Einsichten? Ich denke nicht. Die These, dass es Meme gibt, ist *heuristisch trivial*, zumindest in der Form, in der sie verteidigt werden kann, d.h. unter Berücksichtigung des Problems der materiellen Identifikation. Wie alle Geistes-, Sozial- und Kulturwissenschaften geht die Memtheorie von bestimmten unbeobachtbaren, ideellen Einheiten der Kultur aus. Damit holt sich die Theorie auch die üblichen ontologischen Probleme ins Haus, die sich aus der Annahme solcher Entitäten ergeben. Bis heute ist nicht klar, welchen ontologischen Status mentale bzw. ideelle Einheiten genau haben. Trotz der Anleihen aus einer naturwissenschaftlichen Theorie hat die Memtheorie aber bisher nichts Neues zur Beseitigung der ontologischen Unwägbarkeiten des Postulats ideeller Einheiten angeboten, da sie keine Ressourcen enthält, welche die Grundkoordinaten des Problems der materiellen Identifikation verändern, geschweige denn auflösen würden.

Die These, dass es besagte Meme gibt, re-formuliert schlicht ein altes Postulat der Geistes- und Sozialwissenschaften und trägt nichts Neues dazu bei, Probleme zu lösen, die mit diesem Postulat erst geschaffen werden.

(ii) Sind Meme Replikatoren?

Wenn wir der Memtheorie dieses Problem und die damit verbundene Trivialität nachsehen, kann die Theorie dann zumindest für sich beanspruchen, dass Meme Replikatoren sind?

Der Schluss von gleichem Verhalten auf gleiche Meme ist nicht immer gewährleistet. Wenn ein Kind etwas von einer Elterngeneration lernt (z.B. die Aussprache bestimmter Phoneme) und somit das gruppentypische Verhalten zeigt, kann es trotzdem eine völlig andere Repräsentation dieses Verhaltens entwickeln, wie Boyd & Richerson eingewendet haben.[19] In den Fällen, in denen ein Schluss von gleichem Verhalten zu gleicher kognitiver Repräsentation prekär ist, ist es schwer zu sagen, ob sich ein Mem repliziert hat oder nicht, selbst wenn Menschen gleiches Verhalten an den Tag legen. Dies ist nicht der entscheidende Punkt. Er zeigt lediglich die Bedeutung des Problems der materiellen Identifikation für die zweite These. Entscheidend für die zweite These der Memtheorie ist hingegen Folgendes. Soziales Lernen basiert, wie Boyd & Richerson ebenso festhalten, meist auf dem mehrmaligen Kontakt mit der zu lernenden kulturellen Einheit und ein Kind übernimmt dann in der Regel die häufigste Version einer kulturellen Einheit (im Beispiel: eine bestimmte phonetische Aussprache). Lernen beinhaltet somit keine eins-zu-eins-Relation zwischen einem Original und einer Kopie und erlaubt somit auch keine echte Lineagebildung. Dies ist aber, wie oben ausgeführt, eine zentrale Bedingung für Replikation im engen Sinn und wird von Genen auch erfüllt. In Bezug auf diesen Aspekt ist die Analogie zwischen Genen und Memen also ebenso *falsch*, zumindest für jene Meme, die keine eins-zu-eins Relation zwischen einem Original und einer Kopie aufweisen.

Verteidiger der Analogie können auf Kritik dieser Art mit einem Rückzug auf einen weiten Replikationsbegriff reagieren: Alles was gewährleistet sein muss, um von Replikation von Memen zu sprechen, ist, dass '*etwas*' von Mensch zu Mensch durch soziales Lernen übertragen wird.[20] Dieses Etwas sei das Mem. Dass dies eine unbefriedigende Antwort ist, ist unmittelbar einsichtig, denn es führt die Analogie direkt in die Trivialität: Dass bei sozialem Lernen 'etwas' von Mensch zu Mensch übertragen wird, ist keine neue Einsicht und gleichsam in der Definition von sozialem Lernen enthalten. Die These, dass Meme Replikatoren im weiten Sinne sind, ist somit *heuristisch trivial*: Der heute in der Anthropologie und den Sozialwissenschaften übliche Kulturbegriff geht davon aus, dass der Kern der Kultur aus Ideen und kognitiven Repräsentationen von Verhaltensweisen und Praktiken besteht, die individuell oder sozial gelernt werden und nicht angeboren sind. Die Analogie zu Genen ist lediglich eine Neuformulierung dieser altbekannten definitorischen Annahme in Darwinistischer Sprache: Ähnlichkeiten zwischen Menschen erklären

sich entweder über Natur oder über Kultur und letzteres beruht auf kultureller Vererbung, bzw. sozialem Lernen. *Explanatorisch trivial* wird die Analogie durch dieses Rückzugsgefecht, da mit der These, dass in der Kultur etwas von einer Person zur nächsten übertragen wird, keine Alternative zu den Theorien des sozialen Lernens, die in der Psychologie entwickelt wurden, angeboten wird. Im Gegensatz zur Memetik versuchen diese Ansätze, die Unterschiede zwischen verschiedenen Formen des sozialen Lernens zu verstehen und die beteiligten kognitiven Prozesse im Detail zu erforschen.[21] Wenn soziales Lernen erklärt werden kann, dann mit Hilfe dieser sozialpsychologischen Theorien. Die These, dass Meme Replikatoren im weiten Sinne sind, ist keine Alternative zu diesen Theorien, denn sie operiert nicht auf dem gleichen Erklärungsniveau und ist somit nicht nur heuristisch, sondern auch explanatorisch trivial.

(iii) Können Meme 'egoistisch' sein?

Die dritte der oben eingeführten Thesen besagt, dass Meme 'egoistisch' seien: Sie verbreiten sich angeblich allein aufgrund ihrer eigenen Fitness, unabhängig von den Eigenschaften und Interessen ihrer 'Träger' und deren 'Fitness', d.h. unabhängig von unseren Eigenschaften, Überzeugungen und Präferenzen. Meme müssen nicht auf unsere 'Fitness' Rücksicht nehmen. Die Fitness der Meme sei somit entscheidend, um den Verlauf der biologischen und kulturellen Evolution zu erklären, und nicht unsere Fitness bzw. unser Nutzen. Was wir denken, 'nützt' eigentlich nicht uns, wie auch immer dieser Nutzen definiert ist, sondern den Genen und Memen, d.h. ihrem Überleben. „[A] cultural trait may have evolved in the way it has simply because it is *advantageous to itself*".[22] Diese Egoismusthese scheint unserer traditionellen Erklärung von kultureller Selektion zu widersprechen, denn letztere erklärt die Verbreitung kultureller Einheiten über Eigenschaften, Interessen und Entscheidungen von Individuen, mögen diese rational sein oder nicht.

Wenn wir die Egoismusthese so verstehen, dass sie Kultur auf ein 'survival of the fittest meme' reduziert, dann zeigt sich ganz deutlich, dass die Fitness der Meme *nicht* unabhängig vom menschlichen Geist (mit seinen Eigenschaften, Fähigkeiten, Interessen und Überzeugungen) *sein kann*, da der menschliche Geist die Umwelt der Meme bildet und die Umwelt eines Replikators die Fitness eines Replikators bestimmt. Die These, dass sich ein Mem unabhängig vom Geist des Menschen verbreiten kann, ist so sinnlos wie die These, dass sich Gene unabhängig

von *ihrer* selektiven Umwelt verbreiten können. Sie ist sinnlos, da sie zu einer Tautologie führt, die sich speziell auf 'survival of the fittest x' Erklärungen bezieht. Wenn die Evolutionstheorie die Existenz, d.h. das Überleben bestimmter Typen von Organismen über Fitnessunterschiede, über das Prinzip 'survival of the fittest', erklärt und Fitness dabei als hohe Überlebenswahrscheinlichkeit definiert, dann läuft die Erklärung auf ein 'Überleben der mit höchster Wahrscheinlichkeit Überlebenden' hinaus. Um dieser Tautologie zu begegnen, muss Fitness, eine rein quantitative Eigenschaft, selbst erklärbar sein und entsprechend definiert werden, indem auf *Angepasstheit* der Organismen verwiesen wird und nicht nur auf die Überlebenswahrscheinlichkeit. Angepasstheit, eine qualitative Eigenschaft, ergibt sich aus der Relation der entsprechenden Einheit (Gene, Organismus, Meme, …) zur selektiven Umwelt. Die Angepasstheit der sprichwörtlichen Giraffe mit dem langen Hals ergibt sich aus dem Verhältnis der Giraffe zur Umwelt mit hohen Bäumen und dies, und nur dies, begründet ihre Fitness, d.h. ihre Überlebenswahrscheinlichkeit. Wird der Bezug zur selektiven Umwelt ignoriert, dann wird die These, dass Giraffen mit langen Hälsen überlebten, weil sie eine hohe Fitness hatten, tautologisch. Das gleiche gilt für Meme: 'Survival of the fittest meme' erklärt schlicht gar nichts, solange nicht auf die *Relation zwischen den Eigenschaften der Meme und den Eigenschaften seiner Umwelt verwiesen* wird. Diese Relation erst erklärt, wieso bestimmte Meme eine höhere Fitness aufweisen als andere. Ein Beispiel: Dass das Mem '39ioghdöaghdlokreörk' eine sehr geringe Fitness hat, liegt an der Relation zu bestimmten Eigenschaften unseres Geistes. In einer silikonbasierten Computerumwelt oder im Gehirn von Marsmenschen mit einer ganz anderen Struktur des Geistes könnte dieses Mem durchaus eine sehr große Fitness haben, in der Umwelt des menschlichen Geistes aber geht die Fitness dieses Mems gegen Null. In genau diesem Sinne kann kein Mem unabhängig vom menschlichen Geist sein.[23]

Obwohl diese Abhängigkeit von den Verteidigern der These, dass Meme egoistische Replikatoren sind, bisweilen zugegeben wird, sehen diese nicht, dass sie damit in einem *explanatorischen Dilemma* gefangen sind: Wenn die Relation zwischen Memen und dem Geist als selektiver Umwelt berücksichtigt wird, dann enthält die These gar keinen Widerspruch zur traditionellen Erklärung, denn diese besagt ja nichts weiter, als dass Menschen Meme (ideelle Grundbausteine von Kultur) selektieren. Die Egoismusthese präsentiert diese traditionelle Auffassung

lediglich trickreich in der Terminologie des Genselektionismus, d.h. in neuer Verpackung. Die Egoismusthese wird demnach *heuristisch trivial*. Wenn andererseits die Relation des Mems zum menschlichen Geist nicht berücksichtigt wird, dann wird die Egoismusthese *explanatorisch trivial*, weil sie tautologisch wird.

Sterelny (2006) versuchte gegen diesen Einwand die explanatorische Kraft der Meme zu rehabilitieren. Ein Vergleich zur Nature/Nurture-Debatte kann sein Argument verdeutlichen: Obwohl immer Gene *und* Umwelt herangezogen werden müssen, um die Entwicklung eines phänotypischen Merkmals zu erklären, gibt es phänotypische Merkmale, deren Ausprägung variiert, wenn die Umwelt sich ändert und es gibt solche, deren Ausprägung (für eine gegebene Variation der Umwelt) dies nicht tut. Letztere haben eine sog. 'flache' Reaktionsnorm und werden oft als genetisch determiniert bezeichnet.[24] In ähnlicher Weise versucht Sterelny zu zeigen, dass, obwohl *immer* Meme und der Geist herangezogen werden müssen, um die Verbreitung von Memen zu erklären, es Meme gibt, deren Verbreitungswahrscheinlichkeit stark von idiosynkratischen Eigenschaften des menschlichen Geistes abhängen und solche die das weniger tun.[25] Inwiefern einzelne Meme in diesem Sinne abhängig sind und andere nicht, ist eine empirische Frage, die in diesem Aufsatz nicht gelöst werden kann. Festzuhalten bleibt jedoch, dass dieses Argument keinen Ausweg aus dem oben beschriebenen Dilemma bildet, denn die prinzipielle Abhängigkeit von einer Umwelt wird dadurch nicht aufgelöst.

Der Grund, wieso die Abhängigkeit im Fall der Meme oft übersehen wird, liegt in der falschen bzw. einseitigen Ausbuchstabierung der Analogie: Meme können sich nicht auf die gleiche Art und Weise wie Gene unabhängig vom Organismus (bzw. dem Geist) verbreiten, weil der Organismus im Falle der Meme die Rolle der Umwelt, und nicht nur (bzw. wenn überhaupt) die Rolle des Phänotyps der Replikatoren übernimmt. Wenn überhaupt, ist der Geist Umwelt und Phänotyp zugleich und der Evolutionsprozess ein Prozess der Autoselektion: Der Geist selektiert sich selbst und seine Inhalte. Dies ist eine Idee, die weit davon entfernt ist, unsere moderne Vorstellung vom Menschen als Schöpfer seiner Kultur zu destruieren.

Eine weitere Antwort auf das aufgezeigte explanatorische Dilemma finden wir in einer Art Rückzugsgefecht: Die Egoismusthese eigne sich besonders für Fälle von Irrationalität bzw. kann – im Gegensatz zur tra-

ditionellen Perspektive – diese Fälle erklären, d.h. Fälle, in denen wir etwas tun oder denken, dass wir eigentlich nicht tun oder denken wollen.[26] Dies sind Fälle, wo die Verbreitung einer kulturellen Einheit, eines Mems, durch eine Person bzw. einer Personengruppe mit keinem Nutzen für die Person bzw. die Gruppe korrespondiert. Meme seien Parasiten. Die Verbreitung eines Kettenbriefs oder Suchtverhalten sind einfache und gern verwendete Beispiele für in diesem Sinne 'egoistische' Meme. Solche Meme werden von den Memetikern gerne als 'virus of the mind' bezeichnet.[27] Wir tun etwas, selektieren eine Idee, auch wenn wir dies gar nicht wirklich wollen. Doch auch in solchen Fällen von Irrationalität können und müssen wir, erstens, auf Eigenschaften und Interessen von Individuen verweisen, wenn auch auf irrationale oder verdrängte. Sonst würde die 'survival of the fittest meme' Erklärung wiederum tautologisch werden. Da die traditionelle Perspektive zweitens sehr wohl solche irrationalen 'Gründe' berücksichtigen kann, bietet die Egoismusanalogie durch den Rückzug auf Irrationalität auch hier keine neuen Erkenntnisse.

Es gibt ein drittes Argument zur Abwendung des explanatorischen Dilemmas, dem ich bisher noch nichts entgegen gehalten habe. Das Argument besagt, dass der menschliche Geist nichts weiter sei als ein phänotypischer Effekt *von* Memen bzw. ein Bündel *aus* Memen.[28] Die Memetik müsste dann nicht als neue Erklärung für die Verbreitungsmuster von kulturellen Einheiten, noch als Erklärung für irrationales Verhalten bewertet werden, sondern als eine These über die Natur des menschlichen Geistes. Was die Natur des Geistes ist, und ob wir den Geist auf seine *Inhalte* reduzieren können, sind alte und ehrwürdige Kernfragen der Philosophie, die ich hier nicht versuchen werde zu beantworten. Ich möchte aber zeigen, inwiefern diese Zuspitzung der Memtheorie problematisch ist, und inwiefern sie keinen Ausweg aus dem oben beschriebenen explanatorischen Dilemma bereitstellen kann.

Aus nicht-naturalistischer Perspektive gibt es begründete Argumente für die Behauptung, dass die Existenz von Ideen, kognitiven Repräsentationen etc. ein Subjekt voraussetzt.[29] Mit anderen Worten, ein Jemand muss Ideen denken. Ein solches Subjekt ist aber selbst kein Mem. Ob es ein Subjekt dieser Art gibt, möchte ich hier nicht entscheiden. Ich möchte vielmehr folgenden Punkt in Bezug auf den Subjekteinwand vertreten: Das Subjekt wird nicht erst seit dem Darwinismus als dubiose Entität kritisiert, und die Bündeltheorie des Geistes, die auf ein solches Subjekt

des Denkens verzichtet, fand lange vor der Memtheorie Unterstützung.[30] Dennett selbst vertrat eine solche Theorie anfangs ohne jegliche Referenz auf egoistische Meme.[31] Solange nicht gezeigt werden kann, dass die Memtheorie ein *neues* Argument für die Bündeltheorie eröffnet, und meines Wissens wurde dies bisher noch nicht gezeigt, ist der Verweis auf Meme im Kontext der Philosophie des Geistes *heuristisch trivial*.

Unabhängig davon möchte ich ein weiteres Argument anbringen, diesmal aus naturalistischer Perspektive. Selbst wenn es kein cartesisches Subjekt geben sollte, muss es einen physikalischen spezifizierbaren 'Eigentümer' der Meme geben. So wie Gene Zellen brauchen und Zellen keine Gene sind, brauchen Meme einen Geist und dieser hat Eigenschaften bzw. Fähigkeiten, die nur schwerlich als Meme rekonstruierbar sind.[32] Bewusstsein oder kognitive Fähigkeiten, wie z.B. die Fähigkeit mit abstrakten symbolischen Zeichen umgehen zu können, sind selbst *keine* Meme, da sie nicht durch soziales Lernen erworben werden. Die Fitness von Memen ist von diesen Fähigkeiten des Geistes, die selber keine Meme sind, abhängig und würde stark variieren, wenn diese grundlegenden Eigenschaften nicht mehr zur selektiven Umwelt der Meme gehören würden.

Aber selbst dann, wenn der menschliche Geist nur aus Memen bestehen *würde*, ohne Subjekt und andere Eigenschaften und Fähigkeiten, wäre, so mein letztes kritisches Argument, die Diffusion eines jeden einzelnen Mems *nie* unabhängig von seiner selektiven Umwelt, die von diesem einzelnen Mem zu unterscheiden ist, wie auch Laland & Brown in ihrer Antwort auf das Tautologieproblem argumentieren.[33] Selbst wenn der Geist also ein bloßes Bündel aus Memen wäre, kann dies keinen Ausweg aus dem explanatorischen Dilemma der Egoismusanalogie bieten, unabhängig davon, ob die naturalistische Bündeltheorie zutrifft oder nicht. Die These, dass der Geist nichts weiter als ein Bündel aus Memen ist, kann also nicht benutzt werden, um die Egoismusthese zu retten.

– Was bleibt von der Egoismusthese? Nichts Spektakuläres: Sie besagt, dass wir manchmal irrational sind, d.h. dass wir Dinge tun, die wir nicht tun wollen. Sie besagt, dass unser Geist bestimmte Inhalte hat. Mit anderen Worten, dass es etwas gibt, in welchem Sinne auch immer, das der Geist denkt. Beides stimmt, den Darwinismus brauchen wir dazu jedoch nicht bemühen, weder um dies zu sehen, noch um es zu erklären.

Ein Schaf im Wolfspelz

Die These, dass es Meme gibt und dass sie Replikatoren im weiten Sinne sind (sich durch soziales Lernen 'fortpflanzen'), ist heuristisch trivial. Die Geistes-, Sozial- und Kulturwissenschaften gehen alle seit langem davon aus, unabhängig vom Darwinismus. Die Memtheorie bietet zudem nichts Neues, um die damit verbundenen ontologischen Fragen über den Status der postulierten ideellen Einheiten zu klären, und tritt nicht als Alternative zu den detaillierten Erklärungen wie soziales Lernen funktioniert auf und ist somit auch explanatorisch trivial. Die These, dass Meme Replikatoren im engen Sinne sind, ist für alle jene Meme falsch, die durch mehrmaligen Kontakt mit einem Set aus 'Originalen' erworben werden und somit keine Lineage bilden. Die dritte These, dass die Fitness von Memen kulturelle Diffusionsprozesse erklären könne, ist entweder heuristisch trivial oder sie ist tautologisch und somit explanatorisch trivial: Entweder wird nichts weiter behauptet, als dass Menschen aus diesen oder jenen Gründen bzw. Ursachen bestimmte Meme übernehmen, oder die 'survival of the fittest meme'-Erklärungen werden tautologisch, weil die Rolle des Geistes als selektive Umwelt ignoriert wird.

Die drei Thesen der Memtheorie beinhalten somit keine einzige 'gefährliche Idee' (Dennett 1995), die unser Selbstverständnis von Kultur verändern könnte. Wir haben es somit nicht mit einem 'gefährlichen Wolf' zu tun, sondern vielmehr mit einem 'Schaf im Wolfspelz', mit einem harmlosen Geschöpf, das sich lediglich als gefährlich gebärdet. Wenn wir uns rein auf den *Gehalt* der drei Thesen konzentrieren, dann ist die Memtheorie eine modische Redensart, ohne Potential den Geistes-, Kultur- und Sozialwissenschaften auch nur irgendeine Einsicht hinzuzufügen.

3. Eine wissenschaftstheoretische Einordnung der Memtheorie

Ein Blick auf die Funktion der Theorie

Doch es ist auch eine andere Perspektive möglich, um die Memtheorie zu bewerten. Aufgabe ist dann nicht nur zu verstehen, *was* diese Theorie besagt, welchen *Gehalt* sie hat, sondern es gilt zu verstehen, *wieso* die Theorie postuliert wird, d.h. welcher *Zweck* damit verfolgt wird. Neben den expliziten Erklärungszielen (kulturelles Lernen, Diffusionsprozesse,

Irrationalität, Geist) können dabei auch implizite Ziele bzw. Funktionen betrachtet werden.

Appendix des Genselektionismus

So hat Dawkins (1982a, 112) betont, dass der Wert der Analogie gar nicht in der Erklärung von Kultur liege. Die Analogie könne aber helfen, das Wirken der natürlichen Selektion (auf der Basis von Replikatoren) besser zu verstehen. Diese Strategie hat historische Vorläufer: Um seine eigene Neo-Darwinistische Evolutionstheorie zu verteidigen, griff bereits August Weismann auf die Idee der Kultur zurück. Er konstruierte Kultur analog zu biologischer Vererbung und als Ersatz für Lamarckistische Vererbung, um seine Neo-Darwinistische Evolutionstheorie zu untermauern und gegen Kritiker zu verteidigen. Dabei hatte er natürlich v.a. das Lager derer im Blick, die Lamarckistische Vererbung verteidigten, insbesondere Herbert Spencer, mit dem sich Weismann in den 1890ern um die 'all-sufficiency' der natürlichen Selektion stritt.[34]

Ob eine solche Funktion der Analogie zielführend ist, muss im Einzelnen betrachtet werden. In Dawkins Fall ist festzuhalten, dass die These, dass es neben Genen auch noch Meme, d.h. Kultur, gibt, zwar keine neue These darstellt, wie ich oben dargestellt habe, dass die These aber innerhalb des Paradigmas des Genselektionismus fruchtbar sein kann, um beispielsweise deutlich zu machen, dass Genselektionismus nicht mit Gendeterminismus gleichgesetzt werden darf. Die Memtheorie wäre somit aber nicht als eigenständige Theorie zu behandeln, sondern lediglich als ein Appendix des Genselektionismus und in dieser Funktion in der Tat als informativ. Gleichzeitig steht und fällt die Idee dann natürlich mit dem Genselektionismus.

Eine Interfeldtheorie

Weit wichtiger, da größere Bereiche der Wissenschaft betreffend, ist jedoch eine andere, eine integrative bzw. *inter*disziplinäre Funktion. Die Wissenschaftstheorie geht heute davon aus, dass ausdifferenzierte Disziplinen durch bestimmte 'interfield theories' verbunden sind.[35] Die Memtheorie ist m. E. eine solche Interfeldtheorie. Sie verbindet biologische mit geistes- bzw. sozialwissenschaftlichen Disziplinen, obwohl sie diesen Disziplinen nichts Neues hinzufügt, außer einer gemeinsamen Sprache. Wie Hull (2000, 43, 46) vermerkt, kann die Memtheorie diese Disziplinen zusammenbringen, um den Menschen wieder in seiner Gesamtheit

in den Blick zu bekommen. Die gemeinsame Sprache ermöglicht die dringend notwendige Verständigung über Disziplingrenzen hinweg. Sie erlaubt einen Vergleich der Ansichten und somit auch die gegenseitige Positionierung ohne die arbeitsteilige Struktur der modernen Wissenschaft über Bord zu werfen. In einer Zeit, in der die phänomenale Gesamtheit des Menschen aufgrund der Differenzierung der Lebens- und Humanwissenschaften unwiederbringlich verloren ist, ist dies eine ernstzunehmende und wichtige Funktion. Jenseits des Genselektionismus, d.h. unabhängig von dieser Version des Darwinismus, kann die Memtheorie die Rolle einer Interfeldtheorie einnehmen und als solche auch nützlich sein.

Eine Wittgenstein'sche Leiter zwischen den Disziplinen

Oft hat man jedoch den Eindruck, dass die Motivation hinter der Memtheorie eher 'imperialistischer' Natur ist: Die Evolutionstheorie soll helfen, die Mauern zwischen den sog. „zwei Kulturen"[36] einzureißen und den Bereich jenseits des naturwissenschaftlichen 'Reiches' mit einer Theorie des „unity of design space" zu beglücken.[37] Mit anderen Worten, die Geistes-, Sozial- und Kulturwissenschaftler sollen sich doch bitte endlich von der Evolutionstheorie, einer echten Naturwissenschaft, helfen lassen, um aus ihrem Interpretationsgeflecht und Theorienebeneinander ausbrechen zu können. Sie würden bekommen, was sie aus eigener Kraft bisher nicht entwickeln konnten und aus der Perspektive der Naturwissenschaftler doch so dringend zu brauchen scheinen: ein allumfassendes theoretisches 'Empire', das für Ordnung, Klarheit und Fortschritt im unterentwickelten Hause sorgt. Kurzum, sie könnten eine echte Theorie der Kultur bekommen: die Memetik, so hart und klar und naturalistisch wie die Genetik.

Obwohl die Memtheorie dazu dienen kann, die unterschiedlichen Disziplinen, die sich mit dem Menschen beschäftigen, ins Gespräch zu bringen, sollte dies nicht darüber hinwegtäuschen, dass sie trotzdem einer Wittgenstein'schen Leiter vergleichbar ist, die weggeworfen werden sollte, sobald man zu den Details kommt. Die Details sollten auch nach wie vor in den einzelnen Disziplinen bearbeitet werden. Damit ist einem wissenschaftlichen Imperialismus zwischen Naturwissenschaften einerseits und Geistes- und Sozialwissenschaften andererseits ein Riegel vorgeschoben. Die Memtheorie bildet kein neues, aus der Naturwissenschaft importiertes, theoretisches Fundament für die Sozial- und Geis-

teswissenschaften. Sie ist lediglich eine Theorie *zwischen* den Disziplinen, im echten Sinne des Wortes.

Anmerkungen

1 Mein Dank gilt Hans Rott, Veronika Roth und Konrad Bachmann für ihre hilfreichen Kommentare. Die hier vorgestellten Argumente finden sich ausführlicher in Kronfeldner (im Erscheinen).
2 Siehe Laland & Brown (2002) für eine kurze Zusammenfassung des von der Memtheorie völlig unabhängigen, populationsgenetischen Ansatzes. Richerson & Boyd (2005) ist eine ausführliche neuere Beschreibung dieses Ansatzes. Jablonka & Lamb (2005) integrieren diesen Ansatz in ihre erweiterte Evolutionstheorie, die 'vier Dimensionen' beinhaltet, eine genetische, epigenetische, behaviorale und symbolische Dimension.
3 Siehe z.B. Aunger (2007), Blackmore (2010), Blute (2010), Wimsatt (2010).
4 Strittig war jedoch immer und ist es noch, wie Gene individuiert werden können, d.h. welche Abschnitte einer DNA-Sequenz als einzelne Gene bezeichnet werden sollten. Siehe Beurton et al (2000), Griffiths (2007), Müller-Wille & Rheinberger (2009).
5 Siehe Kronfeldner (im Erscheinen) für eine detaillierte und vergleichende Analyse der einzelnen ontologischen Annahmen über Meme.
6 Siehe Dawkins' ursprüngliche Definition der Meme in Dawkins (1976, 192), die später (1982a, 109–112) präzisiert wurde. Vgl. Hull (1982, 275–6, 310; 2000, 58–61), Dennett (1991, 201–208; 1995, 342–360), Blackmore (1999, 14–16, 63–66). Abweichend von dem dargestellten ideellen Membegriff, setzt Aunger (2002) Meme mit den neuronalen Zuständen gleich. Ein behavioraler Membegriff wurde ebenso vertreten, z.B. von Gatherer (1998). Um die Annahme unbeobachtbarer ideeller Einheiten zu umgehen, werden Meme dabei Verhalten bzw. Artefakten gleichgesetzt. Ein Rad oder ein Buch ist ein Mem, so die Annahme.
7 Siehe Brandon & Burian (1984), die eine gute Übersicht der klassischen Positionen zur Debatte bieten, ob Gene, Individuen, oder auch Gruppen Einheiten der Selektion sein können.
8 Dawkins hat diese Position in Dawkins (1976) vorgestellt und v.a. in Dawkins (1978, 1982a, 1982b) weiterentwickelt.
9 Für eine ausführlichere Debatte siehe Godfrey-Smith (2000).
10 Replikation kann sich nur auf hinreichende Ähnlichkeit beziehen, da sonst Mutation ausgeschlossen würde, was nicht im Sinne von Dawkins sein kann.
11 Siehe Dawkins (1976, 273–4; 1982a, 97–99).
12 Siehe dazu Dawkins (1976, 274). Zu den hier nicht erwähnten, weiteren Bedingungen siehe Kronfeldner (im Erscheinen).
13 Siehe z.B. die berühmt-berüchtigte Stelle in Dawkins (1976, 19).
14 Dennett (1991, 1995, 2000, 2001).
15 Blackmore (1999, 2000).

16 Dies gilt selbst wenn Gene wie Meme gleich schlecht in Bezug auf ein anderes Identifikationsproblem abschneiden: Es ist bei Genen wie Memen schwer festzulegen, welche Abschnitte einer DNA-Sequenz bzw. welche Teile einer Kultur als *einzelnes* Gen bzw. Mem individuiert werden sollen. Siehe Kronfeldner (im Erscheinen) zur Unterscheidung verschiedener Probleme der Identifikation; siehe Hull (2000, 48) für das Argument, dass Gene wie Meme in Bezug auf die Partionierung in einzelne Einheiten gleich schlecht abschneiden.
17 Als Beispiel sei auf Dennett (1995, 352–354) verwiesen, der das Problem der materiellen Identifikation eindeutig zugesteht.
18 Diese Kritik bezieht sich natürlich nur auf den ideellen Membegriff. Auf die Frage wieso ein neuronaler oder behavioraler Membegriff keine Lösung darstellt, kann hier nicht eingegangen werden. Siehe Kronfeldner (im Erscheinen) zu dieser Art Rückzugsgefecht.
19 Siehe Boyd & Richerson (2000, 155–6).
20 Siehe z. B. Dennett (1995, 359).
21 Zum Kulturbegriff siehe z. B. Keesing (1974) oder Kuper (1999).
22 Siehe z. B. Tomasello (1999). Eine gute Übersicht über die Theoriebildung bietet auch Heyes (1994).
23 Dawkins (1976, 200; Hervorh. im Orig.). Siehe auch Dennett (1991, 203-7; 1995, 361–4) oder z. B. Blackmore (1999, 24, 30–32).
24 Das Tautologieproblem wurde bereits von Sober (1992) angeführt und von Sterelny & Griffiths (1999), Wilson (1999) und Conte (2000) wiederholt, jedoch ohne Bezug auf den Unterschied zwischen Fitness und Angepasstheit und teilweise für alle Theorien der kulturellen Evolution gleichermaßen.
25 Siehe Kitcher (2000) und Kronfeldner (2009) zu diesem Themenbereich.
26 Sterelny (2006, 158–162)
27 Dennett (1991, 205), Dennett (2001).
28 Siehe Dawkins (1993) für diese Terminologie zur Bezeichnung irrationalen Verhaltens.
29 Siehe v.a. Dennett (1991, 207–226) oder aber auch Blackmore (1999, 219–234).
30 Clark (1993, 12–14) hat ein solches Argument gegen die Memtheorie vorgebracht.
31 Siehe Candlish (1998) für einen Überblick zur Geschichte der Bündeltheorie des Geistes von Hume bis heute.
32 Siehe z. B. Dennett (1989).
33 Obwohl Clark (1993, 12–14) von einem nicht-naturalistischen Standpunkt aus argumentiert, übernehme ich diesen Vergleich zwischen Geist und Zellen von ihm.
34 Siehe Laland & Brown (2002, 235).
35 Siehe Weismann (1892).
36 Der Begriff geht auf Darden & Maull (1977) zurück.
37 Snow (1969).
38 Dennett (1995, 189).

Literatur

Aunger, Robert, 2002: *The Electric Meme: A New Theory of How We Think*. New York: Free Press.

Aunger, Robert, 2007: Memes. In: Dunbar, Robin; Barrett, Louise (Hrsg.): *The Oxford Handbook of Evolutionary Psychology.* Oxford: Oxford University Press, S. 599–604.

Beurton, Peter J.; Falk, Raphael; Rheinberger, Hans-Jörg (Hrsg.), 2000: *The Concept of the Gene in Development and Evolution: Historical and Epistemological Perspectives*. Cambridge: Cambridge University Press.

Blackmore, Susan, 1999: *The Meme Machine*. Oxford: Oxford University Press.

Blackmore, Susan, 2000: The power of memes. In: *Scientific American* 283, S. 64–73.

Blackmore, Susan, 2010: Memetics does provide a useful way of understanding cultural evolution. In: Ayala, Francisco J.; Arp, Robert (Hrsg.): *Contemporary Debates in Philosophy of Biology*. New York: Wiley, S. 255–272.

Blute, Marion, 2010: *Darwinian Sociocultural Evolution: Solutions to Dilemmas in Cultural and Social Theory*. Cambridge: Cambridge University Press.

Boyd, Robert; Richerson, Peter J., 1985: *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.

Boyd, Richard; Richerson, Peter J., 2000: Memes: Universal acid or a better mousetrap? In: Aunger, Robert (Hrsg.): *Darwinizing Culture: The Status of Memetics as a Science*. Oxford: Oxford University Press, S. 143–162.

Brandon, Robert N.; Burian, Richard M. (Hrsg.), 1984: *Genes, Organisms, Populations: Controversies over the Units of Selection*. Cambridge, MA: MIT Press,.

Campbell, Donald T., 1965: Variation and selective retention in sociocultural evolution. In: Barringer, Herbert R.; Blanksten, George I.; Mack, Raymond W. (Hrsg.): *Social Change in Developing Areas: A Reinterpretation of Evolutionary Theory*. Cambridge, MA: Schenkman, S. 19–49.

Candlish, Stewart, 1998: Bundle theory of mind. In: Craig, E. (Hrsg.): *Routledge Encyclopedia of Philosophy, Version 1.0*. London and New York: Routledge.

Cavalli-Sforza, Luigi L.; Feldman, Marc, 1981: *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton: Princeton University Press.

Clark, Stephen R. L., 1993: Minds, memes, and rhetoric. In: *Inquiry* 36, S. 3–16.

Conte, Rosaria, 2000: Memes through (social) minds. In: Aunger, Robert (Hrsg.): *Darwinizing Culture*. Oxford: Oxford University Press, S. 83–119.

Darden, Lindley; Maul, Nancy, 1977: Interfield Theories. In: *Philosophy of Science* 44, S. 43–64.

Darwin, Charles, 1859: *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: Murray.

Dawkins, Richard C., 1976: *The Selfish Gene*. Oxford: Oxford University Press.

Dawkins, Richard, 1978: Replicator selection and the extended phenotype. In: *Zeitschrift für Tierpsychologie* 47, S. 61–76.

Dawkins, Richard C., 1982a: *The Extended Phenotype: The Gene as the Unit of Selection*. Oxford: Oxford University Press.

Dawkins, Richard, 1982b: Replicators and vehicles. In: King's College Sociobiology Group (Hrsg.): *Current Problems in Sociobiology*. Cambridge: Cambridge University Press, S. 45–64.

Dawkins, Richard, 1993: Viruses of the mind. In: Dahlbom, Bo (Hrsg.): *Dennett and his Critics*. Cambridge: Blackwell, S. 13–27.

Dennett, Daniel C., 1989: The origins of selves. In: *Cogito* 3, S. 163–73.

Dennett, Daniel C., 1991: *Consciousness Explained*. Boston: Little, Brown & Co.

Dennett, Daniel C., 1995: *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. New York: Simon and Schuster.

Dennett, Daniel C., 2000: Foreword to 'Darwinizing Culture'. In: Aunger, Robert (Hrsg.): *Darwinizing Culture: The Status of Memetics as a Science*. Oxford: Oxford University Press, S. v–ix.

Dennett, Daniel C., 2001a: The evolution of culture. In: *Monist* 84, S. 305–324.

Durham, William H., 1991: *Coevolution: Genes, Culture, and Human Diversity*. Stanford: Stanford University Press.

Gatherer, Derek, 1998: Why the thought contagion metaphor is retar-

ding the progress of memetics. In: *Journal of Memetics* 2, S. http://jom-emit.cfpm.org/1998/vol2/gatherer_d.html.

Godfrey-Smith, Peter, 2000: The replicator in retrospect. In: *Biology and Philosophy* 15, S. 403–423.

Griffiths, Paul; Stotz, Karola, 2007: Gene. In: Hull, David L.; Ruse, Michael (Hrsg.): *Cambridge Companion to the Philosophy of Biology*. Cambridge: Cambridge University Press, S. 85–102.

Herder, Johann G., 1784: Ideen zur Philosophie der Geschichte der Menschheit. In: Suphan, Bernhard. (Hrsg.), 1887: *Sämtliche Werke*, Bd. 13. Berlin: Weidmannsche Buchhandlung.

Heyes, Cecilia M., 1994: Social learning in animals. In: *Biological Reviews* 69, S. 207–231.

Hull, David L., 1982: The naked meme. In: Plotkin, Henry C. (Hrsg.): *Learning, Development, and Culture*. Chichester: Wiley, S. 273–27.

Hull, David L., 2000: Taking memetics seriously: Memetics will be what we make it. In: Aunger, Robert (Hrsg.): *Darwinizing Culture: The Status of Memetics as a Science*. Oxford: Oxford University Press, S. 43–67.

Jablonka, Eva; Lamb, Marion J., 2005: *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. Cambridge, MA: MIT Press.

James, William, 1880: Great men and their environment. In: Burkhardt, Frederick H.; Bowers, Fredson; Skrupskelis, Ignas K. (Hrsg.): *The Works of William James, Bd. 6: The Will to Believe and Other Essays in Popular Philosophy*. Cambridge, MA: Harvard University Press, S. 163–189.

Keesing, Roger M., 1974: Theories of culture. In: *Annual Review of Anthropology* 3, S. 73–97.

Kitcher, Philip, 2000: Battling the undead: How (and how not) to resist genetic determinism. In: Kitcher, Philip (Hrsg.): *In Mendel's Mirror: Philosophical Reflections on Biology*. Oxford: Oxford University Press, S. 283–300.

Kronfeldner, Maria E., im Erscheinen: *Who's Afraid of Darwinism? Creativity and Memetics Demystified*. Durham: Acumen.

Kronfeldner, Maria E., 2009: Genetic determinism and the innate-acquired distinction. In: *Medicine Studies* 2, S. 167–181.

Kuper, Adam, 1999: *Culture: The Anthropologists Account*. Cambridge, MA: Harvard University Press.

Laland, Kevin N.; Brown, Gillian R., 2002: *Sense and Nonsense: Evolutionary Perspectives on Human Social Behavior*. Oxford: Oxford University Press.

Müller-Wille, Staffan; Rheinberger, Hans-Jörg, 2009: *Das Gen im Zeitalter der Postgenomik: Eine wissenschaftshistorische Bestandsaufnahme*. Frankfurt: Suhrkamp.

Richerson, Peter J.; Boyd, Robert, 2005: *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press.

Semon, Richard, 1904: *Die Mneme als erhaltendes Prinzip im Wechsel des organischen Geschehens*. Leipzig: Engelmann.

Snow, Charles P., 1969: *The Two Cultures: And a Second Look*. Cambridge: Cambridge University Press.

Sober, Elliott, 1992: Models of cultural evolution. In: Griffiths, Paul (Hrsg.): *Trees of Life: Essays in Philosophy of Biology*. Dordrecht: Kluwer, S. 17–39.

Sterelny, Kim; Griffiths, Paul E., 1999: *Sex and Death: An Introduction to Philosophy of Biology*. Chicago, London: University of Chicago Press.

Sterelny, Kim, 2006: Memes revisited. In: *British Journal for the Philosophy of Science* 57, S. 145–165.

Tomasello, Michael, 1999: *The Cultural Origins of Human Cognition. Cambridge*, MA: Harvard University Press.

Weismann, August, 1892: Gedanken über Musik bei Thieren und beim Menschen. In: Weismann, August (Hrsg.): *Aufsätze über Vererbung und verwandte biologische Fragen*. Jena: Gustav Fischer, S. 587–637.

Wilson, David S., 1999: Flying over uncharted territory. In: *Science* 285, S. 206.

Wimsatt, William C., 2010: Memetics does not provide a useful way of understanding cultural evolution: A developmental perspective. In: Ayala, Francisco J.; Arp, Robert (Hrsg.): *Contemporary Debates in Philosophy of Biology*. New York: Wiley, S. 273–291.

Matthias Rang, Olaf L. Müller

Newton in Grönland

Das umgestülpte *experimentum crucis* in der Streulichtkammer

Zusammenfassung

Newtons *experimentum crucis* hat ein komplementäres Gegenstück, d.h. ein Experiment, in dem die Rollen von Licht und Schatten genau ausgetauscht sind. Statt wie Newton in der Dunkelkammer zu experimentieren, müssen wir das Komplement des *experimentum crucis* in einer Streulichtkammer aufbauen (deren Wände sog. Lambertstrahler sind). Wenn es dieses umgestülpte Experiment wirklich gibt, dann liefert es für jeden newtonischen Beweis einen umgestülpten Gegenbeweis, dessen Konklusion die Heterogenität der Schatten wäre (also die Behauptung, dass nicht weißes Licht, sondern schwarze Schatten eine heterogene Mischung verschiedenfarbiger Strahlen mit unterschiedlichen Brechungseigenschaften seien). Dass Newtons *experimentum crucis* in diesem Sinne umgestülpt werden kann, wird von Newtons eigener Theorie impliziert. Mehr noch, inzwischen ist der empirische Nachweis der Umstülpung gelungen.

Abstract

Newton's *experimentum crucis* has a complementary counterpart, i.e., an experiment in which the roles of light and shadow are switched. Instead of experimenting in the dark room, à la Newton, we propose to perform the complement of the *experimentum crucis* in a bright room (whose walls consist of Lambertian radiators). If Newton's experiment can be turned upside-down, then for each Newtonian conclusion (derived from the original experiment) there is a counterconclusion (derived from the new experiment). Thus, we arrive at the heterogeneity of darkness, which is an unorthodox hypothesis: Darkness and shadows (rather than white light) consist of rays that differ with respect to colour and refrangibility. Newton's own theory implies that in his *experimentum crucis* the roles of light and darkness can be interchanged; moreover, their very interchangeability can be demonstrated empirically.

philosophia naturalis 46 / 2009 / 1

I. Einleitung

Das *experimentum crucis* ist die berühmteste experimentelle Leistung des Physikers Isaac Newton.[1] Bis heute streiten sich die Gelehrten darüber, was Newton genau mit diesem Experiment zeigen wollte, auf welche Weise er dies hat zeigen wollen und wie erfolgreich er dabei war.[2] Klar ist nur, dass Newton mit dem Experiment auf die Heterogenität des Sonnenlichts abzielte. Diesem newtonischen Lehrsatz zufolge besteht Sonnenlicht (entgegen seinem homogenen Anschein) aus Lichtstrahlen mit unterschiedlichen Brechungseigenschaften,[3] ja, aus Lichtstrahlen unterschiedlicher bunter Farben.[4]

Ein Nebenschauplatz der Streitigkeiten um Newtons Experiment hat mit einer bemerkenswerten optischen Symmetrie der Natur zu tun, mit einer Symmetrie, die im Schulwissen der Physik nicht kanonisiert ist. Es handelt sich um eine Symmetrie zwischen weißem Licht und Finsternis sowie zwischen beliebigen Farben und deren komplementären Gegenstücken – also um eine Symmetrie zwischen Schwarz und Weiß, zwischen Grün und Purpur, zwischen Gelb und Blau sowie zwischen Türkis und Rot. Es ist dieselbe farbliche Symmetrie, die zwischen einem Farbphoto und dessen Negativ herrscht (zumindest, wenn man davon absieht, dass die Komplementärfarben der Farbnegative – aus technischen Gründen – zusätzlich rostrot getönt sind).

Wo zeigt die Natur diese Symmetrie? Sie zeigt sie in den prismatischen Experimenten, deren Pionier Newton war. Nehmen Sie ein einfaches prismatisches Experiment Newtons und vertauschen Sie darin die Rollen von weißem Licht und Schatten, drehen es gleichsam optisch um, invertieren es. Dann liefert Ihnen die Natur ein anderes prismatisches Experiment, das in *allen* seinen Aspekten wie ein exaktes Farbnegativ des Ausgangsexperiments aussieht. Die Vertauschung von Weiß und Schwarz führt also zur Umkehrung *aller anderen Farben* des Ausgangsexperiments. Wo z.B. in Newtons Grundexperiment ein Prisma den weißen Sonnenstrahl in Strahlen der Regenbogenfarben Blau, Grün, Rot (mit vielen fein abgestuften Zwischentönen) zerspaltet, da sieht man im invertierten Gegenexperiment das Komplement der Regenbogenfarben, deren Farbnegativ, also die Farben Gelb, Purpur, Türkis, ebenfalls mit Zwischentönen (vergl. Abb. 1 und Abb. 2 mit Abb. 3). Zur Farbterminologie: Um der Kürze willen werden wir bei Newtons Spektrum ebenso wie bei dessen Komplement jeweils nur von den drei erwähnten Farben

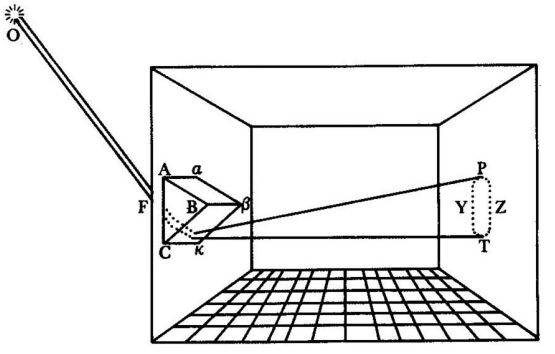


Abb. 1: Newtons Grundexperiment. Ein weißer Lichtstrahl reist von der Sonne O durchs Fensterladenloch F in die Dunkelkammer; dort wird er sogleich beim Eintritt ins Prisma (an der Fläche ACκα) zum Lot hingebrochen und beim Austritt aus dem Prisma (an der Fläche BCβκ) vom Lot weggebrochen. Bei dieser zweifachen Refraktion fächert sich der weiße Lichtstrahl in seine regenbogenbunten Bestandteile auf und zeichnet ein längliches Spektrum PYTZ an die gegenüberliegende Wand. (Aus Newtons *Lectiones opticae* ([LOOL]:50).

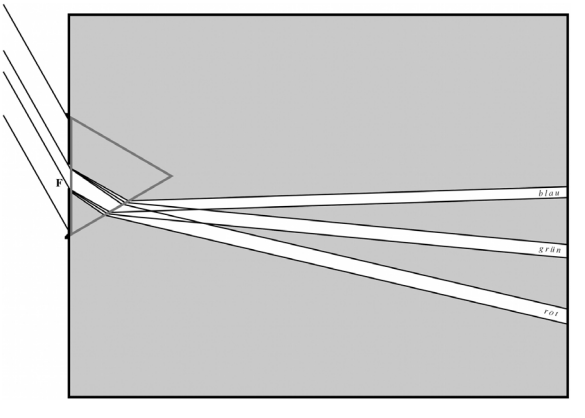


Abb. 2. Schematische Darstellung des Strahlengangs in Newtons Grundexperiment. Wir haben drei Strahlenbündel unterschiedlicher Refrangibilität hell hervorgehoben. Das obere (blaue) Strahlenbündel wird am stärksten von seinem Weg abgelenkt, das untere (rote) am schwächsten; mittlere Ablenkung erfährt das (grüne) Strahlenbündel in der Mitte. Die Farbnuancen zwischen diesen drei Hauptfarben des newtonischen Spektrums haben wir hier ausgeblendet; wir werden uns auch in den anderen Abbildungen auf jeweils drei Farben beschränken. Für diese und die nächsten Abbildungen haben wir Newtons Vereinfachungen beim Zeichnen der Abbildungsstrahlen unendlich entfernter Lichtquellen übernommen.

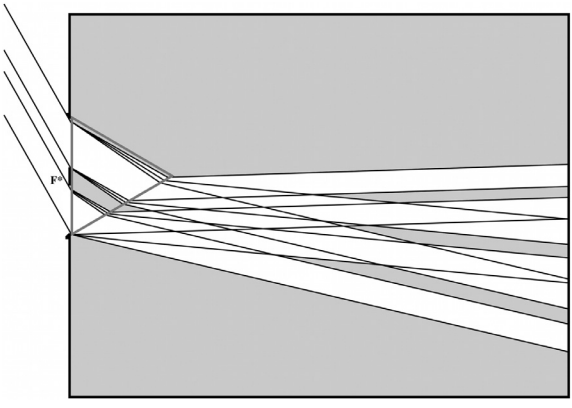


Abb. 3. Schematische Darstellung des newtonischen Strahlengangs im invertierten Grundexperiment.[5] Grau unterlegt haben wir diejenigen Richtungen, in denen sich laut Newton kein eigenes optisches Geschehen fortpflanzt. (Es sind dieselben Richtungen, die in der Abb. 2 weiß hervortreten). Aus newtonischer Perspektive pflanzen sich die drei Lichtstrahlensorten in den hier weiß hervorgehobenen Richtungen jeweils *doppelt* fort, einmal oberhalb des Schattenwerfers F*, einmal darunter.

Blau-Grün-Rot bzw. Gelb-Purpur-Türkis reden; sie treten dort jeweils am deutlichsten hervor. Wer nicht gern auf die *Zwischentöne* in Newtons Spektrum Blau-*Hellblau*-Grün-*Orange*-Rot verzichten möchte, kann diese Zwischentöne überall in unserer Argumentation mitlaufen lassen; dasselbe gilt für die Zwischentöne im Komplementärspektrum Gelb-*Hellrot*-Purpur-*Blau*-Türkis. (Was wir hier kursiv hinzugefügt haben, gibt natürlich ebensowenig wie unsere ärmere Terminologie die volle Farbenvielfalt der beiden Spektren wieder).

Dass sich *einfache* prismatische Experimente invertieren lassen, ist seit Newtons Tagen bekannt; Goethe hat in seiner Newton-Kritik besonders deutlich darauf hingewiesen.[6] Doch bislang konnte nicht endgültig geklärt werden, ob sich auch Newtons *experimentum crucis* invertieren lässt.[7]

Wir möchten ein neues Experiment beschreiben, das dies leistet. Newtons Theorie sagt voraus, dass die Invertierung des *experimentum crucis* funktionieren muss. Sollte das stimmen, so wirft dies Ergebnis ein kritisches Licht auf Newtons Anspruch, mit seinem Experiment einen *eindeutigen* Beweis zugunsten seiner Theorie geliefert zu haben.

II. Newtons berühmtestes Experiment

Bevor wir Newtons *experimentum crucis* beschreiben, möchten wir betonen, dass Newton in seiner Beschreibung des Experiments aus dem Jahr 1672 keinerlei Farben erwähnt. Das gesamte Experiment könnte erfolgreich von einem völlig farbenblinden Physiker durchgeführt werden. Wir werden gleich dennoch allerlei Farben ins Spiel bringen, weil man dadurch besser verstehen kann, was in dem Experiment vor sich geht; das ist ein didaktisches Mittel der Darstellung, kein wesentlicher Zug des Experiments selbst.

Betrachten Sie Abb. 4, die einer Abbildung aus Newtons *Opticks* nachempfunden ist.[8] Newton lässt Sonnenlicht durch ein Prisma ABC scheinen, baut aber – anders als in dem Grundexperiment, das wir in Abb. 1 und Abb. 2 dargestellt haben – unmittelbar hinter diesem Prisma einen Schirm DE auf. Nun können sich die verschiedenfarbigen (und divers refrangiblen) Lichtstrahlen unmittelbar hinter dem Prisma ABC noch nicht weit genug voneinander entfernt haben, um auf dem Schirm DE ein farbiges (und merklich in die Länge gezerrtes) Spektrum zu hinterlassen. Nur an den äußeren Enden des dort aufgefangenen Bildes werden sich Farben zeigen. Oben auf dem Schirm zeigt sich ein blauer Farbfleck (da dorthin aus dem durchs Prisma kommenden Gesamtstrahlenbündel nur die oberen der stärker refrangiblen Lichtstrahlen gelangen können, ohne von irgendwelchen schwach refrangiblen Lichtstrahlen überlagert werden zu können); und am unteren Ende des Bildes auf dem Schirm DE zeigt sich ein roter Farbfleck (da dorthin aus dem durchs Prisma kommenden Gesamtstrahlenbündel nur die unteren der schwach refrangiblen Lichtstrahlen gelangen können, ohne Störung durch irgendwelche stärker refrangiblen Strahlen). Kurzum, der größte Teil des Bildes auf dem Schirm DE ist weiß; und genau in seiner Mitte versammeln sich (i) vom unteren C-Ende des Prismas die stark refrangiblen blauen Strahlen, (ii) vom oberen A-Ende des Prismas die schwach refrangiblen roten Strahlen, und (iii) aus der Mitte des Prismas die mittelmäßig refrangiblen grünen Strahlen (siehe Abb. 5 im Anhang).[9]

Nun hat Newton in der Mitte dieses Schirms ein winziges Loch G gebohrt. Nur Strahlen aus dem eben erwähnten weißen Lichtgemisch können durch das Loch hindurch, und zwar je nach Farbe in unterschiedlicher Richtung.[10] In hinreichender Entfernung hinter dem Loch werden sich alle diese Lichtstrahlen weit genug voneinander entfernt

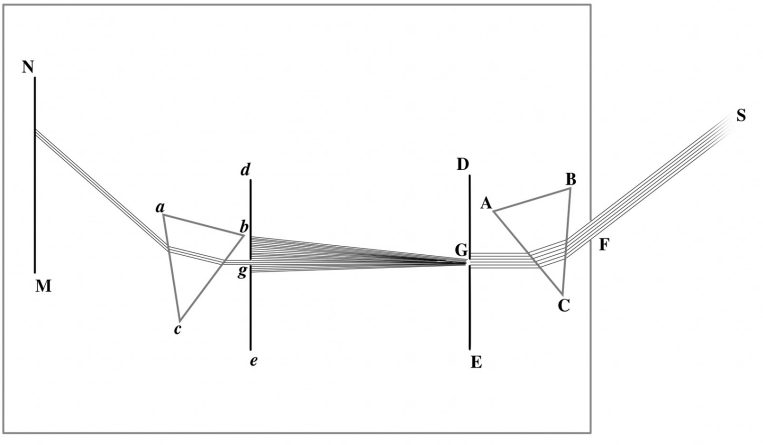


Abb. 4. Nachzeichnung des *experimentum crucis*. Wir haben alle Wände der Dunkelkammer eingezeichnet, die in Newtons Zeichnung zum größten Teil fehlen. Zudem haben wir eine idealisierte Pointe des newtonischen Versuchs berücksichtigt, die in Newtons Zeichnung untergeht: Anders als es seine Theorie im Idealfall verlangt, reisen in Newtons Zeichnung die Lichtstrahlen vom zweiten Prisma abc nicht parallel zum Schirm NM; das haben wir berichtigt. Der empirischen Realität angemessener ist dennoch Newtons Zeichnung. (Einen anderen Fehler der newtonischen Zeichnung haben wir nicht berichtigt; in seiner und in unserer Zeichnung verlaufen alle Strahlen zwischen Prisma ABC und Schirm DE parallel; sie müssten das Prisma aber in unterschiedlicher Richtung verlassen – sonst können sie sich hinter der Blendenöffnung G im Schirm DE kaum auffächern, wie richtigerweise in beiden Zeichnungen angedeutet).

haben, ohne sich noch in die Quere kommen zu können. Das bestätigt der weitere Verlauf des Experiments: Zwölf Fuß hinter dem Schirm DE bringt Newton einen zweiten Schirm de an. Auf diesem Schirm zeigt sich das bereits bekannte newtonische Farbenspektrum in seiner vollen Entwicklung: ganz oben blau, dann grün und rot – mit beliebig feinen Zwischenstufen. Und seine Länge übersteigt seine Breite um ein Vielfaches. Wenn man diese Farbenvielfalt auf drei Farben reduziert, so ergibt sich ein Bild wie in Abb. 6 im Anhang.

Bis hierhin entspricht Newtons *experimentum crucis* dem Grundexperiment aus Abb. 1 und Abb. 2. Fast könnte man sagen, dass das Grundexperiment als Bestandteil im *experimentum crucis* enthalten ist. Zwar zwingt Newton diesmal das Licht durch ein winziges Loch unmittelbar *hinter* dem Prisma; anders als im Grundexperiment, wo das Licht

vor dem Prisma durch ein Loch hindurch musste: durch das Loch im Fensterladen.[11] Aber dieser Unterschied in der Versuchsanordnung fördert keine unterschiedlichen Spektren zutage und bringt zunächst keine wesentlichen theoretischen Unterschiede mit sich.[12]

Jetzt kommt Bewegung ins Spiel: Newton dreht das Prisma ABC langsam um seine Achse hin und her. Was beobachtet er? Auf dem ersten Schirm DE wird das überwiegend weiße Bild (mit seinem blauen bzw. roten Ende) auf- und abwandern. Doch solange Newton das Prisma nicht zu stark dreht, solange also durchs Loch G im ersten Schirm DE immer noch ein Ausschnitt der weißen Mitte des Bildes hindurchkommt, solange wird sich auf dem zweiten Schirm de immer noch das gesamte Farbspektrum sehen lassen. Es wird dort allerdings im Rhythmus der Prismendrehung ebenfalls auf- und abwandern.

Nun hat Newton auch in den zweiten Schirm ein winziges Loch g gebohrt; es hat denselben Durchmesser wie das Loch G im ersten Schirm. Durch dieses Loch kann immer nur ein winziger Ausschnitt des kompletten Lichtspektrums hindurch, und zwar je nach Drehung des Prismas ABC manchmal der Anteil, den das Prisma besonders stark von seinem Weg abgelenkt hat, manchmal der Anteil mit besonders schwacher und manchmal der mit mittlerer Wegablenkung.

Was geschieht jeweils mit diesen diversen Lichtstrahlen, wenn sie noch einmal durchs Prisma gesandt werden? Um das zu untersuchen, bringt Newton hinter dem zweiten Loch g ein zweites Prisma abc an. Und er fängt das Licht, das durch dies Prisma gelangt, in gewisser Entfernung auf einem Schirm NM auf, um es dort zu betrachten.

Hier die Ergebnisse des raffinierten Experiments: Lichtstrahlen, deren Richtung sich beim Weg durchs erste Prisma am stärksten geändert hat, ändern ihre Richtung beim Weg durchs zweite Prisma abermals am stärksten; und Lichtstrahlen, deren Richtung sich beim Weg durchs erste Prisma am wenigsten geändert hat, ändern ihre Richtung beim Weg durchs zweite Prisma wiederum am wenigsten; genauso für Lichtstrahlen, deren Refrangibilität (beim Weg durchs erste Prisma) zwischen den beiden Extremen gelegen hat.[13]

Was genau beweisen diese Ergebnisse? Und inwiefern beweisen sie, dass weißes Sonnenlicht aus Lichtstrahlen mit unterschiedlichen Brechungseigenschaften besteht? Darüber verliert Newton wenig Worte. Für ihn scheint es sich von selber zu verstehen, dass er mithilfe des Experiments nachgewiesen hat, dass die Brechungseigenschaften – die jewei-

lige Refrangibilität – intrinsisch in den Lichtstrahlen stecken, also nicht von externen Kausalfaktoren abhängen (wie z. B. dem Einfallswinkel der Strahlen ins Prisma).

Und in der Tat, alle Lichtstrahlen, die im Experiment durch das zweite Prisma geschickt werden, genügten beim Eintritt ins zweite Prisma exakt denselben externen Bedingungen: Sie treten an exakt derselben Stelle im exakt gleichen Winkel in exakt dasselbe Prisma ein, und doch verlassen sie das Prisma in unterschiedlichen Richtungen! Da dies nicht auf *externen* Bedingungsänderungen beruhen kann, muss die Ursache für deren unterschiedliche Refraktion in den Lichtstrahlen selber liegen, in ihrem Wesen gleichsam. Und das bedeutet offenbar, dass jedem Lichtstrahl stets ein und dieselbe Refrangibilität innewohnt, ganz gleichgültig, durch wieviele Prismen er geschickt wird. Mithin steckten schon vor der ersten prismatischen Refraktion Lichtstrahlen mit ihrer je eigenen Refrangibilität im weißen Sonnenlicht, Q.E.D.

Wohlgemerkt: Aus Newtons Schriften ergibt sich kein eindeutiger Beweisgang; wir haben eben nur einen der denkbaren Beweise skizziert, die Newton im Auge gehabt haben könnte.[14] Und schon über diese Beweisskizze könnte man ausgiebig streiten. Wir wollen diesen Streit links liegen lassen und stattdessen versuchen, Newtons Experiment zu invertieren.

III. Vorsortiertes Weiß! Vorsortiertes Schwarz? – Schwierigkeiten beim Invertieren

Wir möchten jetzt herausarbeiten, an welcher Stelle derjenige steckenzubleiben droht, der versuchen will, Newtons *experimentum crucis* zu invertieren. Wer die Rollen von Licht und Schatten in Newtons *experimentum crucis* vertauschen will, wird vielleicht zuerst den Fensterladen weit aufreißen und anstelle des Fensterladenlochs F einen gleich großen Schattenwerfer F* anbringen. Gehen wir dieser Idee nach, und lassen wir dabei das Prisma ABC unverändert. *Wenn* wir den Schirm DE weit vom Prisma ABC entfernten und das Loch G in diesem Schirm außer Betracht ließen, dann hätten wir ein Komplement des newtonischen Grundexperiments (siehe Abb. 1, Abb. 2, Abb. 3); wir würden dann auf dem Schirm DE das bereits bekannte komplementäre Spektrum auffangen. (Vergleichen Sie Abb. 7 mit Abb. 3). Jetzt aber schieben wir den Schirm immer

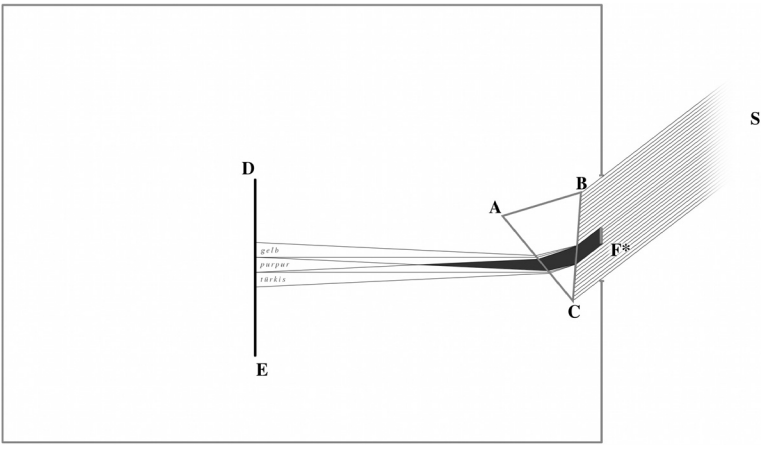


Abb. 7. Versuch einer Invertierung des *experimentum crucis* (Auftakt). Dem breiten Sonnenlichtstrahl, der durchs jetzt weit aufgerissene Fenster in die Dunkelkammer eintritt, setzen wir einen Schattenwerfer F* entgegen, der an die Stelle des Fensterladenlochs F der vorigen Abbildungen tritt. Wenn wir den Auffangschirm DE vom Prisma ABC wegrücken, so taucht auf dem Schirm das Komplementärspektrum auf: oben Gelb, darunter Purpur, unten Türkis. (Die Abstände entsprechen also denen der Grundexperimente, vergl. Abb. 2, Abb. 3).

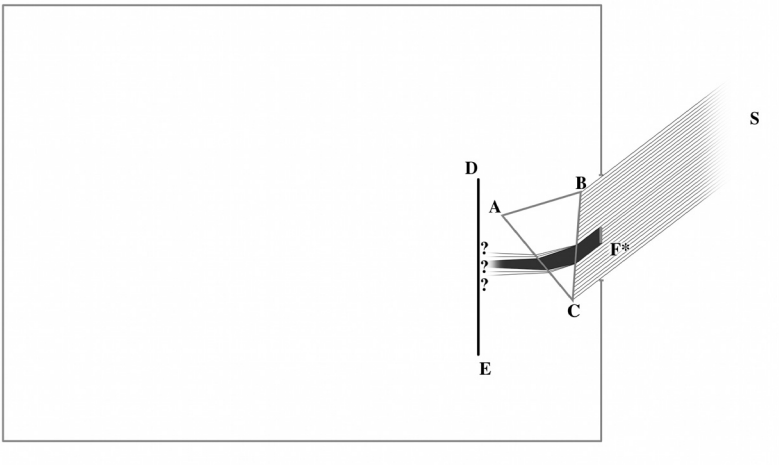


Abb. 8. Versuch einer Invertierung des *experimentum crucis* (Zwischenüberlegung). Anders als in Abb. 7 bringen wir jetzt den Schirm DE näher ans Prisma ABC heran, also dorthin, wo in Newtons *experimentum crucis* die erste Lochblende DGE steht. Was zeigt sich auf dem Schirm?

näher an das Prisma ABC heran. (Siehe Abb. 8). Was erwarten Sie zu sehen? Erwarten Sie das Komplement dessen, was sich in der newtonischen Situation gezeigt hat?

Zur Erinnerung: In der analogen newtonischen Situation (mit Fensterladenloch F anstelle des Schattenwerfers F*) erschien auf dem Schirm ein weißer Lichtfleck, der nur am oberen und unteren Rand farbig gesäumt war. Zwischen dem roten und blauen Ende des Lichtflecks gelangten auf den Schirm alle Lichtstrahlenarten zugleich, vermischten sich und bildeten einen weißen Fleck. Dieses Weiß, das sich unmittelbar hinter dem Prisma auffangen ließ, möchten wir als vorsortiertes Weiß bezeichnen. Es war kein gewöhnliches Weiß. Zwar bestand es aus Lichtstrahlen aller Regenbogenfarben, aber anders als vor deren Eintritt ins Prisma verliefen diese farbigen Strahlen nicht allesamt parallel; vielmehr hatte jede dieser Farbstrahlenarten eine eigene Richtung: Die grünen Lichtstrahlen (mittlerer Refrangibilität) verließen das Prisma allesamt in horizontaler Richtung (und waren nur untereinander parallel); die blauen strebten aufwärts (und waren nur jeweils untereinander parallel), die roten Strahlen strebten abwärts (und waren ebenfalls nur jeweils untereinander parallel). Ausschließlich in diesem vorsortierten Weiß operierte Newton; seine Blendenöffnung G im Schirm DE ließ immer nur irgendwelche Teile des vorsortierten Weiß hindurch.

Man kann mit dieser Blende testen, ob man wirklich in vorsortiertem Weiß operiert oder in unsortiertem Weiß. Wer unsortiertem weißen Licht eine Lochblende entgegenstellt, wird hinter der Lochblende keine Farberscheinungen ausmachen.[15] Anders bei vorsortiertem weißen Licht. Die Blendenöffnung sorgt dafür, dass sich die vorsortierten Lichtstrahlen in unterschiedliche Richtungen aufspreizen, und in gebührendem Abstand von der Blende kann man denn auch sämtliche Farben des newtonischen Regenbogenspektrums auffangen.

Jetzt haben wir genug Material beisammen, um zu erklären, warum der eben anvisierte Versuch der Invertierung nicht funktionieren konnte. Wenn wir nämlich am Fenster die Rollen von Licht und Schatten vertauschen, also das Fensterladenloch F durch einen Schattenwerfer F* ersetzen und dann den Fensterladen aufreißen, wie in Abb. 9 im Anhang angedeutet, so fangen wir zwar unmittelbar hinter dem Prisma ABC das Komplement dessen auf, was wir vorhin beschrieben haben: einen schwarzen Fleck mit farbigen Rändern (am oberen Ende gelblich, am unteren Ende türkis).

Doch lässt sich die schwarze Mitte hinter dem Prisma *nicht* als vorsortiertes Schwarz auffassen, weder gedanklich noch experimentell. Dass es dort kein vorsortiertes Schwarz gibt, zeigt experimentell derselbe Blendentest wie eben: Wer in die schwarze Mitte hinter dem Prisma einen Schirm mit kleiner Blendenöffnung einbringt, kann hinter der Lochblende keine Farben sehen; vorsortiertes Schwarz müsste sich aber (in hinreichendem Abstand von der Blende) in seine verschiedenfarbigen Bestandteile entmischen. Doch nichts davon wird sichtbar.

Warum sich aus dem angeblich vorsortierten Schwarz mittels Lochblende keine Farben (des Komplementärspektrums) hervorlocken lassen, zeigt eine einfache theoretische Überlegung. Sie basiert auf Newtons Theorie und lautet: Die Dunkelheit hinter dem Schattenwerfer F* ist optisch unwirksam; nur das weiße Licht, das am Schattenwerfer vorbeikommt, zieht optische Prozesse nach sich. Über und unter dem Schattenwerfer bewegen sich alle Regenbogenfarbenstrahlen in schöner weißer Eintracht hin zur Prismenoberfläche BC. Die Mitte dieser Fläche bleibt schwarz, aber ihr oberer und unterer Teil wird in gleißend weißes Licht getaucht. Während diese Lichtstrahlenbündel durchs Prisma reisen, werden sie sortiert, und hinter dem Prisma zeigen sich dann *zwei* Bereiche von vorsortiertem Weiß: Der eine liegt oberhalb der schwarzen Mitte, der andere unterhalb. Und dort, wo das obere bzw. untere vorsortierte Weiß an die schwarze Mitte hinter dem Prisma angrenzt, zeigen sich die beobachteten Farbsäume. Sie mögen so wirken, als träten sie aus dem schwarzen Bereich hinter dem Prisma hervor – so, als entmischte sich dort ein angeblich vorsortiertes Schwarz. In Wirklichkeit treten sie aus den zwei weißen (vorsortierten) Nachbarschaften der schwarzen Mitte hervor. Dass es sich so verhält, zeigt der Test mit der Lochblende. Wer die Blendenöffnung in einer der weißen Nachbarschaften der schwarzen Mitte placiert, kann hinter der Blende bunte Farben auffangen – und wer sie in der schwarzen Mitte placiert, sieht keine Farben, nur schwarz.

Wie wir in den nächsten beiden Abschnitten dartun werden, war unser Anlauf für die Invertierung des *experimentum crucis* zu halbherzig. Wir haben bloß ein einziges Element der newtonischen Versuchsanordnung invertiert, und das sozusagen nur *architektonisch*: nämlich die Situation am Fenster der Dunkelkammer. Dort – an der Grenze zwischen dem Innern der Dunkelkammer und ihrer äußeren Umgebung – haben wir ein Fensterladenloch durch einen gleich großen Schattenwerfer ersetzt;

das war nur eine kleine bauliche Veränderung. Wer die *gesamte* newtonische Konstellation invertieren will, muss sich stärker anstrengen. Er muss einerseits die äußere Umgebung der Dunkelkammer in ihr Farbnegativ überführen, andererseits muss er Newtons Dunkelkammer in eine *helle* Kammer verwandeln. Das erste Manöver werden wir im bevorstehenden Abschnitt theoretisch durchspielen, das zweite Manöver im Abschnitt V. (Und wie wir sehen werden, muss dann ausgerechnet an der architektonischen Grenze zwischen Innen und Außen, am Fenster, keine Invertierung vorgenommen werden.) Das Ziel unserer Überlegungen – die vollständige und rein *optische* Invertierung – werden wir im folgenden als *optische Umstülpung* bezeichnen oder kürzer einfach als Umstülpung. Dieser Fachbegriff dient uns zur Abgrenzung von halbherzigen und inkonsequenten (z. B. bloß architektonischen) Invertierungsbemühungen.

IV. Cambridge umstülpen

Analysieren wir als erstes die Situation, die vom Prisma aus gesehen wird. Aus Newtons Dunkelkammer schaut man durch ein ziemlich großes Fensterladenloch in die Cambridger Umgebung. Dominiert wird das gesehene Bild von der gleißenden Sonne. Der Himmel und die Cambridger Umgebung sind im Vergleich zur Sonne geradezu dunkel. (Natürlich bieten auch das Weiß der Wolken und das Himmelsblau keine in sich dunklen Farbeindrücke, aber sie erscheinen im Kontrast zur gleißenden Sonne *vergleichsweise* dunkel. Diese Richtigstellung werden wir nicht jedesmal wiederholen).

Newton hat mit seinem Loch im Fensterladen nur das nachgebaut, was ohnehin tagtäglich zu sehen ist – eine beinahe punktuelle helle Leuchte, umgeben von viel Dunklerem. Fast könnte man den Spieß umdrehen und sagen: Die Sonne wirkt wie ein kosmisches Fensterladenloch. Deshalb blieb die Invertierung, die wir mit dem Schattenwerfer F* versucht haben, auf halbem Wege stehen. Denn dieser Schattenwerfer verdeckt nur in seinem Kernschattenbereich die Sonne, das ist alles.

Um unser Ziel ganz zu erreichen, müssten wir zuallererst die Umgebung (außerhalb der Dunkelkammer Newtons) umstülpen. Wir brauchten eine Umgebung, die wie das Farbnegativ der tatsächlichen Umgebung aussähe: eine strahlend helle Umgebung mit einer tief schwarzen Sonne!

(Die – rein hypothetischen – Beobachtungen, die wir gleich beschreiben werden, entspringen Gedankenexperimenten und werden von Newtons Theorie vorausgesagt; ganz am Ende unseres Textes werden wir ein echtes Experiment aufbieten, mit dem wir demonstrieren wollen, dass unsere Überlegungen zur empirischen Realität passen).

Unter den anvisierten neuen Bedingungen außerhalb der Dunkelkammer lassen wir Newtons Versuchsanordnungen in deren Innern erst einmal unverändert. Was erschiene auf dem Schirm DE, wenn wir ihn weit weg vom Prisma placierten wie in Newtons Grundexperiment? Dort erschiene erst jetzt das *vollkommene* Komplement zum Spektrum aus Newtons Grundexperiment. Selbst die optischen Einflüsse, die etwa das rot leuchtende Nachbardach ursprünglich auf Newtons Spektrum ausgeübt hat und die wir bislang nicht berücksichtigt haben, hätten jetzt ein komplementäres Gegenstück; denn unter den neuen Bedingungen (Umstülpung der gesamten Umgebung der Dunkelkammer) würde ein türkisfarbenes leuchtendes Dach des Nachbarhauses durchs Fensterladenloch in das komplementäre Spektrum hineinfunken.

Wie wird aber der Schirm DE aussehen, wenn wir ihn dicht ans Prisma heranrücken, so wie im *experimentum crucis*? – Er wird keineswegs schwarz sein, sondern er wird ziemlich hell aussehen, sogar heller als in Newtons Experiment. Denn bei Newton kam das Licht fast nur aus Richtung der Sonne; in der umgestülpten Umgebung hingegen kommt das Licht von überall, nur nicht aus Richtung der Sonne. Und weil diesmal insgesamt viel mehr Licht im Spiel ist, wird auch der Schirm heller sein – auch dort, wo bei Newton die weiße Mitte leuchtete und wo wir nun auf eine schwarze Mitte gehofft haben. (Zwar bricht das Prisma dorthin keine Lichtstrahlen, die von der pechschwarzen Sonne ausgehen; aber dieser Lichtmangel in der fraglichen Mitte des Schirms wird bei weitem ausgeglichen von denjenigen Lichtstrahlen, die aus allen anderen Richtungen aufs Prisma treffen und z.T. in die Schirmmitte gebrochen werden. Bedenken Sie, dass diesmal kein nahezu paralleles Sonnenlicht aufs Prisma trifft, sondern sonnenhelles Licht aus allen erdenklichen Richtungen).

Sind wir also beim theoretischen Versuch, vorsortiertes Schwarz zu erzeugen, endgültig gescheitert? Nein; durchdenken wir, was man vom Ort des Schirms DE aus sehen müsste.

Wird das Prisma ABC fortgenommen, so sähe man vom Schirm DE aus die komplementärfarbige Cambridger Umgebung, dominiert von

einer pechschwarzen Sonne in gleißend blendender Umgebung. Nun stellen wir das Prisma ABC wieder an seinen angestammten Ort. Dann erschienen (beim Blick durchs Prisma) in der Himmelsumgebung der pechschwarzen Sonne die komplementären Farben des wirklichen Cambridger Himmels. Und da der ursprüngliche Cambridger Himmel recht dunkel ist, würde die jetzt ins Auge gefasste komplementäre Ansicht von einer (fast) sonnenhellen Umgebung dominiert, an der fast keine Farben entstünden. Nur in der Mitte dieser Ansicht, am optischen Ort der schwarzen Sonne, zeigte sich ein ziemlich dunkles komplementärfarbiges Spektrum. Es müsste offenbar vom Schwarz der umgestülpten Sonne herrühren, denn der sie umgebende helle Himmel wiese keinerlei Neigung zu Farben auf.

Ohne Umstülpung der Umgebung ist beim Blick durchs Prisma (vom Schirm DE aus) im Vergleich zur Sonne eine nahezu schwarze Umgebung zu sehen, die ebenfalls wenig Neigung zur Farbigkeit aufweist. In der Mitte dieser Ansicht ist das Weiß der Sonne hingegen zu einem newtonischen Regenbogenspektrum geworden. Wie ist das möglich? Immerhin erscheint der Ort, von dem aus wir dies sehen – nämlich der Schirm DE – weiß beleuchtet. Unsere Antwort: Das vorsortierte Weiß auf dem Schirm DE ist nur dann gut vorsortiert, wenn *vom Schirm aus* durchs Prisma keinerlei Weiß zu sehen ist, sondern das volle Regenbogenspektrum, umgeben von Dunkelheit. Der Blendentest aus Abschnitt III hat uns dies klargemacht. Denn die Blende DGE samt dahinterliegendem Schirm de funktioniert zusammen wie ein einfach nachgebautes Auge, wie Pupille samt Retina. Auf dem Schirm de erscheint genau das, was man vom Ort G aus durchs Prisma sieht. Kurzum, unser Auge kann im Ergebnis auch ohne Bewaffnung mit Blenden und Schirmen das durchführen, was wir vorhin als Blendentest bezeichnet haben.[16]

Wiederholen wir jetzt solche Blendentests für die umgestülpte Cambridger Umgebung, mit Prisma hinter dem Fensterladenloch! Auf dem Schirm de muss dann genau das erscheinen, was vom Ort G zu sehen wäre – und das ist bis ins feinste Detail das Farbnegativ der nicht umgestülpten Umgebung, dominiert von einem komplementären dunklen Vollspektrum in sonnenheller Umgebung.

Um für diese Situation einen griffigen Namen parat zu haben, werden wir weiterhin von vorsortiertem Schwarz reden, selbst wenn dessen Schwärze nicht buchstäblich zu verstehen ist. Ungeachtet des hellen Farbeindrucks, den sie bieten, werden wir diejenigen Strahlenzonen als

vorsortiertes Schwarz bezeichnen, die beim Blendentest kein newtonisches Farbenspektrum erzeugen, sondern dessen Komplement.

Trotzdem werden Sie fragen: Wenn wir durch die Umstülpung der Umgebung eine solche Zone erzeugt haben, warum erscheint dann der Schirm DE so hell erleuchtet? Warum sehen wir dort nicht Schwarz, das Komplement der weißen Mitte, die wir dort im newtonischen Experiment gesehen haben?

Wir müssen es zugeben – die Mitte des Schirms wird nach der Umstülpung viel heller erscheinen als davor. Sie wird fast so hell gleißen wie die Umgebung außerhalb des Zimmers, sogar fast so hell wie beim direkten Blick in die Sonne. Im Vergleich damit ist das sogenannte vorsortierte Weiß extrem dunkel. Nur: Es war voreilig zu meinen, dass das vorsortierte Weiß aus Newtons Experiment absolut weiß gewesen wäre oder maximal weiß. Also braucht auch dessen Komplement (beim Umstülpen) keineswegs absolut schwarz zu sein oder maximal dunkel. Und wenn, wie sich nun ergeben hat, das vorsortierte „Weiß" dunkler ist als das vorsortierte „Schwarz", so verträgt sich dies überraschende Ergebnis sehr wohl damit, dass das eine das Komplement des anderen ist.

Warum dies weniger verrückt ist, als zunächst scheinen mag, zeigt folgende Überlegung. Je besser eine weiße Strahlenzone vorsortiert ist, desto dunkler erscheint sie auf dem Schirm. Denn um ihre Vorsortierung zu verbessern, müssen wir die zugrundliegende Lichtquelle verkleinern. Umgekehrt steht es beim vorsortierten Schwarz. Es zeigt sich auf dem Schirm in umso hellerer Beleuchtung, je besser es vorsortiert ist.[17] Um das zu begründen, betrachten wir die vorsortierten Zonen als Strahlenzonen im Sinne der newtonischen Theorie. Das vorsortierte Weiß ist eine Strahlenzone, in der jede Strahlenart nur in einer Richtung vorkommt. Wie erreicht man das? Durch Aussortieren *aller anderen* Strahlenarten in dieser Richtung. Die so erzeugte Strahlenzone ist also ziemlich dunkel. In der Zone des vorsortierten Schwarz kommt hingegen jede Strahlenart in genau einer Richtung *nicht* vor, alle anderen Strahlenarten kommen in dieser Richtung vor. Man erreicht dies also durch Aussortieren *nur einer* Strahlenart in der entsprechenden Richtung. Daher ist diese Strahlenzone immer sehr hell, ihrem Namen zum Trotz.

Kurz und gut, wer vorsortierte Strahlenzonen schaffen will, muss immer irgendwelche Strahlenarten aussortieren; vorsortiertes Weiß unterscheidet sich von vorsortiertem Schwarz nur durch Vertauschung von Einsortiertem und Aussortiertem. Damit haben wir aus Newtons

Voraussetzungen abgeleitet, dass beide Vorsortierungen existieren müssen. Aus mengentheoretischer Sicht könnte man sagen, dass sie sich zueinander verhalten wie eine Teilmenge W aus der Gesamtmenge L aller Lichtstrahlenarten aller Richtungen und deren Komplement S = L\W.[18] Der Begriff des vorsortierten Schwarz hat also im Rahmen der newtonischen Orthodoxie guten Sinn. Es bleibt aber zu untersuchen, ob sich solche vorsortierten Strahlenzonen realisieren lassen und ob sie die geforderten Eigenschaften haben. Dafür muss das Experiment befragt werden. Bevor wir auf die Empirie zu sprechen kommen, müssen wir allerdings eine weitere Schwierigkeit aus dem Weg räumen, die nichts mit den optischen Parametern außerhalb der Experimentierkammer zu tun hat, sondern mit den Parametern im Innern der Kammer. Wie wir im nächsten Abschnitt sehen werden, müssen wir die Dunkelkammer optisch umstülpen und in eine Lichtkammer verwandeln.

V. Freunde, baut die Streulichtkammer!

Der Stand der Dinge, den wir im letzten Abschnitt erreicht haben, bleibt unbefriedigend. In beiden Fällen (bei Newton und bei der anvisierten Umstülpung) hatten wir einen Schirm, der durchs Fensterladenloch hell beleuchtet wurde. In der Umgebung der beleuchteten Partien des Schirms war es beidemal dunkel. Sowohl vorsortiertes Weiß als auch vorsortiertes Schwarz erschienen als Helligkeit auf dem Schirm, *umgeben von Dunkelheit*. Und wenn in beiden Fällen eine dunkle Umgebung zu sehen ist, dann kann man den einen Fall kaum als konsequente Umstülpung des anderen ausrufen.

Wer im *experimentum crucis* nur die Fensterläden aufreißt und dann einen Schattenwerfer genau dort installiert, wo ehedem das Fensterladenloch klaffte, der hat, wie wir gesehen haben, das Experiment nicht konsequent genug umgestülpt. Aber hat denn derjenige, der die kosmische Tat vollbringt und ganz Cambridge, ja den Himmel umstülpt, alles getan, was er muss? Bedenken Sie: Die Backen der Lochblenden, mit denen Physiker zu arbeiten pflegen, sind schwarz. Wer die Rollen von Licht und Schatten, von Weiß und Schwarz, von Helligkeit und Dunkelheit konsequent vertauschen will, muss die Blendenbacken weiß erscheinen lassen. Sie müssen also auf jeden Fall weiß angemalt werden, sonst können sie nie und nimmer weiß erscheinen.

Und damit nicht genug. Newton hat sorgsam alle seine (undurchsichtigen) Ausrüstungsgegenstände geschwärzt: schwarz angemalt, mit schwarzem Samt beklebt, usw.[19] Er tat dies, weil er meinte, dass Schwärze und Finsternis soviel bedeuteten wie abwesendes Licht – nur im Finstern ist laut Newton sichergestellt, dass kein störendes Streulicht die Experimente behindert oder verfälscht.[20] Diese newtonische Abwesenheit von Störfaktoren müssen wir vollständig rückgängig machen, wenn wir Newtons *experimentum crucis* umstülpen wollen.

Was ist demzufolge zu tun? Müssen wir auch die Wände der Dunkelkammer weiß anmalen? Ja, aber das genügt nicht. Es genügt nicht, die Dunkelkammerwände weiß anzumalen, denn ohne genug Licht wird sie eine *Dunkel*kammer bleiben. Die Wände der Kammer müssen nicht weiß *sein*, sondern weiß *scheinen* – weiß leuchten.

Wir brauchen eine Streulichtkammer, keine Dunkelkammer. Wie könnte das gehen? Im Abstrakten ist die Sache einfach: An jedem Wandpunkt der Streulichtkammer muss eine Lichtquelle installiert sein, die in alle Richtungen weißes Licht abstrahlt – diffuses Licht. Was bedeutete das konkret? Stellen wir uns Millionen kleiner Glühlämpchen an den Wänden der Streulichtkammer vor – sie müssten sehr klein sein und sich so eng aneinanderdrängen, dass der Eindruck einer zusammenhängenden, gleichmäßig leuchtenden Fläche entsteht. (Physiker reden in solchen Zusammenhängen von Lambertstrahlern, siehe Bergmann et al [O]:673/4). Es gibt bestimmte Folien, die so leuchten – und damit sind wir nahe an einem praktikablen Vorschlag dafür, wie die erforderliche Streulichtkammer realisiert werden könnte. (Wir werden allerdings erst im Abschnitt VII auf echte empirische Resultate eingehen).

Zu Newtons Zeiten gab es weder Glühlampen noch Folien, die diffuses weißes Licht abgeben. Halten wir kurz inne und fragen: Was hätten Newton und seine Zeitgenossen tun können, um die Streulichtkammer zu realisieren, die für die Umstülpung des *experimentum crucis* nötig ist?

Eine Antwort darauf findet sich bei Goethe. Goethe hat viele der hier verhandelten Gedanken vorweggenommen oder jedenfalls Vorahnungen dieser Gedanken kultiviert. In unseren augenblicklichen Zusammenhang passt folgender Ausruf Goethes:

> Freunde, flieht die dunkle Kammer […][21]

In der Tat, wer in die freie Natur ausweicht, der experimentiert – auch – in diffusem Streulicht. Wird derjenige, der Goethes Ratschlag befolgt,

draußen im Freien *immer* geeignete Bedingungen vorfinden, um Newtons *experimentum crucis* umzustülpen? Nein, damit ist nicht zu rechnen. Ob wir im Freien nahe genug an die optischen Bedingungen der anvisierten Streulichtkammer herankommen, hängt vom Wetter ab und von den im Wetter verstreuten Gegenständen. Ideal wäre ein nebelverhangener Schneetag, an dem die Sonne genug Licht spendet, ohne dass sich ihr genauer Ort am Himmel ausmachen ließe. Ideal wäre es, wenn die Lichtverhältnisse so diffus wären, dass der Horizont vage verschwimmt, weil die mattweiß leuchtende Schnee-Ebene dieselbe Helligkeit und denselben Farbmangel aufweist wie der vernebelte Himmel. Es gibt solche Tage.[22]

Wäre Newton Eskimo gewesen, hätte er vielleicht zuerst das umgestülpte *experimentum crucis* ausprobiert und wäre nie auf die Idee gekommen, im diffusen Schwarz seiner Dunkelkammer zu operieren. Bedenken Sie: Eskimos hatten vor Ankunft der Europäer keine Dunkelkammern, in denen sie mit der Sonne hätten experimentieren können. Ihre Iglus ähnelten dem, was wir als Streulichtkammer bezeichnet haben, jedenfalls bei bestimmten Wetterbedingungen. (Soviel zur Entschuldigung für die verspielte Überschrift unseres Aufsatzes).

Bevor wir weitergehen, möchten wir einem Einwand entgegentreten, der sich aufdrängt. Wir haben im vorliegenden Abschnitt dafür plädiert, Newtons *experimentum crucis* konsequenter umzustülpen, als man im ersten Anlauf für nötig halten könnte. Es genügt nicht, die Rollen von Helligkeit und Finsternis nur in den Elementen des Experiments zu vertauschen, die durchs Prisma hindurchgehen; Rollentausche sind auch weiter innen im Experiment nötig, etwa da, wo bei Newton Finsternis herrscht. Und das löst folgenden Einwand aus: Wieso sind uns diese weitergehenden Rollentausche nicht schon vorher begegnet, etwa beim Grundexperiment und seinem komplementären Gegenstück? Müssen wir nicht darauf bestehen, dass beim Umstülpen immer genau dieselben Spielregeln eingehalten werden? – Wir stimmen zu: Jedes Experiment Newtons soll am Ende nach genau denselben Spielregeln umgestülpt werden; das *experimentum crucis* darf keine Sonderbehandlung beanspruchen. Aber das spricht in unseren Augen nicht gegen den jetzt eingeschlagenen anspruchsvollen Weg, auf dem das *experimentum crucis* umgestülpt werden soll. Vielmehr spricht es dagegen, das komplementäre Grundexperiment als *exakte* Umstülpung seines newtonischen Gegenstücks aufzufassen; es bietet nur eine gleichsam *architektonische* Blendeninvertierung.

VI. Newtons Theorie impliziert vorsortiertes Schwarz

Wo stehen wir? In der erdachten Streulichtkammer herrscht viel diffuses weißes Streulicht, das aus allen Richtungen in alle Richtungen strahlt. Wenn die Streulichtkammer leer ist, so ist es darin an jeder Stelle exakt gleich hell, wie man sich leicht klarmachen kann.[23] Wer nun an einer der weiß leuchtenden Wände der Streulichtkammer einen schwarzen Fleck anbringt, verringert die Gesamthelligkeit in der Streulichtkammer und hat mithin gleichsam eine Dunkelheitsquelle geschaffen.[24] In unmittelbarer Nähe des schwarzen Flecks der Streulichtkammer ist es besonders dunkel, und je weiter man sich vom schwarzen Fleck entfernt, desto schwächer wird die bemerkbare Dunkelheit, desto stärker überwiegt der Einfluss des diffusen Streulichts aus allen Richtungen – aus allen Richtungen bis auf *eine*, um genau zu sein, denn vom schwarzen Fleck geht genau kein Streulicht aus. Wie stark also die Dunkelheitsquelle auf einen Beobachter wirkt, hängt vom Abstand zwischen beiden ab – genauso wie bei Lichtquellen. Steigt der Abstand, so sinkt die wahrnehmbare Kraft der Licht- bzw. Dunkelheitsquelle, und zwar in beiden Fällen mit dem Quadrat der Entfernung.[25]

Man könnte sagen, dass der schwarze Fleck Dunkelheit in die Streulichtkammer wirft – aber es ist kein hundertprozentig schwarzer Schatten mit klar umrissenen Grenzen, sondern eine sich graduell ändernde Abschattung oder Verdunklung: dunkel in unmittelbarer Nähe des schwarzen Flecks, heller grau weiter von ihm entfernt. Es handelt sich allerdings nicht um vorsortiertes Schwarz oder Weiß. Denn wer in den verdunkelten Bereich eine Lochblende hineinstellt, wird hinter der Blende keine Farben sehen – weder die Farben des Newtonspektrums noch die seines Komplements.

Jetzt placieren wir ein Prisma in unmittelbarer Nähe des schwarzen Flecks. Wer sich nah hinter das Prisma stellt und durch das Prisma auf den schwarzen Fleck schaut, wird (so sagt Newtons Theorie voraus) den schwarzen Fleck fast ohne merkliche Verfärbung seiner Ränder erblikken. Und was die Retina auffängt, wird (bei geeigneten Lichtverhältnissen) auch ein Schirm hinter einer Blende auffangen können.[26]

An dieser Stelle können wir die Frage nach dem vorsortierten Schwarz aus Abschnitt IV neu aufgreifen. Wir bauen in der erdachten Streulichtkammer das newtonische *experimentum crucis* auf, verwenden aber zunächst die Blende de als Schirm, um mithilfe der Lochblende DGE

den Blendentest zu wiederholen, wie wir ihn für die umgestülpte Cambridger Umgebung in Newtons Dunkelkammer vorgeschlagen hatten. Mithilfe der Streulichtkammer sind wir jetzt beim Umstülpen weiter vorangekommen als vorhin, denn jetzt haben wir auch Newtons Dunkelkammer umgestülpt. Newton hatte an zwei Stellen seiner Gesamtkonstellation eine kleine Quelle der Helligkeit vor vergleichsweise dunklem Hintergrund; einerseits im Außenraum die Sonne, andererseits in seiner Kammer die Fensteröffnung. Er hat konsequent, doppelt, im Dunklen gearbeitet. Wir wollen daher konsequent und doppelt im Hellen experimentieren. Nur so wird es uns gelingen, das *experimentum crucis* vollständig umzustülpen. Also arbeiten wir einerseits mit wenig Dunkelheit vor hellem Himmelshintergrund, andererseits in der Streulichtkammer.

Betrachten wir in Abb. 10 im Anhang den mittleren Fleck p des Auffangschirms de. Was für Strahlen kommen dort an? Unter newtonischer Betrachtung sind das ziemlich viele. Aus jeder Richtung kommen in p Lichtstrahlen aller Regenbogenfarben an – Folge des diffusen Lichts in unserer Kammer.

Aus *jeder* Richtung? Nein, aus *fast* jeder Richtung. Zwar kommen erstens in p von vielen Wänden der Streulichtkammer weiße Strahlen an (also Strahlen aller Regenbogenfarben), aber natürlich nur von den Wänden, die vor dem Auffangschirm de und hinter der Lochblende DGE liegen (also insgesamt von den Lichtkammerwänden, die von p aus sichtbar sind). Und zweitens kommen in p zwar von den weißen Blendenbacken weiße Strahlen an (also Strahlen aller Regenbogenfarben) – aber nicht unbedingt auch aus der Blendenöffnung G. So, wie Auffangschirm, Blende und Prisma angeordnet sind, können auf dem Schirm in p nur ganz bestimmte Lichtstrahlen aus Richtung der Blendenöffnung auftreffen: nämlich solche Strahlen, die zuvor durch die Prismenfläche AC hindurchgereist sind – also Lichtstrahlen, die vom Prisma irgendwie gebrochen wurden und letztlich von der Lichtkammerwand herkommen müssen.[27] Da von dieser Wand der Streulichtkammer – wie von allen ihren anderen Wänden – diffuses Streulicht ausgeht, wird regenbogenbuntes Licht von dort in allen erdenklichen Richtungen durchs Prisma hindurchkommen und dabei in hunderterlei Weise vom Weg abgelenkt werden.

In diesem Durcheinander verliert man schnell die Übersicht. Vielleicht ist es instruktiver, zu fragen, auf welchen Pfaden genau *keine* Lichtstrah-

len irgendeiner Farbe durchs Prisma kommen. Abb. 11 im Anhang zeigt z.B. die grünfreien Pfade, sie sind durch Ketten grüner Minuszeichen angedeutet („Minus" für *abwesendes* Grün). Ein Teil dieser grünfreien Pfade endet an den vorderen Backen der Blende DGE. An der weißen Rückseite dieser Backen setzen sich die grünfreien Pfade deshalb nicht fort, weil diese Rückseiten weiß angemalt und weiß beleuchtet sind, also auch in der fraglichen Richtung grüne Strahlen abgeben. (Man beachte, dass es zur Unterbrechung grünfreier Pfade nicht genügt, ihnen ein Hindernis in den Weg zu stellen. Vielmehr muss die Rückseite des Hindernisses in geeigneter Weise beleuchtet sein. Wäre z.B. die Rückseite des Hindernisses schwarz oder läge sie im Dunkeln, so ginge der grünfreie Pfad durch das Hindernis hindurch!)

Aber bestimmte grünfreie Pfade setzen sich in den Raum hinter der Blende fort, wie der Abb. 11 zu entnehmen ist; nämlich diejenigen grünfreien Pfade, die sich schnurstracks durch die Blendenöffnung hindurch verlängern lassen. Dass diese grünfreien Pfade auch hinter der Blende weiterlaufen, liegt letztlich daran, dass in der Blendenöffnung genau keine weiße opake Oberfläche installiert ist, die weißes Streulicht reflektieren könnte, also auch grüne Lichtstrahlen in derjenigen Richtung abgäbe, die uns interessiert.

Analog für blaufreie Pfade in Abb. 12 und für rotfreie Pfade in Abb. 13. In Abb. 14 (alle im Anhang) sehen Sie die drei vorigen Abbildungen übereinander. Dadurch wird deutlich, dass die drei betrachteten Arten grün-, blau- und rotfreier Pfade nicht exakt in derselben Richtung verlaufen. Wer also einen Schirm weit genug hinter der Blende vor- und zurückschiebt, der wird einen idealen Abstand entdecken können, an dem sich die blau- und rotfreien Pfade so weit vom grünfreien Pfad entfernt haben, dass auf dem Schirm eine hinreichend große grünfreie Zone p aufreißt. (Genauer gesagt, eine Zone, in der exakt aus Richtung der Blendenöffnung keinerlei grüne Lichtstrahlen ankommen; aus allen anderen Richtungen kommt dort allerlei diffuses weißes Licht an, also auch allerlei grünes Licht).

Welchen Farbeindruck erwarten wir, wenn wir auf einen Fleck starren, auf den aus allen Richtungen diffus schwaches Licht aller Farben eintrifft – und aus einer einzigen prominenten Richtung Licht aller Farben *mit Ausnahme des grünen Lichts*?

Grünes Licht ist an dieser Stelle des Auffangschirms *unterrepräsentiert*. Ein regenbogenbuntes Lichtgemisch mit deutlicher Unterrepräsen-

tation einer Lichtsorte bietet den farblichen Eindruck des Komplements der unterrepräsentierten Lichtsorte, also in unserem Fall den Eindruck von Purpur.

Genauso macht man sich klar, warum über dem purpurnen Fleck p auf dem Schirm ein gelber Farbfleck sichtbar wird und unter p ein blauer Farbfleck. Die Blende DGE holt also aus der Mitte der Strahlenzone hinter dem Prisma ein Komplementärspektrum heraus. Und das bedeutet, dass diese Zone das bietet, was wir vorhin vorsortiertes Schwarz genannt haben.

Schauen wir vom Prisma ABC aus auf den Schirm DGE, dann sehen wir dort – sagt Newtons Theorie – die äußeren Randbereiche hell erleuchtet, denn sie werden von überall gleichmäßig bestrahlt. In der Mitte des Schirms, dort wo sich die Blendenöffnung G befindet, ist es etwas dunkler, da sich in diesem Bereich die Abdunkelung der schwarzen Fläche S* fortsetzt. Es kommen dort also weniger Lichtstrahlen an als an allen anderen Orten auf dem Schirm, und daher ist dieser Bereich dunkler als die anderen Schirmorte. Er wird begrenzt von schwach farbigen Säumen, die den Komplementärfarben der Säume bei Newtons vorsortiertem Weiß entsprechen. In der Streulichtkammer erscheint jetzt also auch der Bereich des vorsortierten Schwarz dunkler als die gesamte homogen helle Umgebung. Mithin ist es gerechtfertigt, diesen Bereich als vorsortiertes Schwarz zu bezeichnen. Dass es dort nicht ganz schwarz sein kann, haben wir uns in Abschnitt IV klargemacht. (Überlegen wir uns noch im Vorübergehen, wie die Strahlenzone des vorsortierten Weiß in der Streulichtkammer aussehen wird. Wir hatten bemerkt, dass sie erheblich dunkler erscheinen muss als das vorsortierte Schwarz. Im Vergleich mit der jetzt überall homogenen Helle wirkt sie geradezu schwarz, zeigt aber die farbigen Ränder wie in Newtons Dunkelkammer, nur diesmal umrandet von Hellem).

Wo stehen wir? Wir haben unser Zwischenziel erreicht und endlich mit dem vorsortierten Schwarz in der Streulichtkammer ein echtes Pendant zum vorsortierten Weiß in Newtons Dunkelkammer erreicht, auf dem theoretischen Boden der newtonischen Optik.

Es ist wichtig, sich klarzumachen, dass unser theoretischer Existenznachweis entscheidend von der optischen Gesamtkonstellation abhängt und nur dort triftig sein kann, wo diffuses weißes Licht herrscht, das u.a. von den weißen Backen der Blende DGE zurückgeworfen wird.[28] Wäre die Blende schwarz oder läge ihre Rückseite im Dunkeln, so ginge der

Test anders aus. Aber das ist bei vorsortiertem Weiß genauso, auch hier kommt es auf die optische Umgebung des Tests an. Er funktioniert im Finstern (bei schwarzen Blendenbacken in Dunkelkammern). Aber vermöge einer Blende mit weißen Blendenbacken in der Streulichtkammer lässt sich vorsortiertes Weiß nicht nachweisen. Weiße Hindernisse, die von hinten diffus beleuchtet sind, eignen sich nicht zur effektiven Unterbrechung von Lichtpfaden.

Wenn die bisherigen Betrachtungen ins Schwarze trafen, dann impliziert Newtons Theorie: Das *experimentum crucis* lässt sich zumindest bis zu dem Punkt optisch umstülpen, an dem vorsortiertes Schwarz per Blende DGE in ein komplementäres Spektrum entmischt wird, das auf einem Schirm de aufgefangen werden kann.

Wie steht es mit dem Rest des *experimentum crucis*? Lässt es sich ebenfalls umstülpen? Bevor wir auf diese Frage zurückkommen, ist es vielleicht an der Zeit, genauer zu erklären, an welchen Regeln man sich beim optischen Umstülpen orientieren muss. Wir möchten diese Regeln in erstens geometrische Regeln und zweitens Hell-Dunkel-Regeln einteilen. Sie fordern ein extremes Ideal, taugen also nicht unbedingt für die Praxis.

Zuerst zur Geometrie: Jede Grenzfläche eines fürs umzustülpende Experiment relevanten Körpers (etwa eines Prismas oder einer Kammerwand oder einer Blende) muss auf eine kongruente Körpergrenzfläche des umgestülpten Experiments passen, und auch die geometrischen Gesamtkonfigurationen aller dieser Grenzflächen aus den beiden Experimenten müssen zueinander kongruent sein. Kommt Ihnen der Ausdruck „relevant" zu vage vor? Keine Sorge. Sollte Streit darüber aufkommen, ob ein Körper fürs Experiment relevant sei oder nicht, dann wollen wir ihn sicherheitshalber zu den relevanten Körpern zählen. – Nebenbei möchten wir darauf aufmerksam machen, dass unsere geometrischen Regeln nicht verlangen, Lochblenden in Schattenwerfer zu verwandeln. Im Gegenteil; wo im ursprünglichen Experiment ein Loch klafft, muss auch beim Umstülpen ein Loch gleicher Größe vorgesehen sein. (Die Architektur des Experiments wird also beim Umstülpen genau nicht invertiert).

Zweitens zur Verteilung der Hell-Dunkel-Anteile in den Experimenten. Zur Vereinfachung wollen wir annehmen, dass alle relevanten Lichtquellen des umzustülpenden Experiments weißes Licht liefern. (Das verhält sich so in Newtons *experimentum crucis*, jedenfalls dann, wenn wir

die optischen Effekte des blauen Himmels, der Sterne mit Rotverschiebung, der Cambridger roten Dachziegel usw. aus dem Spiel lassen). Wir markieren alle noch so kleinen Körperoberflächen, an denen Lichtquellen installiert sind, und sorgen ebendort im umgestülpten Experiment für Finsternis. Diejenigen Körperoberflächen, an denen es im umzustülpenden Experiment finster war, statten wir hingegen mit Lichtquellen aus, die weißes Streulicht liefern. (Sollten im umzustülpenden Experiment farbige Lichtquellen vorkommen, so müssten für dessen Komplement an Ort und Stelle Lichtquellen installiert werden, die alle Lichtsorten emittieren *mit Ausnahme* des fraglichen farbigen Lichts).

Wenn wir alles richtig gemacht haben, dann gibt es zu jedem *weißen* Lichtstrahl, der irgendeine Grenzfläche des umzustülpenden Experiments mit einer anderen der involvierten Grenzflächen verbindet, eine *lichtfreie* Verbindungslinie zwischen den kongruenten Grenzflächen aus dem umgestülpten Experiment – und umgekehrt. Aus alledem ergibt sich: Wer die Umstülpung irgendeines Experiments abermals umstülpt, kommt dadurch zum ursprünglichen Experiment zurück. Welcher der beiden Fälle als optische Umstülpung bezeichnet wird, hängt einfach davon ab, womit man begonnen hat. Wäre Newton Eskimo gewesen und hätte er im Iglu experimentiert, so hätten wir in unserem Aufsatz Dunkelkammern, keine Streulichtkammern ins Spiel gebracht, um die grönländischen Experimente Newtons umzustülpen.

VII. Fortgesetzt umgestülpt in Theorie und Empirie: Vollständiges Farbnegativ des *experimentum crucis*

Um die Fortsetzung des newtonischen *experimentum crucis* umzustülpen, hantieren wir mit den Strahlen, die wir im letzten Abschnitt gedanklich aus dem vorsortierten Schwarz hervorgelockt haben.[29] Wir bohren in den Schirm de ein Loch g, das denselben Durchmesser hat wie das Loch G der Blende DGE. Damit haben wir eine zweite Lochblende ins Spiel gebracht; ihre Backen sollen wieder weiß sein. Hinter der zweiten Blende bringen wir ein Prisma abc an, und in gebührendem Abstand hinter diesem zweiten Prisma placieren wir einen Auffangschirm NM, siehe Abb. 15 im Anhang. Was sagt Newtons Theorie für diese Konstellation voraus?

Nehmen wir an, wir hätten das Loch g dort in den Schirm de gebohrt,

wo wir vorher den purpurnen Fleck p des komplementären Spektrums aufgefangen haben. (Dieser Farbfleck lag in der Mitte jenes Komplementärspektrums).

Jetzt verlängern wir in Gedanken diejenige Gerade, die von der Rückseite AC des ersten Prismas durch die Blendenlöcher G und g zur Vorderseite bc des zweiten Prismas führt. Bis zum Punkt g liegt auf jener Geraden jedenfalls das, was wir einen grünfreien Pfad genannt haben; auf diesem Pfad verbreitet sich allerlei Licht, jedoch genau kein grünes Licht. Auch hinter der Blendenöffnung g setzt sich der grünfreie Pfad in gerader Linie fort und trifft im Punkt p aufs Prisma abc.

Die ungrünen Lichtstrahlen, die demzufolge in p aufs Prisma treffen, werden beim Weg durchs Prisma z. T. stärker abgelenkt (blau), z. T. weniger stark (rot), siehe Abb. 15 im Anhang.

Und da auf dem grünfreien Pfad naturgemäß keine grünen Lichtstrahlen (denen mittelgroße Refrangibilität zukommt) reisen, bleibt beim Weg durchs Prisma abc ein mittlerer Pfad grünfrei, der in Abb. 16 im Anhang wieder durch grüne Minuszeichen angedeutet ist.

Es gibt also auf dem Auffangschirm einen Ort π, an dem (aus Richtung der Geraden Gg) keine grünen Lichtstrahlen ankommen. Kommen dort andersfarbige Lichtstrahlen an? Allerdings. Von den diffus leuchtenden weißen Backen der zweiten Blende dge kommt allerlei Streulicht aller Regenbogenfarben, das ebenfalls auf dem Weg durchs Prisma abgelenkt wird; in der Abb. 15 haben wir einen roten und einen blauen dieser Strahlen eingezeichnet.

Man beachte, dass von den weißen Blendenbacken keinerlei grünes Licht durch den Punkt p des Prismas zum Punkt π auf den Schirm gelangen kann; denn solches grünes Licht (mittlerer Refrangibilität) müsste zuvor exakt die beiden Blendenlöcher g und G passiert haben. Und wir haben uns klargemacht, dass diese beiden Blendenlöcher auf einem grünfreien Pfad liegen. Alles das bedeutet, dass im Punkt π allerlei Licht aller Regenbogenfarben auftreffen muss – mit deutlicher Unterrepräsentation grünen Lichts aus besonders prominenter Richtung. Als Farbeindruck resultiert wieder das Komplement der unterrepräsentierten grünen Farbe: purpur.

Welchen Farbeindruck erwarten wir in der Umgebung des purpurnen Farbflecks π? Einen weißen Farbeindruck. Denn dort kommen Lichtstrahlen aller Regenbogenfarben an, auch grüne Lichtstrahlen.

Betrachten wir etwa einen Fleck β knapp oberhalb von π. Dort kommt

erstens blaues Licht aus der Blendenöffnung an, zweitens grünes Licht aus dem Streulicht von oberhalb der Blendenöffnung und drittens rotes Licht von noch weiter oben. Der entscheidende Unterschied zwischen p und β liegt darin, dass in p eine Lichtsorte aus Richtung der Blendenöffnung fehlt (grün), während in β keine Lichtsorte aus dieser Richtung fehlt; dort ist keine Lichtsorte unterrepräsentiert. (Analog für einen weißen Fleck α unterhalb von π).

Alles das bedeutet: Wer den purpurnen Teil p des Komplementärspektrums (der sich durch Refraktion im Prisma ABC hinter einer Blende DGE auf einem Schirm de zeigt) mittels einer Blendenöffnung g aussondert und dann abermals durchs Prisma abc schickt, der wird am Ende einen purpurnen Fleck vor weißem Hintergrund auffangen; das Purpur zerlegt sich also nicht weiter in seine Bestandteile.[30]

Jetzt folgen wir Newtons Beispiel und bringen Bewegung ins Experiment. Newton hat in seinem *experimentum crucis* das erste Prisma ABC um seine Achse gedreht; im komplementären Experiment tun wir dasselbe. Drehen wir also das Prisma ABC so, dass nicht der purpurne Teil des Spektrums durch die Blendenöffnung g fällt, sondern dessen gelber Teil: Das ist der obere Teil des komplementären Spektrums, der aus allem Licht besteht mit Ausnahme blauer Lichtstrahlen (aus Richtung der Blendenöffnung G). Verfolgen wir den zugehörigen blaulosen Pfad auf seinem weiteren Weg durch das Experiment. In Abb. 17 im Anhang ist dieser Pfad mithilfe blauer Minuszeichen angedeutet. Das Prisma abc spaltet die grünen und roten Lichtstrahlen auf, die auf dem blaufreien Pfad durch die beiden Blendenöffnungen G und g hindurchkommen. Alle diese Lichtstrahlen sind weniger refrangibel als blaues Licht, das genau auf diesem Pfad fehlt. Und das bedeutet, dass uns auf dem Schirm NM besonders weit oben ein Farbeffekt ins Auge springen wird – nämlich die Unterrepräsentation blauer Lichtstrahlen. Wir sehen dort einen gelben Fleck, die Komplementärfarbe von Blau. Dass die Umgebung dieses gelben Flecks weiß sein wird, kann man sich wieder so klarmachen wie bei der weißen Umgebung des purpurnen Flecks π (die sich bei der ursprünglichen Ausrichtung des Prismas ABC zeigte).

Für uns wichtig ist nicht die Farbe des aufgefangenen Flecks, sondern sein Ort – etwas, das auch von Farbenblinden registriert werden kann. Er liegt *über* dem Ort π, an dem wir vorher die purpurne *Mitte* des Komplementärspektrums aufgefangen haben. Und da sich gelb auch im ursprünglichen Komplementärspektrum über der purpurnen Mitte fin-

det, hat sich an der räumlichen Beziehung zwischen den beiden Farben durch nochmalige Refraktion nichts geändert.

Allgemeiner gesagt: Die relativen Positionen bestimmter Teile des komplementären Spektrums bleiben bei mehrfacher Refraktion *in der Streulichtkammer* erhalten. Wir haben diese Behauptung zwar nur fürs räumliche Verhältnis zwischen (purpurner) Mitte und oberem (gelben) Extrempol des Komplementärspektrums gezeigt, aber man macht sich schnell klar, dass sich unser Argument leicht auf andere Teile des Komplementärspektrums übertragen lässt, etwa auf dessen unteren (türkisfarbenen) Extrempol.

Mit der allgemeinen Behauptung, die wir eben aufgestellt haben, sind wir am Ziel angekommen. Wir haben auf der Grundlage der newtonischen Theorie nachgewiesen, dass sich Newtons *experimentum crucis* optisch umstülpen lässt. Dort, wo in Newtons *experimentum crucis* Licht vorkommt, herrscht im umgestülpten Experiment Finsternis; und umgekehrt. Mehr noch, alle Farben in Newtons Experiment haben im umgestülpten Gegenexperiment eine komplementäre Entsprechung. Dort, wo im newtonischen Experiment z. B. mittelstark refrangible – grüne – Lichtstrahlen vorkommen, zeigen sich im komplementären Experiment mittelstark refrangible – purpurne – Schattenstrahlen. Genauso für die stark refrangiblen blauen Strahlen Newtons, denen im umgestülpten Experiment stark refrangible gelbe Strahlen entsprechen, und so weiter für alle anderen Strahlenarten. Und diese Komplemente zeigen im *gesamten* umgestülpten Experiment dasselbe Verhalten wie ihre Gegenstücke bei Newton; während bei Newton grüne Lichtstrahlen an *beiden* Prismen mittelstark gebrochen werden, zeigen die purpurnen Strahlen im umgestülpten Experiment ebenfalls an beiden Prismen mittelstarke Brechung. Wenn (laut Newton) den homogenen grünen Strahlen bei jeder Brechung stets ein und dieselbe Refrangibilität zukommt, dann gilt dasselbe für die purpurnen Strahlen.

In einer ersten Version können wir sogar eine experimentelle Realisierung vorweisen. Als Streulichtkammer wurde ein Plexiglasgehäuse verwendet, das mit Blumenseide beklebt war. Es muss von außen mit möglichst vielen weißen Lampen beleuchtet werden, und zwar möglichst einheitlich; dann entsteht im Innern ein gleichmäßig heller Raum. Darin wurde ein verkleinertes *experimentum crucis* aufgebaut. Auf dem Boden des Gehäuses lag ein länglicher schwarzer Papierstreifen, der als Komplement der Sonne verwendet wurde und der von der Helligkeit

der Streulichtkammerwand begrenzt erschien. Aufgrund der Nichtlinearität unseres Hellempfindens ergab die Betrachtung des Schirmbilds NM aber keine befriedigende Ansicht. Wird der Schirm durch eine Kamera ersetzt, so kann das umgestülpte *experimentum crucis* gut dokumentiert werden. Die Ergebnisse sind in Abb. 18 im Anhang gezeigt.

Werden die Lampen ausgeschaltet, die unsere Streulichtkammer beleuchten, so verwandelt sie sich in eine Dunkelkammer. Jetzt muss nur noch der schwarze Papierstreifen durch einen gleich dimensionierten leuchtenden Streifen ersetzt werden, so haben wir ein miniaturisiertes *experimentum crucis* nach Newtons Vorbild. Es zeigt die erwarteten Ergebnisse, die bei sonst vollkommen gleichen Bedingungen mit der Kamera aufgenommen wurden und ebenfalls in Abb. 18 gezeigt werden. Der Vergleich der beiden Bildreihen bestätigt unsere theoretische Untersuchung befriedigend.[31]

Kurz und gut, das neue Experiment in der Streulichtkammer sieht aus wie ein vollständiges Farbnegativ des *experimentum crucis* in der Dunkelkammer, und zwar einerseits in all seinen dynamischen und statischen Aspekten (also mit und ohne Rotation des ersten Prismas), andererseits in all seinen farblichen und geometrischen Aspekten. Newtons wichtigstes Experiment lässt sich optisch umstülpen.

Dies Ergebnis halten wir für bemerkenswert; und Newton hätte darüber gestaunt. Welche weiterführenden Überlegungen könnten Wissenschaftsphilosophen und Physiker daran anknüpfen? Dazu zunächst eine philosophische Andeutung. Unserer Ansicht nach zerstört Newtons Theorie (ebenso wie die optische Empirie, die wir zum Schluss beschrieben haben) die erkenntnistheoretischen Ansprüche, die Newton mit ebendieser Theorie verbunden hat. Damit hätten wir eine *reductio ad absurdum* der Gesamtposition Newtons (die einerseits seine optische Theorie enthält und andererseits erkenntnistheoretischen Optimismus mit Blick auf die Beweisbarkeit dieser optischen Theorie). Aber diese *reductio* bietet nicht alleine negative Konsequenzen; vielmehr liefert sie ein positives Beispiel für die umstrittene These von der Unterbestimmtheit der Theorie durch *ihre* Daten; solche Beispiele sind bislang rar.[32]

Zum Abschluss deuten wir einen weiterführenden Gedanken für die Physik an: Die Symmetrie, die wir in diesem Artikel durch Umstülpung des *experimentum crucis* herausgearbeitet haben, legt es geradezu nahe, die beiden – perfekt symmetrischen – Fälle auch symmetrisch zu erklären. Lichtstrahlenmodelle à la Newton tun dies nicht und sind (im

Falle des umgestülpten Experiments) alles andere als komfortabel, wie wir unfreiwillig vorgeführt haben. Aber die physikalischen Verhältnisse erfordern offenbar keine asymmetrische Erklärung der vorgestellten Experimente. Denn in der Streulichtkammer können mit den hier vorgestellten Umstülpungsregeln alle newtonischen Experimente umgestülpt werden (sogar, wie man sich leicht klarmacht, *alle denkbaren* spektroskopischen Experimente in der Dunkelkammer).

In der Tat, bei der Theoriebildung pflegen Physiker aktiv nach Symmetrien zu suchen – selbst dort, wo die Phänomene dies in viel geringerem Maße nahelegen.[33] Die Ergebnisse unserer Untersuchung zeigen, dass es möglich sein müsste, die optischen Phänomene in einer Theorie zu fassen, deren Symmetrie die Symmetrie der Phänomene abbildet. Newtons Theorie und ihre modernen Nachkommen tun dies nicht. Von zwei geometrisch und optisch äquivalenten Experimentalsituationen zeichnen sie *eine* als primäre aus und leiten die andere daraus ab. Aber nichts spricht dafür, dass die umgestülpte Situation sekundär ist![34]

Anmerkungen

1 Newton beschreibt das Experiment sehr knapp in Newton [NTaL]:3078-3079. Das Experiment taucht ausführlicher auch in den *Opticks* auf, aber ohne den Namen „experimentum crucis" oder dessen englische Übersetzung, siehe Newton [O]:31-33 (= Book I, Part I, Proposition II, Experiment 6).

2 Siehe z. B. Sabra [ToLf]:231 ff, Westfall [NDHF], Westfall [NHCo], Westfall [NRtH], Lohne [EC], Gruner [DFL], Shapiro [GAoN], Lampert [NvG].

3 Siehe Newton [NTaL]:3079. In den *Opticks* kommt diese Behauptung ebenfalls vor, siehe Newton [O]:21 (= Book I, Part I, Proposition II).

4 Diese zweite Behauptung steht in Newton [NTaL]:3083, Punkt 7. In den *Opticks* findet sie sich nicht als Proposition mit eigener Nummer, sie folgt aber logisch aus der ersten Behauptung und aus dem ersten Teilsatz der Proposition II des zweiten Teils des ersten Buchs der *Opticks* (siehe Newton [O]:78).

5 Streng genommen bietet dieses Experiment keine perfekte Umstülpung des Grundexperiments; es genügt nicht den Spielregeln, die wir am Ende des Abschnitts VI explizieren werden. Denn Blenden werden bei echten Umstülpungen genau nicht in ihr Gegenteil (einen Steg) verwandelt.

6 Siehe Goethe [EF]:§215, Goethe [ETN]:§132, Goethe [EzGF]:68. Siehe auch Kirschmann [USSK]:197 ff, Kirschmann [USS], Kirschmann [USSF], Bjerke [NBzG], Müller [GPUb], Nussbaumer [zF]:177/8, Rang / Grebe-Ellis [KS]. Bjerke, Kirschmann, Nussbaumer, Rang und Grebe-Ellis brin-

gen objektive Versionen invertierter Experimente (in denen das Spektrum auf einen Schirm geworfen wird); Goethe bringt teils objektive, teils subjektive Versionen (in denen anstelle des Auffangschirms die menschliche Netzhaut steht; mehr zum Unterschied zwischen subjektiven und objektiven Experimenten in Fußnote 16). Die frühesten beiden invertierten Versionen prismatischer Experimente (die Newton gekannt hat) finden sich in zwei Briefen des Jesuiten Lucas. Beide Versionen sind subjektive Experimente. Die erste Version findet sich in einem Brief vom 17.5.1776 (datiert auf den 27.5.1776), siehe Newton [CoIN]/II:8-12, dort Punkt 7 auf p. 11. Der Brief ging über Oldenburg an Newton und wurde von Newton in mehreren Briefen beantwortet (vergl. Newton [CoIN]/II:8), allerdings erwähnt Newton den Punkt 7 in seinen Antworten nicht. Die zweite Version findet sich in einem Lucas-Brief vom Februar 1677/8 (siehe Newton [CoIN]/II:246-251), und zwar dort auf p. 249. Der Brief ging über Hooke an Newton und wurde von Newton am 5.3.1677/8 beantwortet (vergl. Newton [CoIN]/II:254-260). Newton reagiert dort auf diese zweite Version des invertierten Experiments, siehe p. 257. In den *Opticks* bringt und erklärt Newton dasselbe (subjektive) Experiment, siehe Newton [O]:104 (= Book I, Part II, Proposition VIII, Problem III).

7 Die Frage liegt seit langer Zeit in der Luft. Vor einem halben Jahrhundert hat der norwegische Physiker Torger Holtsmark zusammen mit dem norwegischen Dichter und Publizisten André Bjerke in einem Studienkreis Invertierungen der meisten newtonischen Experimente angesteuert; sie sind in einem brillanten Buch Bjerkes wiedergegeben; siehe Bjerke [NBzG]. Die Invertierung des *experimentum crucis* wird dort allerdings nicht ausführlich genug diskutiert, um es anderen Arbeitsgruppen zu ermöglichen, die Sache zu replizieren (vergl. Bjerke [NBzG]:87). Ein Jahrzehnt später hat Holtsmark einen eigenen Anlauf genau zum *experimentum crucis* veröffentlicht und Vorschläge dafür unterbreitet, wie man dort die Rollen von Licht und Schatten vertauschen müsste (siehe Holtsmark [NECR]). Ob er dieses Experiment wirklich durchgeführt hat, geht aus dem Aufsatz nicht hervor. Holtsmarks schwedischer Kollege Pehr Sällström hat dann die fraglichen Experimente durchgeführt und in einem spektakulären Experimentalfilm dokumentiert, der erst kürzlich fertiggestellt worden ist; siehe Sällström [MS]. (Diese Experimente haben gewisse Nachteile; sie erfüllen nicht alle Wünsche, die man beim Invertieren hegen sollte. Siehe Fußnote 29). – Abgesehen davon haben sich deutsche Physiker aus der Tradition der phänomenologischen Optik an die Invertierung des newtonischen *experimentum crucis* herangetastet (Grebe-Ellis [NECa], Rang [MS]). Zudem hat der Wiener Maler und Farbexperimentator Ingo Nussbaumer eine Reihe aufsehenerregender Experimente durchgeführt und veröffentlicht, die in eine ähnliche Richtung weisen, also z.T. ebenfalls darauf abzielen, Newtons Experimente zu invertieren. Mehr dazu unten in Fußnote 29.

8 Das newtonische Original findet sich dort als *figure 18* auf der Tafel „LIB. I.TAB.IV. Par.I." und gehört zum sechsten Experiment des ersten Teils des ersten Buchs von Newtons *Opticks*. Eine ältere Skizze desselben Experiments stammt aus Newtons zweitem Brief an Pardies vom 9.7.1672; (siehe

Newton [INPL]:101; die Abbildung entnimmt der Herausgeber Cohen den *Philosophical Transactions* 85 (1672)).

9 Dass die breite Mitte des Bildes auf dem Schirm DE weiß ist, erwähnt Newton weder in seiner Darstellung des *experimentum crucis* aus dem Jahr 1672 noch in der Parallelpassage aus den *Opticks*. – Dies hat viele seiner Leser in die Irre geführt, wie wir in der nächsten Fußnote anhand eines prominenten Newton-Kenners dartun werden.

10 Selbst dem sonst so gewissenhaften Shapiro unterläuft an dieser Stelle ein grober Fehler. Er schreibt: „The hole […] in the first board allowed only a small portion of the *spectrum* cast by the prism placed in front of it to be transmitted, but when this prism was rotated different portions would be transmitted" (Shapiro [GAoN]:73/4; unser Kursivdruck). Denselben Fehler macht Shapiro auch an anderer Stelle (siehe [ESoN]:212). Um den Fehler namhaft zu machen, möchten wir zunächst daran erinnern, dass man auf der *dem Fensterladenloch zugewandten Seite* des ersten Schirms DE noch kein Spektrum sehen kann (nur ein weißes Bild mit farbigen Enden); denn der Schirm DE mit dem Loch G steht viel zu nah am Prisma ABC, um das Licht dort schon spektral aufzufächern. Anders steht es, wenn man *von der Rückseite* des Schirms DE durchs Loch G in Richtung des Fensterladenlochs und des Prismas ABC schaut. Wer diese Blickrichtung wählt und sich unmittelbar hinter dem Loch G placiert, wird durch dieses Loch das volle Newtonspektrum erblicken – und das auch dann, wenn das Prima ABC um seine Achse gedreht wird; anders als Shapiro nahelegt, werden durch diese Rotation also durchaus keine unterschiedlichen Portionen des Spektrums durchs Prisma geschickt. Wer diese Blickrichtung dagegen beibehält, sich vom Loch G entfernt und kurz vor dem Schirm de stehen bleibt, der wird dann in der Tat nur jeweils verschiedene Portionen des Spektrums sehen, je nach Rotation des Prismas ABC.

11 Auch im *experimentum crucis* schickt Newton das Licht zuallererst durch ein Fensterladenloch F, siehe Abb. 4. Das Loch darf aber so groß sein, dass es das gesamte Prisma ausleuchtet. In den *Opticks* spricht Newton daher auch von einem sehr breiten Loch („in the window-shut a much broader hole", siehe Newton [O]:31 (= Book I, Part I, Proposition II, Experiment 6)). In Newtons älteren Abbildungen fehlen Loch und Fensterladen; dort sieht es so aus, als käme das Licht fürs Experiment geradewegs und ungehindert von der Sonne. Daher darf man vermuten, dass Fensterladen und Fensterladenloch für Newtons *experimentum crucis* nicht wesentlich sind. Wozu sind sie überhaupt da? Offenbar kann das Experiment nicht gut im Freien stattfinden (weil es sonst von Licht aus allen Richtungen gestört würde); die Aufgabe des Fensterladenlochs besteht also nur darin, überhaupt Licht in die Dunkelkammer hineinzulassen, und zwar grob in ein und derselben Richtung. Ganz anders im Grundexperiment aus Abb. 1 und Abb. 2; dort bietet das Loch einen entscheidenden Parameter des Experiments. In Newtons ursprünglicher Präsentation aus dem Jahr 1672 kommt nicht deutlich heraus, welche unterschiedliche Rolle und Größe das Fensterladenloch in den beiden Experimenten hat; dort kann das *experimentum crucis* wie eine Fortsetzung des Grundexperiments wirken, mit identischem Lochdurch-

messer. Beim Grundexperiment sagt Newton: „[...] and made a small hole in my window-*shuts*" (Newton [NTaL]:3076, unser Kursivdruck); beim *experimentum crucis* heißt es dagegen: „I took two boards, and placed one of them close behind the prism at the *window*" (Newton [NTaL]:3078, unser Kursivdruck). Ob Newton in der Zwischenzeit die Fensterläden wieder aufgemacht hat oder ob er ein größeres Loch in den Fensterladen gebohrt hat, sagt er uns nicht.

12 An diesem Punkt seiner ausführlicheren Darstellung aus den *Opticks* verweist Newton daher zurück auf das dritte Experiment, also auf den dortigen Nachfolger des Grundexperiments aus dem Jahr 1672 (Newton [O]:21–24; 32 (= Book I, Part I, Proposition II, Experimente 3 und 6)). – Warum wir vorsichtigerweise gesagt haben, dass die neue Lochposition nur „zunächst" keine wesentlichen Unterschiede mit sich bringt, wird in Abschnitt III deutlich werden; dort tauchen einige Probleme auf, die bei der ursprünglichen Lochposition nicht aufgetaucht wären. In Abschnitt VII werden diese Probleme gelöst, und die Behauptung oben im Text gilt dann ohne Einschränkung.

13 Wer den Weg nachvollziehen will, den die einzelnen Lichtstrahlen Newton zufolge in dem Experiment zurücklegen, sollte sich die 4. Folie aus Lamperts „Lösung 2" ansehen, siehe Lampert [zWF]:113.

14 Lampert hat bei Newton eine ähnliche Beweisidee ausgemacht; er grenzt diese Idee erfolgreich von allerlei suboptimalen Beweisrekonstruktionen ab, siehe Lampert [NvG]:264–267 *et passim*.

15 Nur wenn die Blende eine extrem kleine Öffnung hat und das durchreisende Licht *beugt*, zeigen sich Farben. Wir werden von nun an alle Beugungsphänomene unberücksichtigt lassen.

16 Aus historischen Gründen spricht man bei einem Experiment, bei dem ein Bild direkt von der Retina des Auges aufgefangen wird, von einem *subjektiven* Experiment. Wird hingegen das Bild zunächst von einem Schirm aufgefangen, so spricht man von einem *objektiven* Experiment. Wir werden uns Newtons Vorliebe für objektive Experimente weitgehend anschließen. Dennoch werden wir Sie manchmal einladen, sich vorzustellen, dass Sie *in* den objektiv ausgeführten Experimenten von bestimmten Punkten aus mit einem gedachten Auge beobachten. Durch diesen Kniff lassen sich die Experimente besser verstehen. Das gedachte Auge wirkt hier immer wie ein Richtungsanalysator; es löst die Gesamthelligkeit eines Raumpunktes nach den richtungsspezifischen Anteilen auf. Jederzeit könnte es durch einen apparativen Richtungsanalysator ersetzt werden, ohne dass sich an unserer Argumentation etwas ändern würde. Newton hat sowohl seine Augen als auch einfachere apparative Richtungsanalysatoren eingesetzt. Der Schirm hinter einer Blende ist immer ein solcher Richtungsanalysator. Fest steht: Nur wer beide Beobachtungsmodi – den subjektiven und den objektiven – miteinander kombiniert, dringt zur vollkommenen physikalischen Beschreibung vor; der subjektive Beobachtungsmodus liefert die richtungsspezifische Abbildungsstruktur eines Raumpunktes, der objektive die daraus resultierende integrierte Gesamthelligkeit im fraglichen Raumpunkt.

17 Das vorsortierte Weiß ist deshalb dunkler als das vorsortierte Schwarz, weil das Sonnenlicht in dem Experiment vom Dunklen abgegrenzt wird; erstens erfüllt die Sonne selbst nicht den gesamten Himmel, vielmehr scheint sie vor dem Hintergrund des fast ganz finsteren Weltalls. Nehmen wir an, die Sonne würde leuchtend das ganze Weltall ausfüllen, dann gäbe es hinter dem Prisma keine vorsortierte Strahlenzone. Diese unsortierte Zone wäre perfekt weiß. Sobald wir sie prismatisch vorsortieren lassen wollen, müssen wir die Lichtquelle begrenzen, und dadurch verringern wir die Helligkeit in jener Zone. Kurz, vorsortierte Strahlenzonen sind nur mit Helligkeitsverlusten zu haben, in der technischen Optik spricht man von Lichtbündelbegrenzung, vergl. Bergmann et al [O]:105, 270. (Theoretische Ausnahme: Falls die Lichtquelle – anders als die Sonne und anders als jede real existierende Lichtquelle – unendlich stark leuchtete, wäre die Helligkeit der durch sie erzeugten vorsortierten Strahlenzone auch dann unendlich groß, wenn diese Strahlenzone per Prisma mithilfe von Begrenzungen erzeugt würde). Das vorsortierte Schwarz kann aber nur aus kleinen, entfernten dunklen Flächen entstehen und muss mithin von Weißem bzw. Leuchtendem begrenzt werden.

18 Für die vorsortierten Zonen können wir daher schließen, dass die nur gedanklich existierende Zone des perfekt vorsortierten Weiß keine Intensität hätte, demnach schwarz wäre und dass umgekehrt die Zone des vorsortierten Schwarz im Extremfall einfach nur weiß wäre. Das klingt paradox, besagt aber im Fall des vorsortierten Weiß nichts Rätselhaftes: Das Einsortierte geht gegen 0, das Aussortierte geht hingegen gegen 1 (sofern die Gesamtheit der unsortierten Lichtstrahlen auf den Wert 1 normiert wird). Im Falle des vorsortierten Schwarz wären die Verhältnisse genau umgekehrt. Zwar mag man es daher bei genauem Hinsehen terminologisch irreführend finden, dass wir der newtonischen Heterogenität des weißen Lichts die unorthodoxe Heterogenität der *Finsternis* entgegensetzen. Doch aus propädeutischen Gründen bleiben wir bei unserer Terminologie, da sie die Komplementarität der Verhältnisse deutlich herausstreicht, die der Sache nach vorliegt.

19 So schreibt Newton: „[…] ‚tis requisite that the Room be made *very dark*, least any scattering light, mixing with the colour, disturb and allay it" (Newton [NTaL]:3087, unser Kursivdruck). Was das heißt, sagt Newton weit deutlicher in den *Opticks*, in denen er schon beim ersten Experiment und gleichsam ein für allemal alles mögliche schwärzt, siehe Newton [O]:17 (= Book I, Part I, Proposition I, Experiment 1).

20 Goethe hat unseres Wissens als erster darauf aufmerksam gemacht, dass die newtonische Ausschaltung von Störeinflüssen keineswegs so nebensächlich ist, wie sie bei Newton daherkommt. Siehe Goethe [ETN]:§35.

21 Aus *Zahme Xenien VI*, zitiert nach Schöne [GF]:209.

22 Als ob er einen Beitrag zu unserem augenblicklichen Problem hätte liefern wollen, schreibt einer der Pioniere der phänomenologischen Optik (in einem anderen Zusammenhang): „Die Erfahrung bei einer Bergwanderung, wenn im frischen Schnee plötzlich Nebel aufkommt, belehrt uns: Es kann blendend hell sein, aber wir haben keine Ahnung mehr, wo es aufwärts, wo

es abwärts geht" (Maier [LiGv]:324). Das Phänomen ist als „white-out" bekannt und kann manchmal in arktischen Regionen beobachtet werden: „A condition that occurs in polar regions, when all visual clues to direction and distance are lost, leading to a dangerous state of disorientation. When the sky is uniformly overcast and snow cover is complete, the horizon, clouds, and the surface itself cannot be distinguished. An equal amount of light is received from all directions, so no shadows are cast and all depth perception is lost. Contrary to popular belief, a white-out does not necessarily involve blizzard conditions, i.e. blowing snow" (Dunlop [DoW]:252 mit weggelassenen Querverweisen). Siehe auch Blüthgen [AK]:94, Putnins [CoG]:66/7.

23 Siehe Rang [HaBz]:49 ff.

24 Viele Autoren, die sich mit Invertierung und Umstülpung newtonischer Experimente abgeplagt haben, suchten nach einem finsteren Analogon zur herkömmlichen Rede von Lichtquellen. So hat Bjerke der newtonischen „Lichtlehre" ein polemisches Analogon gegenübergestellt – die „Opakik", siehe Bjerke [NBzG].

25 Rang [HaBz]:49ff.

26 Anders als im Fall des pechschwarzen Komplements der Sonne kann man diesmal bei geeigneten Abmessungen (des Flecks sowie des Abstandes des Prismas von diesem Fleck) beim Blick durchs Prisma zwischen den Farberscheinungen eine schwarze Mitte sehen, also zwei getrennte Kantenspektren anstelle des damaligen (komplementären) Vollspektrums. (Dass wir beim prismatischen Blick auf die pechschwarze Sonne keine schwarze Mitte sehen, liegt erstens an den Dispersionseigenschaften unserer Prismen und zweitens daran, dass die Sonne – so wie ihr pechschwarzes Komplement – am Himmel immer denselben Winkel von rund 30 Bogenminuten einnimmt; weder ändert sich der Sonnendurchmesser nennenswert, noch unser Abstand von der Sonne). Wir geben zu: Diese schwarze Mitte im Kantenspektrum ist kein vorsortiertes Schwarz. Schadet das unserem Argumentationsziel? Keineswegs! Denn wir behaupten (und können empirisch nachweisen): Bei nochmaliger Vertauschung der Rollen von Licht und Finsternis (also in einer newtonischen Situation) entsteht unmittelbar hinter dem Prisma ebenfalls eine weiße Mitte, die ebenfalls kein vorsortiertes Weiß ist! Denn ein Blendenloch wie G im Schirm DGE macht nichts anderes als das Auge; es nimmt eine Richtungsanalyse der Beleuchtungsrichtungen am Ort des Auges bzw. Lochs vor. Im Rahmen der newtonischen Argumentation gilt: Am fraglichen Ort liegt keine ausreichende Lichtbündelbegrenzung vor, dort überlagern sich lauter Punktlichtquellen; die Superposition all der Einzelspektren erzeugt ein spektral völlig unsortiertes Weiß. Daher war Goethe im Recht, als er darauf verwies, dass die newtonische Theorie der Superposition lauter spektral zerlegter Punktlichtquellen nicht empirisch belegt werden kann. Newton dürfte sich dieser Schwierigkeit bewusst gewesen sein; das erklärt vielleicht, warum er so wenig Wert auf den fraglichen Bereich des vorsortierten Weiß gelegt und stattdessen das komplizierte *experimentum crucis* konstruiert hat.

27 In aller Allgemeinheit stimmt die Notwendigkeit („müssen") dieser

Behauptung nicht, und das aus zwei Gründen. Erstens könnten die fraglichen Lichtstrahlen zwar von der Prismenoberfläche AC herkommen, aber ohne aus dem Innern des Prismas herauszutreten; sie könnten das Ergebnis einer Reflexion an der Prismenoberfläche sein. Und zweitens könnte man die erforderlichen weißen Strahlen auch punktgenau an Ort und Stelle leiten, dann müssen sie nicht von der Lichtkammerwand ausgehen.

28 Der Ausdruck „nur" in diesem Satz bringt eine Übertreibung mit sich, die streng genommen nicht stimmt. Hier gilt dasselbe *caveat*, das wir in der vorigen Fußnote angebracht haben.

29 Hier unterscheidet sich der Ansatz, den wir favorisieren, am schärfsten von dem Ansatz, den Bjerke, Holtsmark und Sällström verfolgt haben. Auch dort, wo sie Newtons Experiment umstülpen, experimentieren sie im vorsortierten Weiß; ihre Umstülpung beginnt sozusagen erst an späteren Stationen des *experimentum crucis*. Siehe Holtsmark [NECR]:1234/5. Sehr überzeugende Varianten des umgestülpten *experimentum crucis* stammen von Ingo Nussbaumer. Eines seiner *objektiven* Experimente hat Nussbaumer am 17.6.2009 an der Humboldt-Universität zu Berlin erstmals einer größeren Öffentlichkeit vorgeführt; Nussbaumer sondert aus einem komplementären Spektrum mittels einer Spiegel*steg*vorrichtung einzelne farbige Abschnitte heraus und bettet sie in weißes (sowohl vorsortiertes wie unsortiertes) Licht ein, um diese weiß eingebetteten Farben des Komplementärspektrums durch ein zweites Prisma zu schicken. Mit diesem Experiment kann gezeigt werden, dass sich die Farben Türkis, Purpur und Gelb (aus dem komplementären Spektrum) in einer umgestülpten Situation ebenso ablenken lassen wie deren newtonische Gegenstücke Rot, Grün und Blau. Dass sich die Farben des komplementären Spektrums in einer umgestülpten Situation ebenso ablenken lassen wie die Farben des normalen Spektrums in der newtonischen Situation, haben Bjerke und Nussbaumer mittels *subjektiver* Experimente bereits gezeigt (Bjerke [NBzG]:66, 86, 88, Nussbaumer [zF]:86, 104, 150, 151, 188, insbesondere Tafeln VIII und IX). Mehr noch, Nussbaumer hat insgesamt sechs weitere strukturerhaltende Transformationen der newtonischen Experimente entdeckt, die paarweise komplementär sind. (Siehe Nussbaumer [zF]:159/160, sowie Tafeln XVIII bis XX).

30 Dass es sich nicht in seine newtonischen Bestandteile blau und rot zerlegt, hängt von den optischen Gesamtbedingungen in der *Streulicht*kammer ab; in einer *Dunkel*kammer lassen sich die newtonischen Bestandteile des Purpur dagegen sehr wohl prismatisch auffächern. Schickt man den purpurnen Teil des komplementären Spektrums in der Dunkelkammer erst durch eine Blende mit schwarzen Backen und dann durch ein Prisma, so kann man auf einem Schirm (in gebührendem Abstand vom Prisma) die Farben blau und rot auffangen. Das jedenfalls sagt Newtons Theorie voraus.

31 Eine ausführlichere Darstellung und Diskussion dieses Experiments findet sich in einem gesonderten Artikel, siehe Rang [HaBz]. Dort werden auch ausführlicher Hellräume als Invertierung von Dunkelräumen beschrieben und in ihren Eigenschaften diskutiert. Die dortige Argumentation unterscheidet sich von der hier vorgelegten dadurch, dass sie sich nicht an der

newtonischen Theorie orientiert, sondern die Sache mit den Begriffen einer phänomenologischen Beschreibung behandelt.

32 Der Ausdruck „underdetermination" stammt von Quine, siehe Quine [WO]:78; *locus classicus* ist Quine [oEES]. Mehr dazu demnächst in Müller [PE]. Weitere wissenschaftsphilosophische Schlüsse aus den vorgeführten Symmetrien bietet Müller [IA].

33 Dass Physiker in ihren Theorien und Experimenten nach Symmetrien suchen, wird oft mit dem Schönheitssinn der Physiker in Verbindung gebracht und mit dessen erkenntnistheoretischer Funktion. Eine Fallstudie zu diesem Thema liefert Müller [FK].

34 Dieser Aufsatz geht auf eine informelle Tagung in Adlershof vom Februar 2009 zurück, bei der Newtons *experimentum crucis* analysiert wurde. Im Lichte einiger Adlershofer Gespräche mit Matthias Rang verfasste Olaf Müller ein Arbeitspapier, um einen Gedankengang zu fixieren, dessen Vollendung ihm jahrelang nicht gelungen war, wegen der Schwierigkeiten mit dem vorsortierten Schwarz; erst in Adlershof erfuhr er, dass sich Matthias Rang längst die Ressourcen zur Lösung dieser Schwierigkeiten erarbeitet hatte, und so bildeten dessen (bis dahin unveröffentlichte) Ideen zum Hellraum eine wichtige Grundlage für Olaf Müllers Arbeitspapier. Das stark gekürzte Arbeitspapier ist dann gemeinsam von Matthias Rang und Olaf Müller überarbeitet worden. Wir danken Johannes Grebe-Ellis, Marc Müller, Ingo Nussbaumer und einem anonymen Gutachter für zahlreiche Anregungen und Verbesserungsvorschläge. Olaf Müller dankt Anna Welpinghus, Thomas Schmidt, Eric Oberheim und Timm Lampert für Kritik an früheren Fassungen seines Arbeitspapiers. – Die Abbildungen hat Matthias Herder zunächst nach Ideen von Olaf Müller entworfen; sie wurden von ihm nach den Wünschen der beiden Autoren mehrfach überarbeitet.

Anhang 1: Farbige Abbildungen

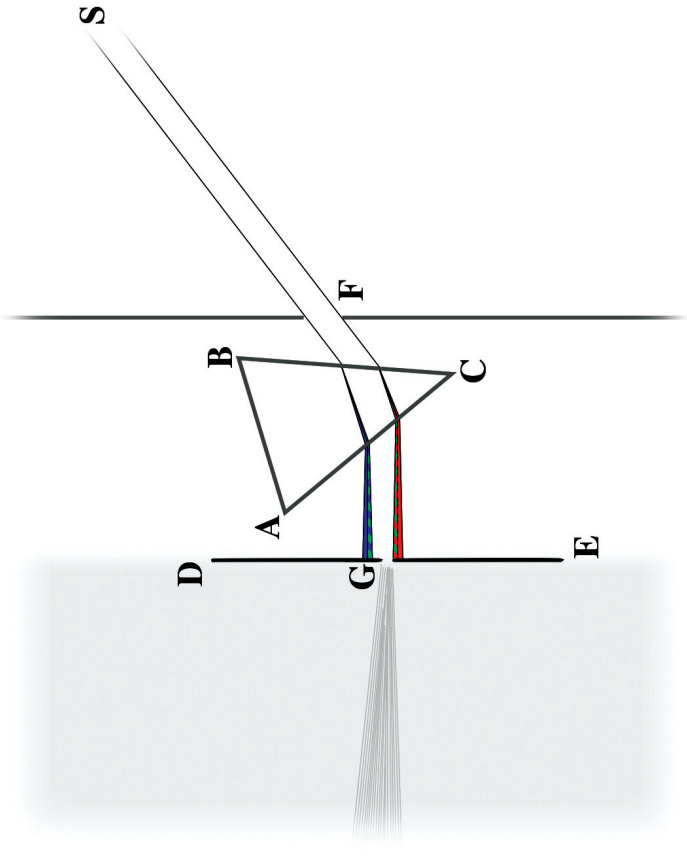


Abb. 5. Auftakt des *experimentum crucis* (vergrößerter Ausschnitt aus Abb. 4). Da der Schirm DE nah am Prisma ABC aufgebaut ist, zeigen sich hier die sogenannten Kantenspektren. Ganz oben haben wir eine blaue Zone eingezeichnet, an die eine hier grünblau schraffierte Zone angrenzt (die vom menschlichen Auge türkis wahrgenommen wird, also als ein Zwischenton, den wir im Newton-Spektrum weiter konsequent nicht berücksichtigen wollen). Entsprechend an der unteren Kante des gebrochenen Lichtbündels; die rote Zone ganz unten grenzt an eine hier grün-rot schraffierte Zone an (die vom menschlichen Auge gelb wahrgenommen wird). Die – additiven – Mischungsregeln, die wir hier voraussetzen, sind von den RGB-Bildschirmen her wohlbekannt.

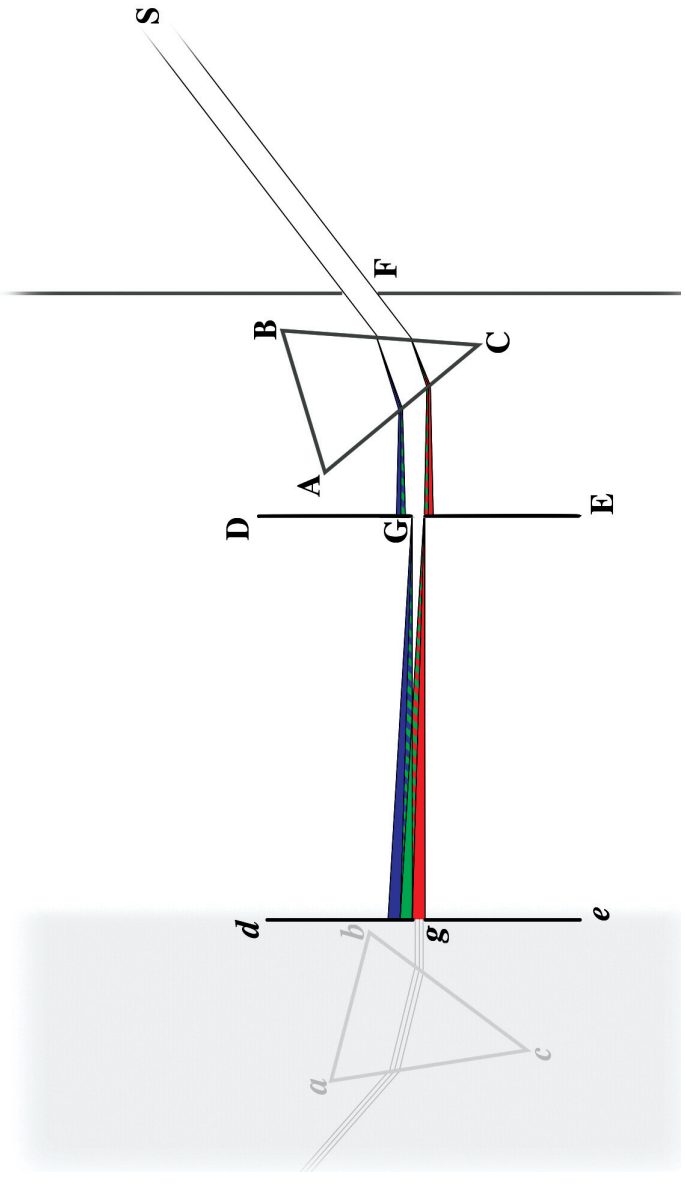


Abb. 6. Fortsetzung des *experimentum crucis*. Wir führen das Loch G der Lochblende DGE in die Zone des vorsortierten Weiß ein und fangen auf dem Schirm de das newtonische Spektrum auf. Der Schirm ist so weit vom Loch G entfernt, dass sich dort bereits das gesamte newtonische Grundspektrum zeigt.

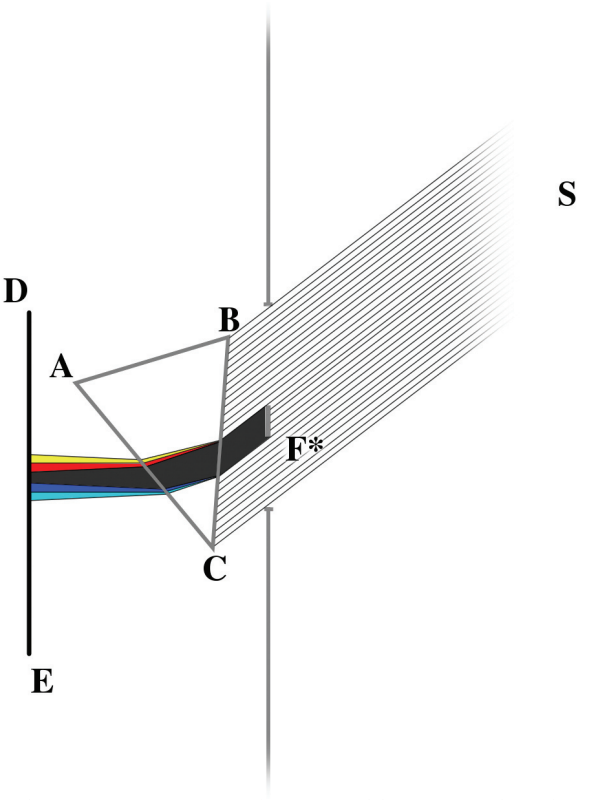


Abb. 9. Die komplementären Kantenspektren in Abb. 5. Wir geben die Farben an, die sich tatsächlich zeigen. Ganz oben ein gelber Rand (daneben rot), ganz unten ein türkisfarbener Rand (daneben blau), in der Mitte schwarz.

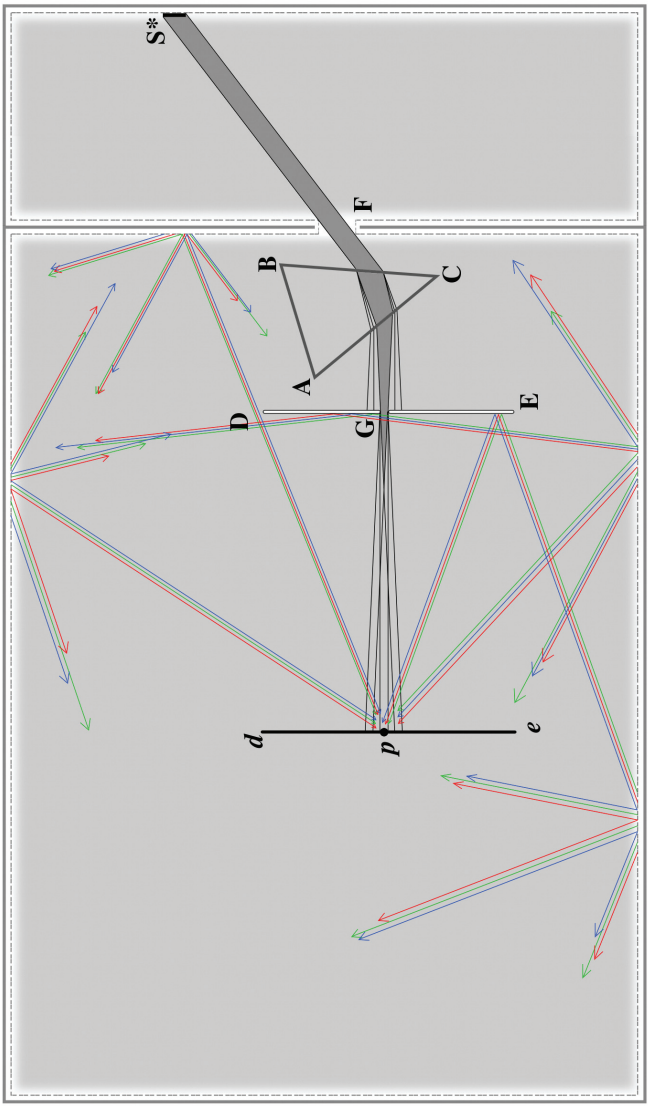


Abb. 10. Streulichtkammer – neuer Anlauf für die Umstülpung des *experimentum crucis*. Der Schattenwerfer F* des ersten gescheiterten Anlaufs (Abb. 9) wird wieder durch ein Fensterladenloch F ausgetauscht; rechts davon stülpen wir den Himmel um, setzen also an die Stelle der Sonne S (bei Newton) deren pechschwarzes Komplement, die Unsonne S*. Die Wände im Innern der Streulichtkammer (linker Teil der Abbildung) sind Lambertstrahler, sie emittieren newtonische Lichtstrahlen aller Farben, und zwar in alle Richtungen. Die Blende DGE hat dieselbe Geometrie wie bei Newton, aber ihre Blendenbacken sind nicht schwarz, sondern weiß.

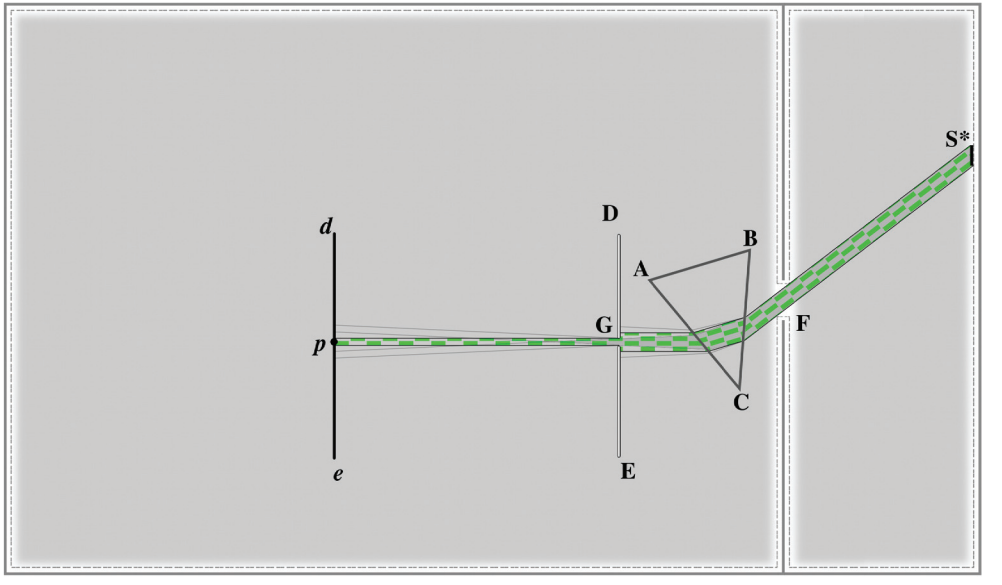


Abb. 11. Grünfreie Pfade in der Streulichtkammer (symbolisiert durch grüne Minuszeichen). In genau einer Richtung – nämlich aus Richtung der pechschwarzen Unsonne S* – dringen keinerlei grüne Lichtstrahlen durchs Fensterladenloch in die Streulichtkammer ein. Wo diese grünfreien Pfade die erste Prismenfläche BC schneiden, kommt zwar grünes Streulicht an (von den Wänden der Streulichtkammer und von der weißen Umgebung der Unsonne; hier nicht eingezeichnet). Aber es kommt aus *anderen* Richtungen; dies diffuse grüne Streulicht wird an der Prismenfläche BC also genau nicht auf die grünfreien Pfade gebrochen, die wir im Innern des Prismas eingezeichnet haben. (Die grünfreien Pfade ändern also an BC ihre Richtung genau so, wie grünes Licht an BC gebrochen würde, wenn es auf einem grünfreien Pfad auf das Prisma aufträfe). Entsprechend für die zweite Prismenfläche AC, an der unsere grünfreien Pfade zwar ihre Richtung abermals ändern, ohne allerdings mit richtungsgleichem grünen Licht aufgemischt zu werden. Die grünfreien Pfade werden von den Blendenbacken der Blende DGE unterbrochen, da deren Rückseiten weiß sind und weiß (also u.a. grün) bestrahlt werden. Dieser Effekt tritt nur beim Loch G nicht ein. Das bedeutet, dass sich die grünfreien Pfade gleichsam durch das Loch G hindurchstehlen können. Sie enden auf dem weißen Schirm de im Punkt p. Dort trifft viel grünes Licht aus allen Richtungen ein, nur nicht aus der einen – horizontalen Richtung – die wir jetzt verfolgt haben. Ergebnis: Unterrepräsentation grünen Lichts in p.

Abb. 12. Blaufreie Pfade in der Streulichtkammer. In genau derselben Richtung, wie rechts in Abb. 11 mittels grüner Minuszeichen angedeutet, dringen auch keinerlei blaue Lichtstrahlen durchs Fensterladenloch in die Streulichtkammer ein (nämlich wieder aus Richtung der pechschwarzen Unsonne S*). Wo auch diese blaufreien Pfade die erste Prismenfläche BC schneiden, kommt zwar zusätzlich noch blaues Streulicht an; *dies diffuse blaue Streulicht wird aber an der Prismenfläche BC nicht in dieselben Richtungen gebrochen wie das grüne Streulicht. Im Innern des Prismas nehmen die blaufreien Pfade also einen leicht anderen Verlauf als die grünfreien Pfade*. Sie pflanzen sich nach denselben Regeln fort, die wir bei Abb. 11 erklärt haben, enden aber auf dem weißen Schirm de *oberhalb des Punktes p*. Dort trifft viel blaues Licht aus allen Richtungen ein, nur nicht aus der einen – leicht schrägen Richtung – die sich in Abb. 12 verfolgen lässt. Ergebnis: Unterrepräsentation blauen Lichts oberhalb von p.

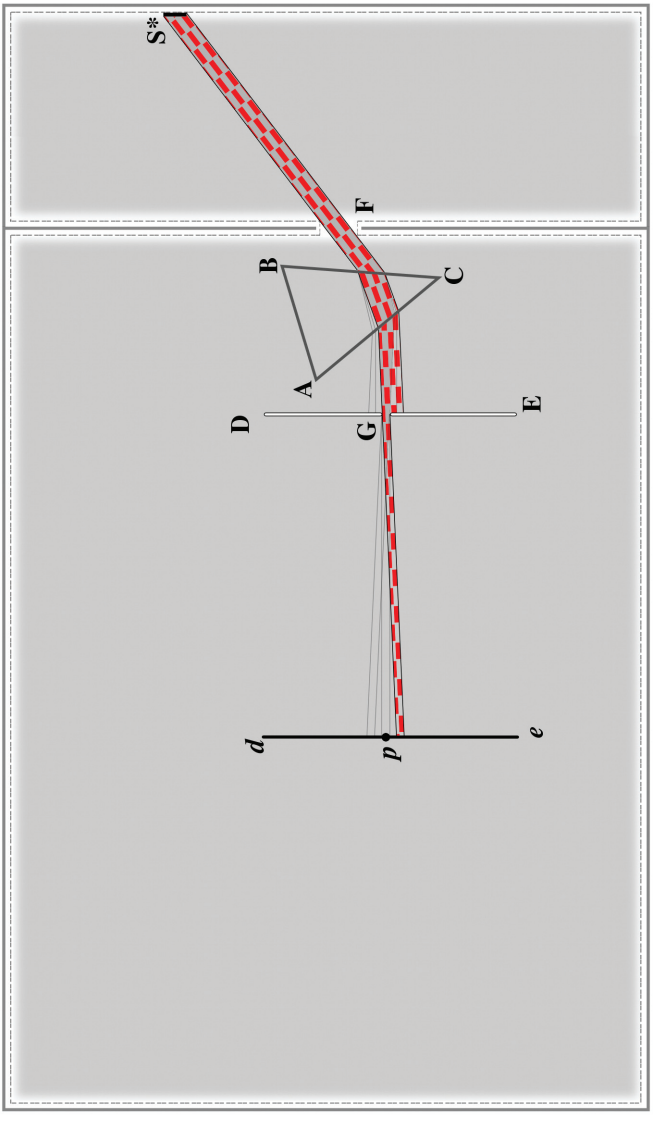


Abb. 13. Rotfreie Pfade in der Streulichtkammer. Der Weg rotfreier Pfade lässt sich nach denselben Regeln verfolgen wie (bei Abb. 11 und Abb. 12) für grün- und blaufreie Pfade erläutert. Ergebnis: Unterrepräsentation *roten* Lichts *unterhalb* von p.

Abb. 14. Das Komplementärspektrum in der Streulichtkammer. Wenn wir die Ergebnisse aus Abb. 11, Abb. 12 und Abb. 13 übereinanderlegen, wird deutlich: Die blau-, grün- und rotfreien Pfade entfalten sich (in der Streulichtkammer) nach demselben Muster wie die blauen, grünen und roten Pfade in der Dunkelkammer (vergl. Abb. 6). Wo die blaufreien Pfade oberhalb von p auf den Schirm de treffen, zeigt sich Gelb, die Komplementärfarbe von Blau. Wo die grünfreien Pfade bei p auf den Schirm de treffen, zeigt sich Purpur, die Komplementärfarbe von Grün. Und wo die rotfreien Pfade unterhalb von p auf den Schirm de treffen, zeigt sich Türkis, die Komplementärfarbe von Rot.

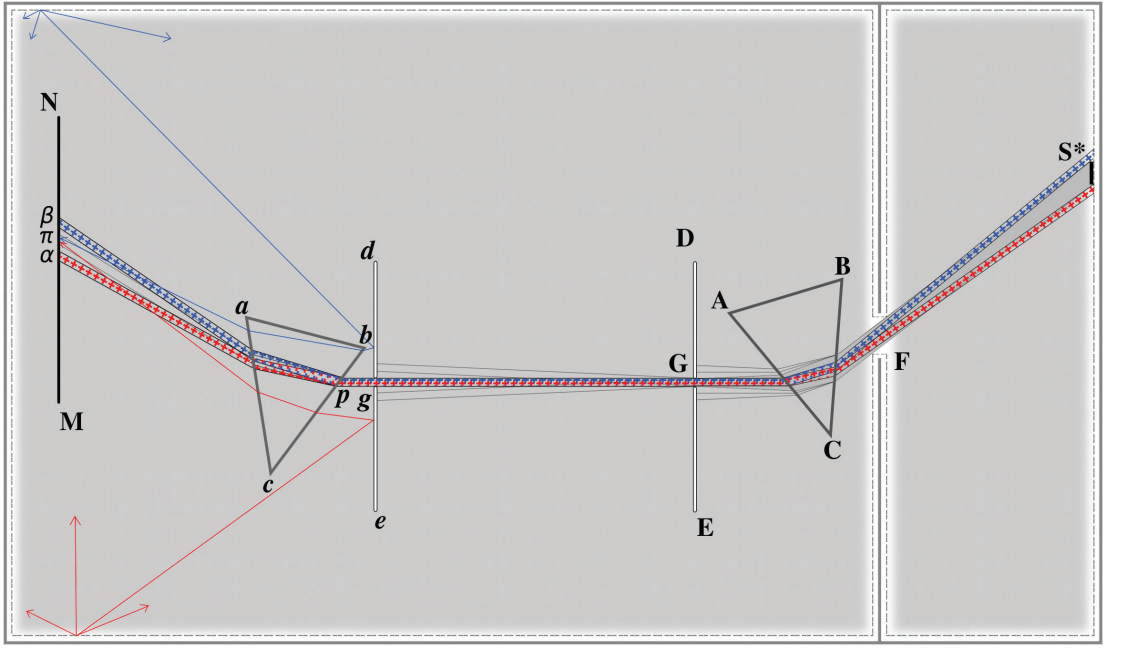


Abb. 15. Effekte des zweiten Prismas bei Umstülpung des *experimentum crucis* in der Streulichtkammer. Die grünfreien Pfade stehlen sich durchs zweite Loch g der Blende dge genauso hindurch wie durchs erste Loch G der Blende DGE (vergl. Erläuterungen bei Abb. 11). Hier haben wir zwei newtonische Lichtsorten eingezeichnet, die auf diesen grünfreien Pfaden durch die Streulichtkammer reisen, einerseits blaue, andererseits rote (jeweils angedeutet durch farbige Pluszeichen). Sie gelangen durchs Loch g hindurch und werden beim Weg durchs Prisma abc in verschiedene Richtungen abgelenkt. Wo sie hinter dem Prisma auseinanderdriften, reißt eine Lücke auf, in der es natürlich auch noch an grünem Licht (der erforderlichen Richtung) mangelt; das zeigt sich insbesondere bei π. – Da es in der Streulichtkammer nirgends an blauem oder rotem Licht mangelt, haben wir *pars pro toto* einige weitere blaue und rote Lichtstrahlenbündel eingezeichnet. Hätten wir alle blauen oder roten Lichtstrahlenbündel einzeichnen wollen, so hätten wir das Bild überbevölkern müssen. Wie man aber sehen kann, kommen sogar bei π blaue und rote Lichtstrahlenbündel an.

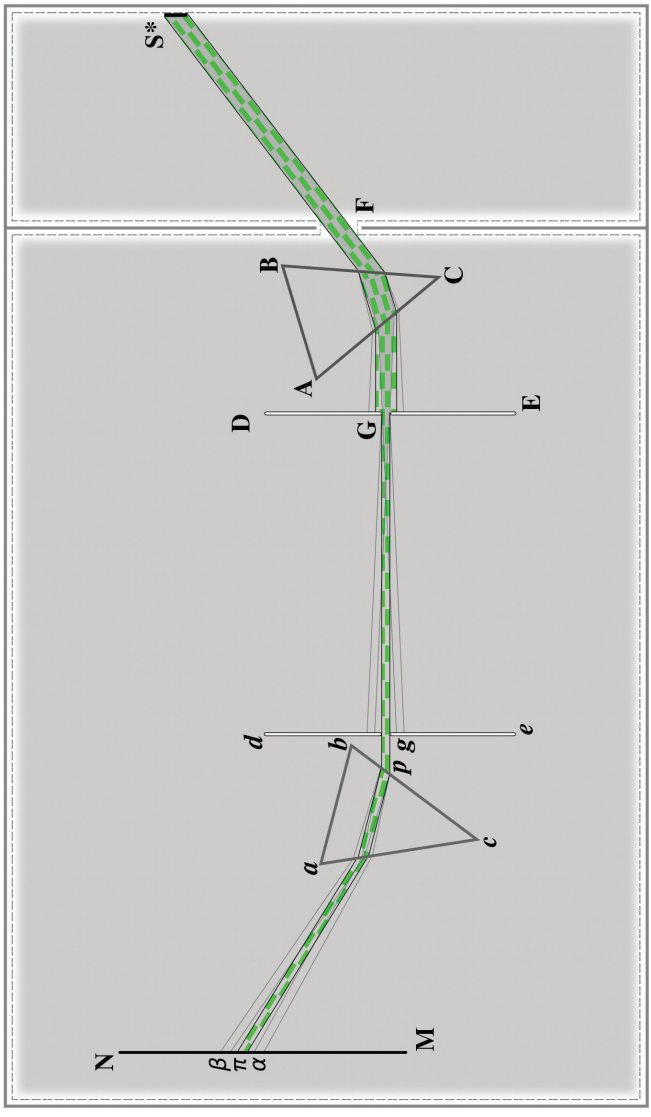


Abb. 16. Grünfreie Pfade bei vollständiger Umstülpung des *experimentum crucis* in der Streulichtkammer. Was wir in Abb. 15 mittels blauer und roter Lichtstrahlen gezeigt haben (d. h. in unserer reduzierten Sprechweise mittels aller Lichtstrahlen mit Ausnahme der grünen), zeichnen wir hier abermals ein, diesmal mithilfe des Konzepts grün*freier* Pfade.

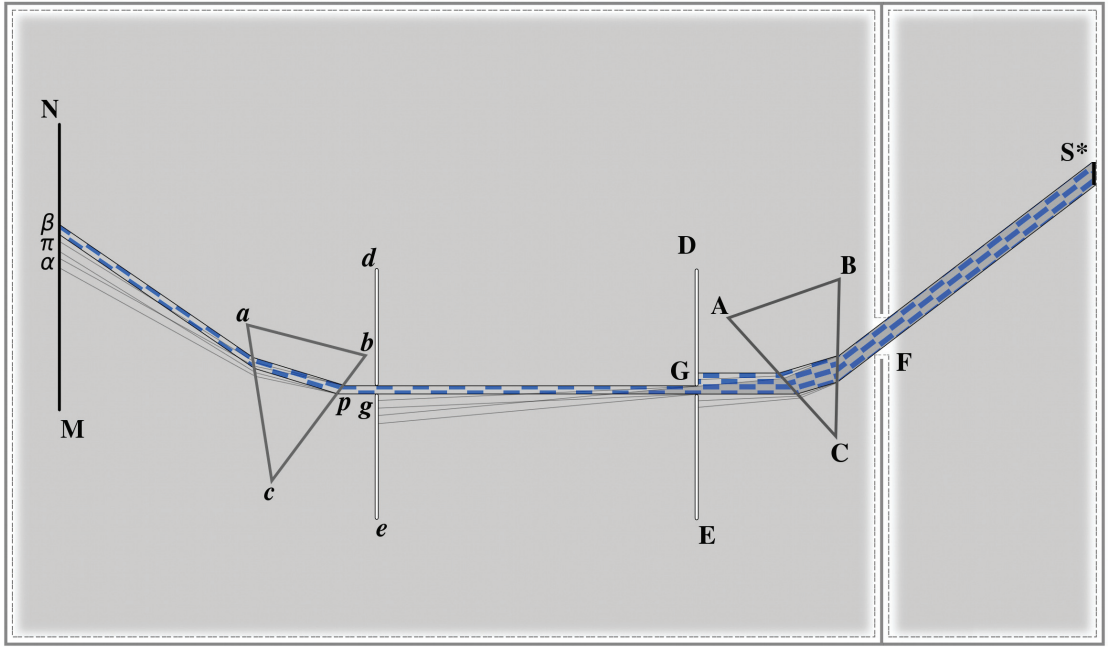


Abb. 17. Drehung des ersten Prismas bei Umstülpung des *experimentum crucis.* Sobald wir das erste Prisma ABC um seine Achse rotieren lassen, fällt ein anderer Teil des in Abb. 14 hergeleiteten Komplementärspektrums durchs Loch g der Blende dge, z.B. der Bereich, der sich mithilfe blaufreier Pfade analysieren lässt (Abb. 12). Diese blaufreien Pfade treffen im selben Winkel und an derselben Stelle aufs Prisma abc wie vorher die grünfreien Pfade (Abb. 16). Sie verhalten sich nach denselben Spielregeln wie vorher die grünfreien Pfade, enden aber auf dem Schirm NM etwas oberhalb der Stelle π, an der die grünfreien Pfade endeten.

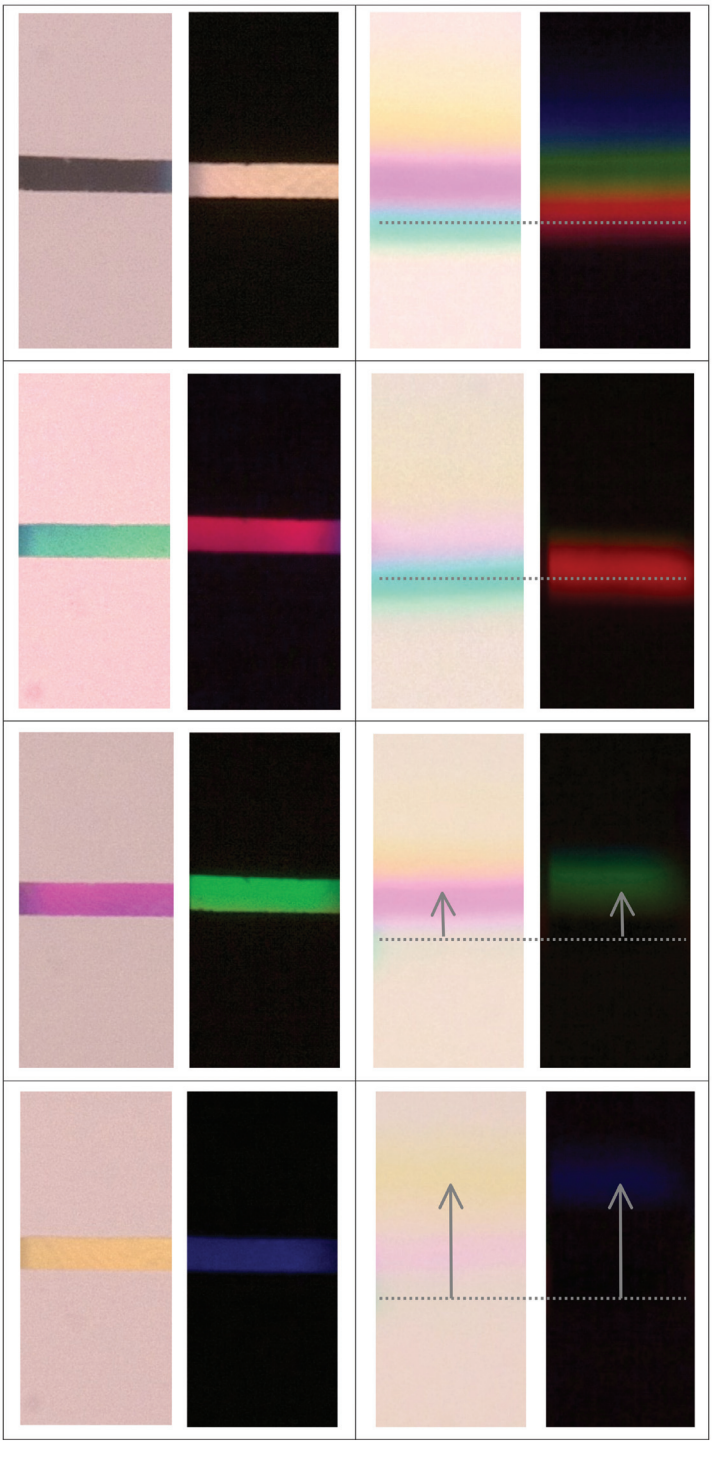


Abb. 18 (vorige Seite). Photographien des *experimentum crucis* und seines umgestülpten Gegenstücks in der Streulichtkammer. Kästchen mit dunkler Bildumgebung beziehen sich immer auf Newtons *experimentum crucis*, Kästchen mit heller Umgebung immer auf dessen Umstülpung. Linke Spalten: Hier ist photographiert, was die Kamera registriert, die vom Ort des Prismas abc durchs zweite Blendenloch g aufs erste Blendenloch G gerichtet ist (jeweils für beide Experimente nebeneinander). Das Prisma abc ist für diese Aufnahmen entfernt. Rechte Spalten: Photographien vom Schirm NM *durch* das Prisma abc auf die in der linken Spalte gezeigten Ansichten. Oberste Reihe: Anstelle der Sonne bzw. umgestülpten Sonne wurde eine ausgedehnte dunkle bzw. helle Fläche verwendet. Zweite bis vierte Reihe: Für das *experimentum crucis* in der Dunkelkammer wurde ein heller Spalt verwendet, für das umgestülpte Experiment ein gleich dimensionierter dunkler Steg. Innerhalb jeder Reihe bleibt die Prismenstellung ABC gleich. Doch von Reihe zu Reihe wird das Prisma ABC verdreht; dann wandert beim Blick durchs Prisma abc das farbige Bild auf und ab, wie durch die Pfeile angedeutet (rechte Spalte). Für beide Experimente ergeben sich gleiche geometrische Verhältnisse, und das bei komplementären Farb- und Helligkeitsverhältnissen.

Anhang 2: Tabelle für die Übersetzung der newtonischen Theorie in ihr Gegenstück

Begriffe aus der newtonischen Optik	Begriffe aus deren Umstülpung
Phänomenologische Begriffe:	*Phänomenologische Begriffe:*
Blau	Gelb
Grün	Purpur
Rot	Türkis
Weißes Licht	Schwarzer Schatten
(vor-) sortiertes Licht	(vor-) sortierter Schatten
Störendes Streulicht	Störender Streuschatten
Theoretische Begriffe:	*Theoretische Begriffe:*
Lichtstrahl	Schattenstrahl
Heterogenes Licht	Heterogene Finsternis
Homogenes Lichtbündel	Homogenes Schattenbündel
Experimentelle Begriffe:	*Experimentelle Begriffe:*
Kausal wirksamer optischer Faktor	Kausal wirksamer optischer Faktor
Prisma	Prisma
Spiegel	Spiegel
optische Begrenzung	optische Begrenzung
optisch (un-)durchlässige Struktur	optisch (un-)durchlässige Struktur
Lochblende	Lochblende
Blende mit schwarz erscheinenden Backen	Blende mit weiß erscheinenden Backen

Literatur

Bergmann, Ludwig; Schaefer, Clemens, 2004 [O]: *Optik*, 10. Auflage. Berlin: de Gruyter. (= Niedrig, Heinz (Hrsg.): *Lehrbuch der Experimentalphysik.* Bd. 3).

Bieri, Hanspeter; Zwahlen, Sara Margarita (Hrsg.), 2008 [TOAW]: *„Trinkt, O Augen, was die Wimper hält…". Farbe und Farben in Wissenschaft und Kunst.* Bern: Haupt Verlag.

Bjerke, André, 1961 [NBzG]: *Neue Beiträge zu Goethes Farbenlehre*. Stuttgart: Freies Geistesleben.

Blüthgen, Joachim, 1980 [AK]: *Allgemeine Klimageographie*. Berlin/ New York: de Gruyter.

Dunlop, Storm, 2005 [DoW]: *A dictionary of weather*. Oxford: Oxford University Press. (Erschien zuerst 2001).

Glasauer, Stefan; Steinbrenner, Jakob (Hrsg.), 2007 [F]: *Farben*. Frankfurt/Main: Suhrkamp.

Goethe, Johann Wolfgang von, 1810 [EF]: *Entwurf einer Farbenlehre. Des ersten Bandes erster, didaktischer Teil.* In: Goethe [SzN] I.4, S. 11–266. (= LA I 4, S. 11–266).

Goethe, Johann Wolfgang von, 1810 [ETN]: *Enthüllung der Theorie Newtons. Des ersten Bandes zweiter, polemischer Teil.* In: Goethe [SzN] I.5. (= LA I 5).

Goethe, Johann Wolfgang von, 1810 [EzGF]: Erklärung der zu Goethes Farbenlehre gehörigen Tafeln. In: Goethe [SzN] I.7, S. 41–115.

Goethe, Johann Wolfgang von, 1955 [SzN] I.4: *Die Schriften zur Naturwissenschaft. Erste Abteilung, Band 4. Zur Farbenlehre: Widmung, Vorwort und didaktischer Teil.* Bearbeitet von Rupprecht Matthaei. Weimar: Hermann Böhlaus Nachfolger. (= LA I 4).

Goethe, Johann Wolfgang von, 1958 [SzN] I.5: *Die Schriften zur Naturwissenschaft. Erste Abteilung, Band 5. Zur Farbenlehre: Polemischer Teil.* Bearbeitet von Rupprecht Matthaei. Weimar: Hermann Böhlaus Nachfolger. (= LA I 5).

Goethe, Johann Wolfgang von, 1957 [SzN] I.7: *Die Schriften zur Naturwissenschaft. Erste Abteilung, Band 7. Zur Farbenlehre: Anzeige und Übersicht, statt des supplementaren Teils und Erklärung der Tafeln*. Bearbeitet von Rupprecht Matthaei. Weimar: Hermann Böhlaus Nachfolger. (= LA I 7).

Grebe-Ellis, Johannes, in Vorbereitung [NECa]: Newtons experimentum crucis aus der Perspektive Goethes.

Gruner, Sol M., 1973 [DFL]: Defending Father Lucas: A consideration of the Newton-Lucas dispute on the nature of the spectrum. In: *Centaurus* 17, S. 315–329.

Holtsmark, Torger, 1970 [NECR]: Newton's experimentum crucis reconsidered. In: *American Journal of Physics* 38, S. 1229–1235.

Kirschmann, August, 1924 [USS]: Das umgekehrte Spektrum und die Spektralanalyse. In: *Zeitschrift für Instrumentenkunde* 44, S. 173–175.

Kirschmann, August, 1926 [USsF]: Das umgekehrte Spektrum und seine Farben sowie seine Bedeutung für die optische Wissenschaft. In: Krüger (Hrsg.) [LF], S. 411–442.

Kirschmann, August, 1917 [USsK]: Das umgekehrte Spektrum und seine Komplementärverhältnisse. In: *Physikalische Zeitschrift* 18, S. 195–205.

Krüger, Felix (Hrsg.), 1926 [LF]: *Licht und Farbe.* Beck'sche Verlagsbuchhandlung: München. (= *Neue Psychologische Studien* Band 2. Drittes Heft).

Lampert, Timm, 2008 [NvG]: Newton vs. Goethe: Farben aus Sicht der Wissenschaftstheorie und Wissenschaftsgeschichte. In: Bieri et al. (Hrsg.) [TOAW], S. 253–278.

Lampert, Timm, 2000 [zWF]: *Zur Wissenschaftstheorie der Farbenlehre. Aufgaben, Texte, Lösungen*. Bern: Bern Studies for the History and Philosophy of Science, Libri Books on Demand.

Lohne, Johannes A., 1965 [EC]: „Experimentum crucis". *Notes and records of the Royal Society of London* 23, S. 169–199.

Maier, Georg (Hrsg.), 2004 [BSS]: *Blicken, sehen, schauen: Beiträge zur Physik als Erscheinungswissenschaft.* In: Grebe-Ellis, Johannes (Hrsg.). Dürnau: Verlag der Kooperative Dürnau.

Maier, Georg, 1996 [LiGv]: Das Licht im Gestrüpp von Missverständnissen. In: Maier [BSS], S. 324–338.

Müller, Olaf L., 2010 [FK]: Farbspektrale Kontrapunkte: Fallstudie zur ästhetischen Urteilskraft in den experimentellen Wissenschaften. In: Nussbaumer [RE].

Müller, Olaf L., 2007 [GPUb]: Goethes philosophisches Unbehagen beim Blick durchs Prisma. In: Glasauer et al. (Hrsg.) [F], S. 64–101.

Müller, Olaf L., 2008 [IA]: Innen und Außen: Zwei Perspektiven auf analytische Sätze. In: *Philosophia naturalis* 45, S. 5–35.

Müller, Olaf L., eingereichtes Manuskript [PE]: Prismatic equivalence:

A new case of underdetermination? Goethe vs. Newton on the Prism Experiments.

Newton, Isaac, 1960 [CoIN]/II: *The correspondence of Isaac Newton. Volume II. 1676–1687.* Turnbull, H. W. (Hrsg.). Cambridge: Cambridge University Press.

Newton, Isaac, 1958 [INPL]: *Isaac Newton's papers & letters on natural philosophy and related documents.* Cohen, Bernard (Hrsg.). Cambridge, Massachusetts: Harvard University Press.

Newton, Isaac, 1984 [LOOL]: *Lectiones opticae – Optical lectures*. In: Newton [OPoI]/I, S. 46–279.

Newton, Isaac, 1671 [NTaL]: A new theory about light and colors. *Philosophical Transactions* 80 (February 19, 1671/2), S. 3075–3087. (Abgedruckt in Cohen (Hrsg.) [INPL], S. 47–59). [Zitat nach der Originalpaginierung].

Newton, Isaac, 1704 [O]: *Optics: or, A treatise of the reflections, refractions, inflections and colours of light.* In: Newton [OQEO]/IV, S. 1–264. [Erschien zuerst 1704; lateinisch 1706; überarbeitete englische Ausgabe 1717/8; die vierte englische Auflage, in der noch Newtons handschriftliche Änderungen der dritten Auflage berücksichtigt sind erschien 1730].

Newton, Isaac, 1984 [OPoI]/I: *The optical papers of Isaac Newton.* Volume I. In: Shapiro, Alan E. (Hrsg.). Cambridge: Cambridge University Press.

Newton, Isaac, 1964 [OQEO]: *Opera quae extant omnia.* Stuttgart: Friedrich Frommann Verlag. [Faksimile-Neudruck der Ausgabe von Samuel Horsley, London, 1779–1785, in fünf Bänden].

Nordmeier, Volkhard; Grötzebauch, Helmuth (Hrsg.), 2010 [TFDP]: *Tagungsband der Frühjahrstagung 2009 der Deutschen Physikalischen Gesellschaft.* Berlin: Lehmanns Media.

Nussbaumer, Ingo, 2010 [RE]: *Rücknahme und Eingriff: Malerei der Anordnungen.* Nürnberg: Verlag für moderne Kunst.

Nussbaumer, Ingo, 2008 [zF]: *Zur Farbenlehre. Entdeckung der unordentlichen Spektren.* Wien: Edition Splitter.

Orvig, S. (Hrsg.), 1970 [CoPR]: *The climate of the polar regions*. Amsterdam/London/New York. (= World Survey of Climatology Volume 14).

Putnins, P., 1970 [CoG]: The climate of Greenland. In: Orvig (Hrsg.) [CoPR], S. 3–128.

Rang, Matthias; Grebe-Ellis, Johannes, 2009 [KS]: Komplementäre Spek-

tren. Experimente mit einer Spiegel-Spalt-Blende. In: *Mathematisch-naturwissenschaftlicher Unterricht* 62 Heft 4, S. 227–230.

Rang, Matthias, 2009 [HaBz]: Der Hellraum als Bedingung zur Invertierung spektraler Phänomene. In: *Elemente der Naturwissenschaft* 90 (2009), S. 46–79.

Rang, Matthias, erscheint im Frühjahr 2010 [MS]: Mehrfachanwendung von Spiegelspaltblenden und Prismen – eine moderne Form von Newtons *experimentum crucis*. In: Nordmeier et al. (Hrsg.) [TFDP].

Sabra, Abdelhamid I., 1981 [ToLf]: *Theories of light from Descartes to Newton.* Cambridge: Cambridge University Press. [Erschien zuerst 1967].

Sällström, Pehr, 2010 [MS]: *Monochromatische Schattenstrahlen. Ein Experimentalfilm.* Dreisprachige DVD; Englisch, Deutsch, Schwedisch. Kassel. (In Produktion).

Schöne, Albrecht, 1987 [GF]: *Goethes Farbentheologie.* München: Beck.

Shapiro, Alan E., 1980 [ESoN]: The evolving structure of Newton's theory of white light and color. In: *Isis* 71, No. 2, S. 211–235.

Shapiro, Alan E., 1996 [GAoN]: The gradual acceptance of Newton's theory of light and color, 1672–1727. In: *Perspectives on Science* 4, No. 1, S. 59–140.

Westfall, Richard S., 1966 [NDHF]: Newton defends his first publication: The Newton-Lucas correspondence. In: *Isis* 57, S. 299–314.

Westfall, Richard S., 1962 [NHCo]: Newton and his critics on the nature of colors. *Archives internationales d'histoire des sciences* 15, S. 47–58.

Westfall, Richard S., 1963 [NRtH]: Newton's reply to Hooke and the theory of colors. In: *Isis* 54, S. 82–96.

Francisco Antonio Doria, Manuel Doria

On formal treatments for general relativity

Abstract

We give a unified axiomatic treatment for the several mutually contradictory theories in the domain of general relativity, and examine a few non-trivial consequences of our proposal. We also consider our proposed axiomatization in the light of Nagel's concept of reducibility.

Zusammenfassung

Es werden verschiedene, sich gegenseitig widersprechende, allgemein relativistische Theorien einheitlich axiomatisch behandelt und einige nicht triviale Folgerungen entwickelt. Die vorgeschlagene Axiomatisierung wird dann mit Ernest Nagels Konzept der Reduktion in Verbindung gebracht.

1. Introduction

When one discusses reducibility between axiomatic theories, one usually starts from, say, some theory T which is supposed to have an adequately nice axiomatization and try to reduce to it (in some convenient sense) a theory T_2, whose formal structure is also supposed given. Quine and others have aptly defended a message which remains in the orthodoxy of philosophy of science, that the domain of physics is supposed to be the only maximally exhaustive scientific inquiry on nature with all other scientific disciplines being local extensions on the general framework established by physicists. However, when one is to investigate the state of the art of axiomatized bedrock scientific theories of physics and apply the conceptual analysis of philosophers of science of the last decades such as intertheoretic reduction in more specific examples, the task is not without difficulty.

Take for instance the widespread view of the „layered cake" approach to scientific theories where we see all disciplines in some scientific domain as a series of layers where layer $i + 1$ (which corresponds to theory T_{i+1})

reduces to layer i, that is to say, to theory T_i (we will soon add more detail to this conception). Does that general viewpoint hold at least for physics when we consider the actual relations among its explicitly given subareas and (possibly conflicting) subdomains?

The case of general relativity

General relativity, or Einstein's gravitational theory, is a well-established domain in contemporary physics. But, what does a physicist *actually* mean when he (or she) refers to a „general relativistic theory"?

In fact, such a reference points to an heterogeneous collection of theories, some of them mutually contradictory, and yet with common traits, say, tensor calculus as the underlying language, gravitation described by a metric tensor or by a curvature tensor or by some kind of affine linear connection, together with some version of general covariance. Then the question is: can we treat in a unified way those theories that belong to the realm of general relativity according to a physicist's viewpoint?

Moreover, can we find some general theory that would encompass all that is, let us say, pragmatically recognized as a general relativistic theory? Which is the relation between such a general theory – if it exists – and particular instances, namely the current various theories of gravitation that stemmed from Einstein's original construction? The point is: in actual scientific practice we frequently have to deal at the same time with two or three different theoretical constructs for the same set of phenomena; how are to juggle them together as formal objects in a reasonable way?

Goals of the paper

To summarize it:

– We use the Suppes predicate technique to axiomatize general relativistic theories within Zermelo-Fraenkel set theory (plus the axiom of choice, ZFC). This axiomatization has already been used by the first author with N. C. A. da Costa in order to derive undecidable sentences in physical theories (Costa 1991, Costa 1992, Costa 1990).
– We give new examples of the incompleteness phenomenon in general relativistic theories for our axiom system, and notice that such a construction is universal for Suppes predicate based axiomatics.
– We try to understand our techniques in the light of Nagel's analysis of the reducibility of theories.

Details

We show here how to give a kind of unified formal, axiomatic treatment for the domain of general relativity (GR) that includes in a single formulation all the different, possibly conflicting and contradictory theories that belong to the GR domain. Our ideas may be extended to other domains in physics, such as the manifold (and eventually contradictory) quantum theories, for example.

Essentially what we do is: we offer a formal characterization for all theories in the GR domain, and see the various particular cases as *representations* or *interpretations* for the prescribed syntactic structure that we have chosen as our axiomatic background. As we will see we follow in that the usual practice of physicists.

Our starting point is a recently introduced axiomatization for Einstein's gravitation theory: we informally sketch here an axiomatic treatment for general relativity that has been used by one of the authors (Costa 2007, Costa 1990) in order to obtain metamathematical results that might be relevant for physics.

We will see that the actual task of dealing with the domain of general relativity in axiomatic terms may turn out to be quite complicated due to the mathematical objects one has to consider, and we suggest that the best approach consists in presenting a stripped-down axiom system that reflects some kind of „bare minimum" collection of concepts for the field, and then explore several, possibly unrestricted, interpretations of that stripped-down axiom system within the set theoretic universe. Of course there will be some arbitrariness in that so to say 'minimal theory' we start from, as it arises from a pragmatic choice.

We will then show that while our approach offers a particular construction, it is in fact very general, and our conclusions apply to a large class of axiom systems that might be constructed for the same domain or analogous domains in physics. It is a kind of ‚universal' treatment for the axioms of general relativistic theory, in a sense to be made precise.

The rationale behind our analysis goes as follows: even for a well-established, essentially mathematical domain in physics such as general relativity, we have to start out of not just a single theory, but out of a family of theories which are very different in their scope and contradictory in their results, and which sometimes might even have rather fuzzy, indistinct contours that separate them from other areas in physics. We try here to put some order into that situation.

Moreover our approach has a kind of universal property: any other axiom-atization of general relativity with the help of Suppes predicates will have the same undecidable sentences as the ones presented here. That fact is discussed in the concluding section of this paper.

Note
The present paper is part of an ongoing effort by the first author to tackle Hilbert's 6th Problem, which is the ugly duckling among the Hilbert Problems. It started with Costa (1990), led to the results in Costa (1991), and was presented in detail in Costa (1992). A recent review is Costa (2007).

2. A concept of reducibility

Most discussions of intertheoretic reduction in philosophy of science stem from the concept of reduction introduced by Ernest Nagel in his seminal work *The Structure of Science* (Nagel 1979). The nagelian account of reduction is a form of logical deduction requiring axiomatic treatment of both the reducing and reduced scientific theories (Schaffner 1967). Let's be more specific.

A theory T_2 is said to be reducible to a theory T_1 in the nagelian sense if and only if the axioms for T_2 are logically derivable from those in T_1. The simplest path in which this could work has very low applicability in the natural sciences, that is to say, the case of *homogeneous* reduction.

Homogeneous reduction is feasible when we find a direct correspondence between all the entities in the formal treatment of the reduced theory T_2 with the objects (terms, etc) in the reducing theory T_1.

However the most important, truly novel contribution of Nagel deals with the case where homogeneous reduction is not possible, even if one feels that somehow there should be a reduction of T_2 to T_1. In that case there is a discontinuity between the vocabulary of T_1 and T_2, particularly when T_2 requires entities for which there is no identification to be found in the terms of T_1. The approach we ought to follow here is to use the representational power of the reducing theory T_1 to introduce entities in T_1 that precisely mirror the properties of those from the reduced theory T_2 which have resisted naive correspondence (Klein 2009).

Formally, it means that one should connect each currently irreducible

predicate of T_2 to a nomologically coextensive predicate in the base theory T_1 with the following biconditional Kim 2005:

$$\forall x_1, \ldots, x_n [M_1(x_1, \ldots, x_n) \leftrightarrow M_2(x_1, \ldots, x_n)].$$

(Here M_i refers to T_i.) This procedure is the *heterogeneous* reduction. Whether or not this particular model of intertheoretic reduction is suitable has been disputed (see for instance (Kim 2000, Chapter 4) but given its pervasiveness in the literature, it will still be the focus of our analysis.

Such is the process that satisfies nagelian connectability between two different scientific theories. As it is well-known, the quintessential example of the nagelian program of intertheoretical reduction in physics concerns the relationship between statistical mechanics and thermodynamics. The concept of heat which is totally absent in the axiomatic treatment of statistical mechanics but necessarily present in thermodynamics can be described in the language of statistical mechanics as „mean molecular kinetic energy", allowing for reduction.

What we propose to do

We will see that our approach in the present paper may be seen as a variant of Nagel's reducibility, or connectability between two theories. We will build for the case of general relativity a kind of „master" theory of which every particular theoretical construction in general relativity appears as a representation or interpretation in a precise formal sense.

3. Construction of a first axiom system for GR

We will just sketch the main ideas here; the clerical details in the construction can be found in Costa (2007).

The axiomatization for general relativity – GR in short – that we discuss here has been briefly presented in Costa (2007), Costa (1990). General relativity is usually taken as the theory of the Einstein gravitational equations,

$$R_{\mu\nu} - (1/2)g_{\mu\nu}R = kT_{\mu\nu}$$

where the left-hand side of the equation is the Ricci-Einstein tensor, and $T_{\mu\nu}$ is the energy-momentum tensor that acts as a source for the Ricci-Einstein tensor.

In order to proceed with our axiomatics we must say:
– Which is the arena where the Einstein equations sit?
– How do we obtain $T_{\mu\nu}$?

The set-theoretic environment where we find the gravitational equations is a domain within a 4-dimensional real differentiable manifold. (By differentiable one usually means, at least of class C^1, since C^1 manifolds always have an atlas of class C^∞ by a well-known result.) Such manifolds are objects with a very rich structure which must be specified when we give our axioms for GR.

The energy-momentum tensor introduces a complication. For our theory will have different Lagrangian densities depending on the kind of interaction, if any, that there is between geometry and physics (we must of course consider the $T_{\mu\nu} = 0$ case too, where the Einstein equations reduce to the vanishing of the Ricci tensor $R_{\mu\nu} = 0$). And we cannot simply postulate in our axioms a Lagrangian density such as:

$$\mathcal{L} = \sqrt{(-g)}\, R + \mathbf{\Phi},$$

where Φ are matter terms, since that would be too vague.

We will copy an idea that has been first used in the so-called C*-algebra formulation for quantum mechanics (Emch 1972). In the present case the axiomatics gives us a general algebraic, „syntactic", framework, and each possible physical situation will be given by some interpretation (within ZFC set theory) of that syntactic framework; in the case of the C* algebra approach, each set of (conveniently defined) equivalent representations describe a particular physical system (see below).

Axiomatics for GR: an informal description

We use Suppes predicates in the da Costa-Chuaqui formulation (see Costa (1990), Costa (2007) for the guidelines). This means that our axiomatics will essentially describe a set and some of its subsets in (a model of) ZFC.

We must start from a single basic set, that we take to be R, the real numbers. Out of that set we build our objects. Roughly:
– Out of R we obtain R^n, any finite integer n.
– We also obtain sets of maps $f\colon \mathsf{R}^n \to \mathsf{R}^n$.
– We can restrict those to differentiable maps.
– We can define coordinate domains and build a differentiable manifold,

if we wish to ‚sit' the Einstein equations on a full differentiable real 4-manifold; if not, it is enough to have a neighborhood $U \subset \mathbb{R}^4$.

– We can define all sorts of bundles, principal and associated, over the 4-manifold (we must build a Lie group, which can be made out of similar prescriptions, or we can take a Lie group as a basic set).

In our case we will require the linear group GL(4, R) as our fiber and obtain the usual associated bundles by an easy construction.

– Vectors, tensors, connection forms are defined as cross-sections of those bundles. So we obtain the Ricci-Einstein tensor $G_{\mu\nu} = R_{\mu\nu} - (1/2)\, g_{\mu\nu}R$.
– The motion equations are given by Einstein's equations,

$$R_{\mu\nu} - (1/2)g_{\mu\nu}\, R = kT_{\mu\nu}$$

where $T_{\mu\nu}$ has properties (symmetry, smoothness) induced by the Ricci-Einstein term besides those specific of the matter-energy distribution it describes.

Notice that we have used a Dirac-like formulation in Costa (2007), an idea that was introduced by Penrose in 1960 and then taken up by F. A. Doria in 1975 (Doria 1975). But for our current purposes it is enough to postulate the Einstein equations, with a rhs that satisfies the properties induced by the geometric lhs.)

Axiomatics for GR: a more rigorous presentation

We are not going to give a fully explicit axiomatics for GR. But the main technical points are clear in what follows.

General relativity is a theory of gravitation that interpretes this basic force as originated in the pseudo-Riemannian structure of spacetime. That is to say: in general relativity we start from a spacetime manifold (a 4-dimensional, real, adequately smooth manifold) which is endowed with an pseudo-Riemannian metric tensor of Lorentzian +2 signature. Gravitational effects originate in that tensor.

Given any 4-dimensional, noncompact, real, differentiable manifold *M,* we can endow it with an infinite set of different, nonequivalent pseudo-Riemannian metric tensors with a Lorentzian signature (that is, – + ++). That set is uncountable and has the power of the continuum. (By nonequivalent metric tensors we mean the following: form the set of all such metric tensors and factor it by the group of diffeomorphisms of *M;*

we get a set that has the cardinality of the continuum. Each element of the quotient set is a different gravitational field for M.)

Therefore, neither the underlying structure of M as a topological manifold, nor its differentiable structure determines a particular pseudo-Riemannian metric tensor, that is, a specific gravitational field. From the strictly geometrical viewpoint, when we choose a particular metric tensor g of Lorentzian signature, we determine a g-dependent reduction of the general linear tensor bundle over M to one of its pseudo-orthogonal bundles. The relation

$$g \mapsto g\text{-dependent reduction of the linear bundle to a pseudo-orthogonal principal bundle is 1-1.}$$

We now follow our recipe:

– We take as basic sets a 4-dimensional real differentiable manifold of class C^k, $1 \leq k \leq +\infty$, and the Lorentz pseudo-orthogonal group $O(3,1)$.
– We form the principal linear bundle $L(M)$ over M; that structure is solely derived from M, as it arises from the covariance properties of the tangent bundle over M. From $L(M)$ we fix a reduction of the bundle group $L(M) - P(M, O(3,1))$, where $P(M, O(3,1))$ is the principal fiber bundle over M with the $O(3,1)$ group as its fiber.
  Those will be our derived sets. We therefore inductively define a Lorentzian metric tensor g on M, and get the couple (M, g), which is spacetime.
  (Notice that the general relativity spacetime arises quite naturally out of the interplay between the theory s „general covariance" aspects, which appear in $L(M)$, and – as we will see in the next section – its „gauge-theoretic features, which are clear in $P(M, O(3,1))$.)
– Field spaces are:
  – The first is the set of all pseudo-Riemannian metric tensors, $\mathcal{M} \subset C^k(\odot^2 T_* (M))$, where $C^k(\odot^2 T_* (M))$ is the bundle of all C^k symmetric covariant 2-tensors over M.
  – Also out of M and out of adequate associated bundles we get $\mathcal{A}$, the bundle of all Christoffel connections over M, and $\mathcal{F}$, the bundle of all Riemann-Christoffel curvature tensors over M.
– We need the space of source fields, $\mathcal{I}$, that includes energy-momentum tensors, and arise out of adequate associated tensor bundles over M.

$\mathcal{G}$ is the group of C^k-diffeomorphisms of M.

– If $\mathcal{K}$ is any of the field spaces above, then $\mathcal{K}/\mathcal{G}$ is the space of physically distinct fields.
– Finally the dynamics are given by Einstein's equations (as mentioned there is also a Dirac-like formulation for those, first proposed by R. Penrose in 1960 as a neutrino-like equation; but see Doria (1975)). Of course one can take here the Einstein equations themselves, or a Lagrangian density

$$\mathcal{L} = \sqrt{(-g)}R + \text{matter fields} + \text{cosmological terms}.$$

The quotient $\mathcal{K}/\mathcal{G}$ is the way we distinguish concrete, physically diverse, fields, as for covariant theories one has that any two fields related by an element of $\mathcal{G}$ „are" the „same" field.

The Suppes predicate that describes that theory comes out easily, as a conjuction of formulae that describe the construction of the preceding objects as presented plus a statement of the dynamical law for the theory.

4. Informal interpretations for the axioms of GR

Consider the version of our axiomatization where the motion equations are given out of a Lagrangian density $\mathcal{L}$. Specifically in this case we allow for the interaction of the gravitational field with some matter field that originates the energy-momentum tensor $T_{\mu\nu}$. Then:

– A preliminary caveat: we restrict our interpretations of the axiomatic structure for GR to a set in a model for ZFC which we take to be of sufficiently large cardinality to include all bundles, maps, and function spaces we require in our formal treatment. Thus the model for our axiom system will be a ZFC set.
– Consider the theory that deals with the gravitational field plus test particles endowed with mass that move under the action of gravitation. We can axiomatize that theory as an extension of General Relativity to which we have added axioms that describe test particles. (A test particle gives a zero contribution to the energy-momentum tensor.)
Or we can – roughly – construct a model $\mathbf{M}^{\varepsilon}$ for GR within ZFC itself, but with nonstandard elements, namely nonstandard integers and reals – and whatever is induced by our nonstandard construction. A test

particle p will give a contribution $\varepsilon t^{p}_{\mu\nu}$ to $T_{\mu\nu}$, where $\varepsilon \approx 0$ is infinitesimal. Then map our new theory $\mathbf{M}^{\varepsilon} \rightarrow \mathbf{M}$ via *st*, the map that sends any object over its standard part. M is also a model for GR, and the infinitesimal elements go onto 0 via *st*.

– Consider now the theory that deals with gravitation plus all possible spacetimes. A spacetime is a differentiable real 4-manifold endowed with a Lorentzian metric tensor or, equivalently, with a 1-foliation with differentiable leaves. Recent results (quoted below) allow us to relate the physics on a spacetime to the classes of spacetimes that support it (contrary to what is sometimes asserted, the Einstein equations *per se*, being defined for a specific coordinate patch when we compute their results, cannot determine the global geometry of a spacetime).
However electromagnetic test fields can be used to determine that geometry, as they determine the (closed) spacetime manifold's intersection form (Scorpan 2004), since such test fields determine de Rham's group $H^2(M, R)$, where M is spacetime. The relation is given by:

$$Q_M(F, F') = \int_M F \wedge F'$$

where F, F' are real cocycles (electromagnetic fields) in the de Rham group $H^2(M, R)$, and Q_M is the intersection form. As it is well-known there are relations that restrict the possible intersection forms for differentiable 4-manifolds (Scorpan 2004). F and F' must be taken as test fields, as they are supposed not to contribute to the energy-momentum tensor.
In order to deal with the relation between general relativity in its axiomatized form, and a theory such as above that deals with all possible space-times, one may proceed as in the previous example, and consider a non-standard model, where we will take infinitesimal fields εF, $\varepsilon F'$ etc.

– Specific theories within the realm of GR are usually dealt with by taking some coordinate-dependent formulation of the Einstein field equations instead of a Lagrangian density as its starting point. Such is the case of, say, Gödel's cosmological models, or of the recent inflationary models. The usage of „model" can be made rigorous in this case [Gödel 1949, Lemoine 2008].

– A more complicated example can be given through Cho's formalization of Kaluza-Klein theories (Cho 1975) by considering a principal bundle $P(M, G)$ over spacetime, where G is a Lie group. We can define

a rather natural metric tensor g_{ab} over P that splits as a direct sum of the metric tensor over spacetime plus the Lie group metric tensor obtained out of its structure constants.

We then form the lagrangian density $\mathcal{L}_P = \sqrt{(-g)}\, R_P$, where g is the metric tensor over P and R_P its associated curvature scalar. That lagrangian density then splits as a sum of three terms: Einstein's lagrangian density, the lagrangian density for a gauge field, and a term that depends on the structure constants of G and which may be interpreted as a cosmological term.

That theory can be easily axiomatized as an expanded version of Einstein's theory: namely we take P instead of spacetime as the differentiable manifold where our objects such as R_P and so on are defined; the action integral is then

$$\int_P \mathcal{L}_P d\omega_P,$$

where $d\omega_p$ is a volume element in P. Of course Einstein's gravitational theory is now a particular case in that construction. On the other hand, since the Kaluza-Klein-Cho theory leads to a lagrangian density with an Einstein lagrangian density term, it can be taken as a model for our axiomatic system for general relativity.

– Still another nontrivial example can be found in the Brans-Dicke [Brans 1961] scalar-tensor theory of gravitation. One notices that Einstein's theory can be obtained out of the Brans-Dicke lagrangian density if we put the scalar field $\phi = 1$.

  Now as the Brans-Dicke lagrangian density includes a term $\sqrt{(-g)}\phi R$ (and not just $\sqrt{(-g)}R$), it differs from the Einstein lagrangian density, and it cannot be obtained as a model for our axiomatic system for general relativity.

– A further example is given by the Einstein-Schrödinger theory of the nonsymmetric field. It is a theory with several versions: the original one by Einstein (Einstein 1967) was preceded by theories like the one by Einstein and Straus in 1946, and the contributions by Schrödinger a few years later. The lagrangian density postulated by Einstein for that theory is different from the standard general relativity lagrangian density $\sqrt{(-g)}R$) and as it is well-known there are several problems concerning the theory's physical interpretation – an early effort to correct that difficulty has been done by Sciama (Sciama 1961) in 1961; a recent one appears in Shifflett (Shifflett 2005).

All those examples deal with theories that are accepted as belonging to general relativity, albeit in an extended sense for that designation.

5. Unified treatment for GR

Our examples show the huge variety of theories that fit under the name „general relativity." Can we given a unified treatment to such a disparately-looking field?

We will follow an idea that stems from quantum mechanics, and that is best explified by the GNS construction in the so-called algebraic formulation for quantum mechanics (Emch 1972) already referred to.

Basically, the GNS construction follows from a theorem that can be thus paraphrased: given the algebra A of observables in some quantum theory, and given each state ϕ, then there is a representation f_ϕ of A by self-adjoint operations on some real or complex Hilbert space. Here we will describe a kind of „master theory" L_{GR} so that every theory usually accepted to be, or at least almost all theories considered to be in the realm of GR appear as interpretations of that master theory.

In the present case, we will describe the syntax of a general relativity theory as:

– We start from a differentiable manifold M; as a real manifold, its dimension $n = 4$.
– We consider the tensor bundle over M, which will give us the mathematical objects that describe the physics we are interested in.
– The dynamics is given by a Lagrangian density $\mathcal{L}$ that is a functional of a metric tensor g over M, or of a connection form on M obtained out of some adequate principal bundle.
– There is a splitting $\mathcal{L} = \mathcal{L}_4 + \mathcal{L}_{matter}$, where $\mathcal{L}_4$ describes the strictly gravitational part of the theory.
– There is a group $\mathcal{G}$ of transformations of M so that physically identical objects are in the same orbit under the action of $\mathcal{G}$.

When properly axiomatized call the above the L_{GR} formatization. The Suppes predicate that axiomatizes that general theory can be obtained as in Section 3. Now each example in Section 4 can be given a formal treatment per se; some of them may deal with richer mathematical structures than those in the formalization given in the present section. However in each case there are easy-to-build maps:

$$L_{\mathrm{GR}} \mapsto \text{formalized example}$$

so that we can look at each example as an interpretation of L_{GR}. More precisely: for each possible particular case as described above, L_{GR} is given a model $\mathrm{M}(L_{\mathrm{GR}})$ which is a set in ZFC.

Reducibilities revisited

The general Einstein equations:

$$R_{\mu\nu} - (1/2)\, g_{\mu\nu}\, R = \kappa T_{\mu\nu}$$

are of course a set of nonlinear partial differential equations. Particular solutions are given when we add boundary conditions and restrict the form of the metric tensor $g_{\mu\nu}$. We call the map that goes from the Einstein equations to one of its solutions a *particularization map.*

Now suppose that **M** and **M**' are models for L_{GR}, and moreover that:

1. ι: $\mathbf{M} \subset \mathbf{M}'$, where ι is a ZFC function.

2. For each sentence ϕ of L_{GR} which is true in **M**, the induced $\iota(\phi)$ holds true of **M**'.

We say that **M** is *homogeneously reducible* to **M**' whenever ι is an identity map on its image, or the inverse of a particularization map (see above). Whenever ι isn't such a map we would have a *heterogeneous reduction* in the sense of our constructions. The structure of reducibilities through ι is a tree-like structure.

Of course $\subseteq$ induces a partial ordering on the ZFC set of all such models **M**, **M**', … for our master theory L_{GR}, and we pale at the structural complexity of such a poset. With some abuse of language with respect to **M** and **M**' (which are always taken as interpretations for L_{GR}), we can note $\mathbf{M} \leq \mathbf{M}'$ for any ι-reduction.

Notice that the structure of that (po)set of interpretations for L_{GR} is very far from the layered-cake (linearly ordered set) arrangement mentioned at the beginning of this paper.

Examples

We give three simple examples. Given the condition,

$$\forall x_1,\ldots,\, x_n\, [M_1(x_1,\ldots,\, x_n) \leftrightarrow M_2(x_1,\ldots,\, x_n)]$$

for reducibility, we can put $M_2 = M_1 \circ \iota$, where $\circ$ denotes function composition.

– The Gödel universe < Einstein's general relativity.
  In fact the Gödel universe is obtained from Einstein's theory by the specification of a given metric tensor. It is an example of an homogeneous reduction. Notice that the map ι maps the Gödel metric onto the arbitrary metric $g_{\mu\nu}$ in the standard, empty space, formulation for Einstein's GR. More precisely, we can see Gödel's universe, or the Schwarzschild solution as a particularization of the general relativistic formulation.
– One example for heterogeneous reduction might be: formulate Einstein's GR with the help of a tetrad basis. Then identify the tetrad basis to the Dirac gammas γ_μ, plus the Clifford algebra condition, $\gamma_\mu \gamma_\nu + \gamma_\nu \gamma_\mu = 2g_{\mu\nu}$, where $g_{\mu\nu}$ is GR's metric tensor. The identification map thus described is of course different from an identity map, and ensures that the γ-tetrad formalism is heterogeneously reducible to Einstein s GR. (One uses the γ-tetrad formalism to add Dirac's equation to the Einstein GR lagrangian density.)
– Brans-Dicke theory ⋦ Einstein s general relativity.
  For their dynamic laws do not coincide, as they have different Lagrangian densities.

Notice that the main point in this construction is: we start from semantics in order to define reducibilities.

6. A few applications

The goal of an axiomatic treatment is twofold:
– We first want the axiomatics to help us in claryfiying basic concepts in the theory or theories we handle.
– We wish to consider metamathematical results for those axiom systems.

We now give an example of the second goal in the preceding listing. Consider the following situation: as it is well-known, the vast majority of accepted models for the universe possess a global time structure, that is, they can be differentiably split as $C \times \mathbb{R}$, where C is a compact 3-manifold, and there is a metric tensor that splits accordingly as $g_{ij} \oplus g_{oo}$, $i, j = 1,2,3$, and o the time coordinate, in an adequate coordinate system.

It is also well-known that the so-called Gödel solution (Gödel 1949) has no global time. Then, if we consider all possible spacetime manifolds,

what can we say about the existence or not of global time, from the metamathematical viewpoint?

Now: the first author proved the following theorem (with N. C. A. da Costa):

Proposition 6.1 *If* ZFC *(Zermelo-Fraenkel set theory with the Axiom of Choice) has a model* N *with standard arithmetic part, then there is an arithmetic term f so that:*

1. ZFC $\nvdash \beta = 0$ *and* ZFC $\nvdash \beta = 1$.
2. $\mathbf{N} \models \beta = 0$. □

Then:

Proposition 6.2 *Given any axiomatization for General Relativity through Suppes predicates within ZFC, there is a metric tensor g over* $\mathbb{R}^4$ *with the usual differential structure so that:*

1. ZFC $\nvdash$ „*g has global time.*".
2. ZFC $\nvdash$ „*g doesn't have global time.*".
3. N= „*g doesn't have global time.*"

Proof: Put $g = \beta n + (1 - \beta)h$, where h is Gödel's metric tensor and n is the Minkowski tensor. □

Some criticism has been raised against that construction, as it will turn out that for a model with nonstandard arithmetic we will have that g describes Minkowski space, and it is difficult to make sense of nonstandard objects in such a context. However consider the following argument:

– The axiomatic machinery of ZFC can be implemented as a Turing machine Mzfc so that, for a formal sentence ξ:
  – If ξ is a theorem of ZFC, $M_{ZFC}(\xi)$ stops in the „accept" state (or, say, outputs 1 and stops).
  – If $\neg\xi$ is a theorem of ZFC, $M_{ZFC}(\xi)$ stops in the „reject" state (or, say, outputs 0 and stops).
  – If ξ and $\neg\xi$ are both undecidable, then M_{ZFC} never stops over ξ or $\neg\xi$ as inputs.
– This holds if and only if there is a Diophantine polynomial:

$$p_{ZFC}(\langle \xi, j \rangle, x_1, x_2, \dots,) = 0$$

so that the machine stops over ξ and outputs j if and only if p_{ZFC} has roots. An abuse of language: ξ should be seen here as its Gödel number.

– Now one knows that, if CH is the Continuum Hypothesis, ZFC + CH and ZFC + ¬CH are both consistent (if ZFC is consistent) and have models with standard arithmetic portions, granted our hypothesis about ZFC.
– Put ξ = CH, and use Richardson's map to obtain a β_{CH} so that $\beta_{CH} = 0$ iff CH holds, and $\beta_{CH} = 1$ in ¬CH.
– Write: $g = \beta_{CH}\eta + (1 - \beta_{CH})^{\eta}$

Then:

Proposition 6.3 *Given any axiomatization for General Relativity through Suppes predicates within ZFC, there is a metric tensor g over* R^4 *with the usual differential structure so that:*

1. ZFC $\nvdash$ „*g has global time.*".
2. ZFC $\nvdash$ „*g doesn't have global time.*".
3. **L** $\models$ „*g doesn't have global time.*", *where* **L** *is Gödel's constructive universe.* □

About the preceding result: there will be models with standard arithmetic for both sentences in the undecidable pair we have considered.

Remark 6.4 The preceding examples seem contrived, and are so, in fact. However, since equality for expressions in the language of ZFC is algorithmically undecidable, given an arbitrary *k* there is no decision procedure to check whether $k = g$, for *g* as above.

That is to say, there are infinitely many expressions for metric tensors over R^4 with the standard differential structure so that we will never be able to al-gorithmically decide whether they exhibit global time or not. □

Now the main point in all those axiomatic treatments is:

Proposition 6.5 *If* $\mathcal{A}$ *and* $\mathcal{A}'$ *are axiomatizations for general relativity with the help of Suppes predicates (as described above) then* $\mathcal{A}$ *and* $\mathcal{A}'$ *have the same undecidable sentences.* □

Therefore any such axiomatic treatment of GR will coincide when we consider metamathematical results such as undecidability and incompleteness. The actual choice of a specific axiom system will then become a matter of taste; they will all be equivalent from this viewpoint.

7. Acknowledgments

The authors wish to thank the Reserch Group on Fuzzy Sets and the Advanced Studies Research Group at the Production Engineering Program for support, as well as their chairmen Prof. Carlos A. Cosenza and Prof. R. Bartholo; both authors also thank Prof. Saul Fuks of the History and Philosophy of Science Program for his constant interest in their work. The authors finally acknowledge continuous help from the Brazilian Academy of Philosophy and from its president Prof. J. R. Moderno. F. A. Doria thanks N.C.A. da Costa for many discussions on foundational issues in physics.

Finally F. A. Doria wishes to acknowledge support from CNPq-Brazil, Philosophy Section.

Note

1 Collaborator, Fuzzy Sets Laboratory, PIT.

References

Brans, C. H.; Dicke, R. H., 1961: Mach's Principle and a relativistic theory of gravitation. In: *Phys. Review* 124, pp. 925–935.

Cho, Y. M., 1975: Higher-dimensional unifications of gravitation and gauge theories. In: *J. Math. Physics* 16, pp. 2029–2035.

da Costa, N.C.A.; Doria, F. A.; de Barros, J. A., 1990: A Suppes predicate for general relativity and set-theoretically generic spacetimes. In: *International Journal of Theoretical Physics* 29, pp. 935–961.

da Costa, N.C.A.; Doria, F.A., 1991: Undecidability and incompleteness in classical mechanics. In: *International Journal of Theoretical Physics* 30, pp. 1041–1073.

da Costa, N.C.A.; Doria, F. A., 1992: Suppes predicates for classical physics. In: Echeverrla, J. et al.: *The Space ofMathematics.* Walter de Gruyter.

da Costa, N.C.A.; Doria, F. A., 2007: Janus-faced physics: on Hilbert's 6th Problem. In: Calude, C. (ed.): *Complexity and Randomness: from Leibniz to Chaitin (Festschrift in Honor of G. Chaitin on his 60th Birthday).* World Scientific.

Doria, F. A., 1975: A Weyl-like equation for the gravitational field. In: *Lett. Nuovo Cimento* 14, pp. 480–482.

Einstein, A., 1967: *The Meaning of Relativity*. Methuen.
Emch, G. G., 1972: *Algebraic Methods in Statistical Mechanics and Quantum Field Theory.* John Wiley.
Gödel, K., 1949: An example of a new type of cosmological solution of Einstein's field equations. In: *Rev. Mod. Physics* 21, pp. 447–450.
Klein, C., 2009: Reduction without reductionism: a defense of Nagel on connectability. In: *Phil. quarterly* 59, pp. 39–53.
Kim, J., 2000: *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation.* The MIT Press.
Kim, J., 2005: *Physicalism, or Something Near Enough.* Princeton Monographs in Philosophy.
Lemoine, M.; Martin, J.; Peter, P. (eds.), 2008: *Inflationary Cosmology.* Springer.
Nagel, E., 1979: *The Structure of Science.* Hackett.
Schaffner, K., 1967: Approaches to reduction. In: *Philosophy of Science* 34, pp. 137–147.
Sciama, D. W., 1961: On the interpretation of the Einstein-Schrödinger unified field theory. In: *J. Math. Physics* 2, pp. 472–477.
Scorpan, A., 2004: *The Wild World of 4-Manifolds.* AMS.
Shifflett, J. A., 2005: Einstein-Schrödinger theory using Newman-Penrose tetrad formalism. arxiv gr-qc0403052v6.

Gregor Betz

What range of future scenarios should climate policy be based on? Modal falsificationism and its limitations

Abstract

Climate policy decisions are decisions under uncertainty and are, therefore, based on a range of future climate scenarios, describing possible consequences of alternative policies. Accordingly, the methodology for setting up such a scenario range becomes pivotal in climate policy advice. The preferred methodology of the Intergovernmental Panel on Climate Change will be characterised as „modal verificationism"; it suffers from severe shortcomings which disqualify it for scientific policy advice. Modal falsificationism, as a more sound alternative, would radically alter the way the climate scenario range is set up. Climate science's inability to find robust upper bounds for future temperature rise in line with modal falsificationism does not disprove that methodology, rather, this very fact prescribes even more drastic efforts to curb CO_2 emissions than currently proposed.

Zusammenfassung

Klimapolitische Entscheidungen werden unter Unsicherheit getroffen. Sie basieren daher auf einer Spanne von Zukunftsszenarien, welche mögliche Folgen der alternativen klimapolitischen Handlungsoptionen beschreiben. Wie diese Spanne von Zukunftsszenarien aufzustellen sei, wird damit zu einer zentralen methodologischen Frage der wissenschaftlichen Klimapolitikberatung. Die entsprechende, vom IPCC bevorzugte Methode, die als „modaler Verifikationismus" charakterisiert werden kann, besitzt deutliche Schwächen und ist im Grunde genommen für die wissenschaftliche Politikberatung ungeeignet. Ein modaler Falsifikationismus stellt zwar eine überzeugende Alternative dar, hätte aber eine radikale Umgestaltung der Art und Weise zur Folge, wie die Spanne von Klimaszenarien aufgestellt wird. Die Tatsache, dass die Klimawissenschaften derzeit offenbar nicht in der Lage sind, robuste obere Grenzen für den zukünftigen globalen Temperaturanstieg im Rahmen eines modalen Falsifikationismus anzugeben, führt diese Methodologie nun keineswegs ad absurdum; vielmehr ergibt sich hieraus die Forderung nach noch drastischeren Anstrengungen, die globalen CO_2-Emissionen zu senken.

philosophia naturalis 46 / 2009 / 1

1. Introduction

Within the last decades, it has emerged as a consensus view among climate scientists that mankind is changing earth's climate by altering the composition of the atmosphere, in particular the concentration of greenhouse gases (GHGs) such as carbon dioxide (CO_2), the main GHG, or methane (CH_4).[1] It is equally well established that the atmospheric concentration of CO_2 has reached levels unprecedented in the last 650.000 years: While the CO_2 concentration varied between 180 and 300 ppmv[2] during the ice-age cycles (Siegenthaler et al., 2005), in 2005, the CO_2 concentration attained 379 ppmv (IPCC, 2007a).

Although this represents sufficient ground to consider climate change a serious problem, climate policy decisions ultimately require some sort of foreknowledge: What will the consequences be if we continue business as usual? And what is going to happen if we reduce our GHG emissions? Conditional predictions of future climate change, more than any other type of scientific knowledge, represent the most relevant policy information.[3]

The climate system, however, is not reliably predictable in a deterministic way – a fact widely acknowledged in the climate community and beyond. Even robust probability forecasts are currently out of reach as we will see below. Therefore, reliable foreknowledge about the climate system is modal. Climate policy decisions are decisions under uncertainty, or ignorance (e.g. Knight, 1921). They have to rely on ranges of future scenarios that project the possible consequences of alternative climate policies. These ranges of future climate scenarios represent one of the most important pieces of information in scientific climate policy advice, in particular in the reports of the Intergovernmental Panel on Climate Change (IPCC).

Hence, the question of how this scenario range should be constructed becomes a central issue in the methodology of scientific climate policy advice. It is not merely a question of interest to a philosopher of science. It is also of utmost political relevance since what turns out to be a rational climate policy – given a normative evaluation of the possible consequences – crucially depends on what future scenarios we consider as possible at all when deliberating climate policy measures.

This paper attempts to further our understanding of this issue by reconstructing and evaluating the methods currently employed to set up

the future climate scenario range. Section 2 describes the methods used by the IPCC and their recent evolution. These methods can be summarised in the methodological principle of modal verificationism. The evaluation of this principle in section 3, however, exposes major problems. Thus, section 4 proposes modal falsificationism as an alternative methodology for setting up future scenario ranges. To which extent this principle can be applied in scientific climate policy advice and where its limitations are to be found is discussed in section 5.

2. Modal verificationism as the preferred IPCC methodology

In this section, I will describe the methods the IPCC has used to set up the scenario range to be considered in climate policy, and how these methods have evolved in the last years. Moreover, I attempt to summarise the methodology in a fundamental principle. I shall consider two types of climate scenarios: global mean temperature scenarios which project the global average temperature change in the 21st century, and sea-level rise scenarios projecting global mean sea-level rise for the same period.

1. The IPCC's Third Assessment Report (TAR) 2001

The TAR (IPCC, 2001) set up the range of future temperature change scenarios as follows: Different global GHG-emission scenarios, systematically studied and grouped in separate IPCC reports,[4] provided the basis to calculate alternative developments of the CO_2-concentration in the 21st century. These alternative paths of CO_2-concentration then served as boundary conditions for simulation runs with different complex climate models,[5] so-called general circulation models (GCMs),[6] which calculated the future development of global mean temperature anomaly. Hence, a future climate scenario was obtained by combining an emission scenario with a specific climate model. Figure 1 depicts the range of temperature scenarios thus obtained, the bars at the right hand side of the diagram indicate the range of simulation results for one and the same emission scenario, i.e. the uncertainty range due to the alternative climate models. This is the spectrum of future climate scenarios the TAR prominently displays as the range policy decisions should be based on.

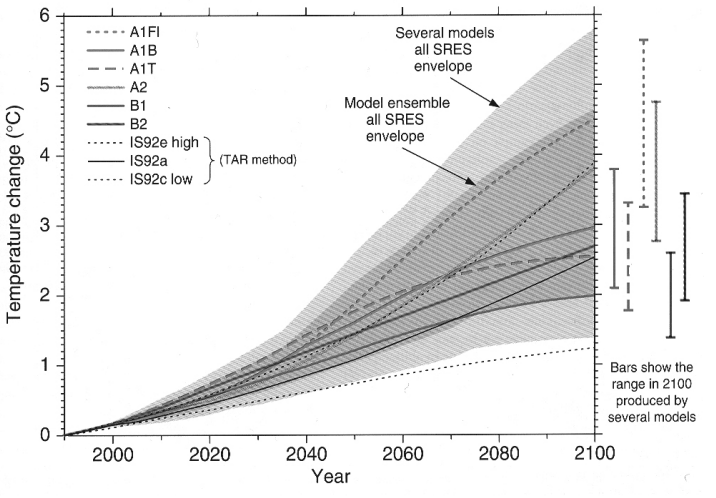


Figure 1: Future global warming scenarios of the TAR. Source: IPCC (2001, p. 555).

While the construction of the sea-level rise range is, in general, very similar to the temperature case, there are some interesting differences. The sea-level rise projections include, in addition to the explicitly modelled sea-level rise, contributions from glaciers and land ice-caps, or the thawing of permafrost. These contributions were not explicit variables in the GCMs, but were calculated separately, based on the simulation results. Thus, a sea-level projection is obtained by combining an emission scenario, a climate model, and a set of assumptions about additional contributing factors. The overall range in figure 2 shows the entire set of scenarios thus calculated. The bars on the right indicate the uncertainty range for each individual emission scenario.

As a first point to note, general circulation models play a central role in these procedures. Those complex models are the main tool for investigating the possible range of future climate change.

Secondly, real climate processes are only taken into account when setting up the scenario range as described above, if they are explicitly modelled in a GCM, or, regarding sea-level rise, at least roughly calculable. Processes which are not modelled are ignored. This obviously holds for the temperature projections; but it is equally true for the sea-level scenarios. Ice-dynamical changes in the Greenland and West Antarctic ice sheets, for instance, i.e. the acceleration of ice flow triggered by climate

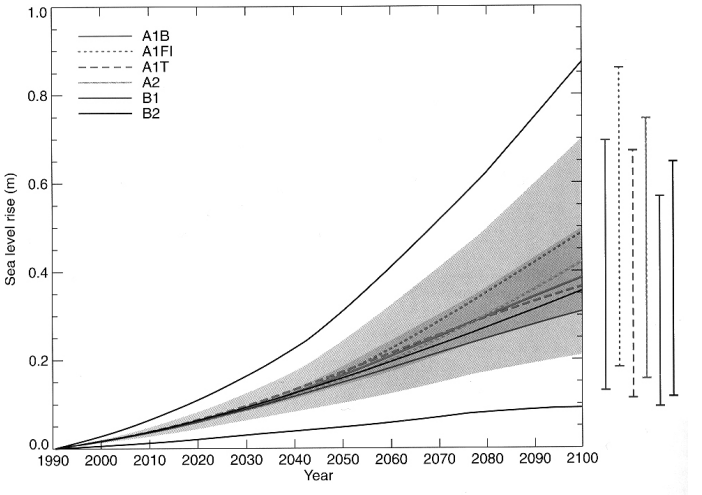


Figure 2: Future sea-level rise scenarios of the TAR. Source: IPCC (2001, p. 671).

change, were so poorly understood that they were not even included in the calculation of the additional contribution to sea-level rise, let alone in the GCMs.

If we assume that the different emission scenarios and climate models are at least possibly correct (i. e. consistent with our physical background knowledge), a model simulation with result R represents a positive proof that R is a possible future climate scenario.[7] Therefore, the IPCC methods can be construed as an implementation of the following methodological principle:

Modal verificationism. It is scientifically shown that a certain statement about the future is possibly true if and only if it is positively shown that this statement is compatible with our relevant background knowledge.[8]

2. 2001–2007: the range of simulation-results explodes

Advances in computing power as well as innovative designs of climate simulation studies (distributing hundreds of simulations on desktop computers worldwide) enabled climate scientists to investigate ever more different versions of climate models.[9] Systematic variation of uncertain parameters led to an explosion of the climate scenario range in the years following the publication of the TAR, or, more specifically, an explosion of the estimated range of climate sensitivity. „Climate sensitivity" refers

to the response of the climate system to a doubling of the CO_2-concentration, and the uncertainty regarding climate sensitivity therefore describes that part of the uncertainty regarding future climate change which is not due to the un-known future evolution the CO_2-concentration. The climateprediction.net study, initiated, amongst others, by the British Hadley Centre, may count as a milestone of this research. Stainforth et al. (2005) report that their model ensemble includes climate models exhibiting a climate sensitivity as high as 11 K (the range of climate sensitivity in the TAR amounts to 2.0–5.1 K (IPCC 2001, p. 560))! Stainforth et al. conclude:

> [Our] results demonstrate the wide range of behavior possible within a GCM and show that high sensitivities cannot yet be neglected as they were in the headline uncertainty ranges of the IPCC Third Assessment Report (for example, the 1.4–5.8 K range for 1990 to 2100 warming).

These results, the authors add, are not too surprising as

> [statistical] estimates of model response uncertainty, based on observations of recent climate change, admit climate sensitivities defined as the equilibrium response of global mean temperature to doubling levels of atmospheric carbon dioxide substantially greater than 5 K. But such strong responses are not used in ranges for future climate change because they have not been seen in general circulation models.

What the authors refer to in this last sentence is nothing but the methodology of modal verificationism which makes GCMs the gold standard of climate change research.

If the IPCC held on to the methods it employed in the TAR, the range of future climate scenarios should encompass a much wider spectrum of scenarios in subsequent reports. Yet, this is not the case, as we shall see in the next section.

3. The IPCC's Fourth Assessment Report (4AR) 2007

Figure 3a shows the range of future temperature scenarios published in the 4AR. Because the graph itself does not display the results from all emission scenarios, the range it indicates is slightly smaller than the corresponding range in the TAR (figure 1); but the bars at its right, comprising all emission scenarios, span a slightly wider range than in the TAR, namely from 1.1–6.4 K. The scenario range does, however, not include the extreme results that have been obtained in previous simulation stud-

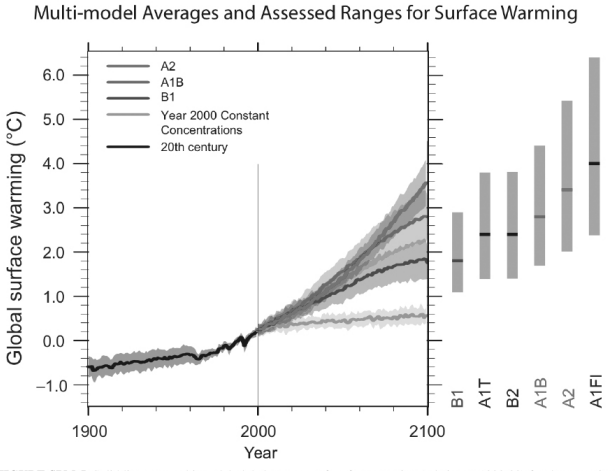
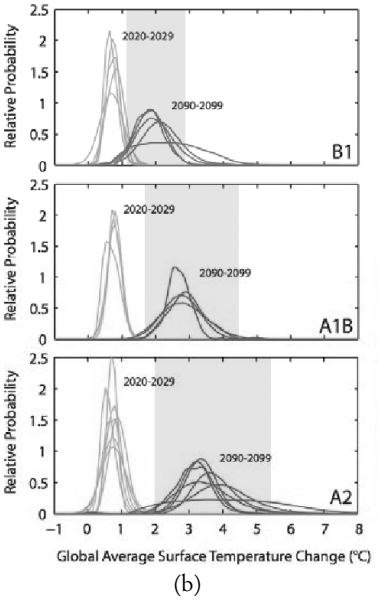


FIGURE SPM-5. Solid lines are multi-model global averages of surface warming (relative to 1980-99) for the scenarios A2, A1B and B1, shown as continuations of the 20th century simulations. Shading denotes the plus/minus one standard deviation range of individual model annual averages. The orange line is for the experiment where concentrations were held constant at year 2000 values. The gray bars at right indicate the best estimate (solid line within each bar) and the *likely* range assessed for the six SRES marker scenarios. The assessment of the best estimate and *likely* ranges in the gray bars includes the AOGCMs in the left part of the figure, as well as results from a hierarchy of independent models and observational constraints. {Figures 10.4 and 10.29}

(a)



(b)

Figure 3: Future global warming scenarios of the 4AR. Gray areas in panel (b) have been added by the author and indicate the ‚likely ranges' from panel (a). Source: IPCC (2007a, p. 14,15).

ies! This is because the IPCC implements a modified methodology in its 4AR: It estimates (and merely reports) the „likely" range of temperature change. The Summary for Policymakers (SPM) defines „likely" as an assessed likelihood, using expert judgement, greater than 66% (IPCC 2007a, p. 4). From the description of figure 3a in the SPM, it is not fully clear how this likelihood has been assessed and how exactly the range was set up. The full report (IPCC, 2007b), however, is more specific, explaining that

> [the reported] range results from an expert judgement of the multiple lines of evidence […], and assumes that the models approximately capture the range of uncertainties in the carbon cycle. The range is well constrained at the lower bound since climate sensitivity is better constrained at the low end […], and carbon cycle uncertainty only weakly affects the lower bound. The upper bound is less certain as there is more variation across the different models and methods, partly because carbon cycle feedback uncertainties are greater with larger warming. (p. 810)

As part of the evidence on which the expert judgement is based, the 4AR uses probabilistic assessments of future climate change obtained with Bayesian methods (IPCC, 2007b, pp. 809f.). Figure 3b displays probability density functions (pdfs) of temperature change as calculated in different studies.[10] These pdfs are typically obtained by assigning prior pdfs to the parameters of a GCM which are then updated given observation in order to obtain a posterior probability for the parameters and, ultimately, climate sensitivity.[11] The application of Bayesian methods to assess likely future climate change is highly problematic as the posterior pdfs significantly depend on the priors chosen.[12] This is the reason why, as noted above, climate policy involves decision making under uncertainty – and not merely under risk.

So, obviously, the IPCC has departed from pure modal verificationism in its 4AR: Some scenarios, which are derived, via model simulations, from the relevant background knowledge, are not considered possible. Due to this methodological adjustment, the IPCC was in a position to reiterate an only slightly modified scenario range compared to its previous reports. But has the IPCC entirely turned its back on modal verificationism? Not quite so, as being the result of a model simulation is actually still a necessary condition for being considered as possible and thence being included in the scenario range. Having in fact tightened the criteria for considering a scenario as possible, the IPCC has merely

dropped the idea that being the result of a model simulation is a sufficient condition for doing so. In sum, the IPCC still seems to implement what one can call weak modal verificationism, i. e. a methodology according to which a future scenario is considered as possible (in policy deliberations) only if it has been positively shown that it is possible.

	4AR	TAR
SRES scenario	„Model-based range excluding future rapid dynamical changes in ice flow“	
	m at 2090–2099 relative to 1980–1999	m at 2100 relative to 1990
B1	0.18–0.38	0.10–0.56
A1T	0.20–0.45	0.12–0.67
B2	0.20–0.43	0.12–0.65
A1B	0.21–0.48	0.13–0.70
A2	0.23–0.51	0.16–0.75
A1FI	0.26–0.59	0.18–0.86

Table 1: Comparison of sea-level rise scenarios of the IPCC's TAR and 4AR. Source: IPCC (2001, 2007a)

Next, we shall consider how the IPCC constructed sea-level rise scenarios in the 4AR. Neither the SPM nor the full 4AR contain a diagram with sea-level scenarios; rather, sea-level change projections are shown in tables. Table 1 compares the projections of the 4AR with those of the TAR. The first thing to notice is that the 4AR provides a „model-based“, not a „likely“ range (as in the temperature case) of future sea-level rise scenarios: Regarding sea-level rise, the 4AR fully implements the methodology of modal verificationism (is this because no explosion of simulation results risked to blast the scenario range?). Moreover, the 4AR ranges are tighter than the TAR ranges throughout. This is due to the fact that, as the IPCC explains, (i) 2090–2099 values instead of 2100 values are used and (ii) the different uncertainties affecting sea-level rise projections (e. g. uncertainties in land ice models, temperature projections, etc.) are combined in a different way (IPCC, 2007b, p. 875).

3. Why modal verificationism is so problematic

The SPM itself openly addresses the problems modal verificationism faces:

> Models used to date do not include uncertainties in climate-carbon cycle feedback nor do they include the full effects of changes in ice sheet flow, because a basis in published literature is lacking. The projections include a contribution due to increased ice flow from Greenland and Antarctica at the rates observed for 1993–2003, but these flow rates could increase or decrease in the future. For example, if this contribution were to grow linearly with global average temperature change, the upper ranges of sea-level rise for SRES scenarios [shown in Table 1 above] would increase by 0.1 m to 0.2 m. (IPCC, 2007a, p. 14 f.)

In other words, because of a lack of knowledge, these processes are simply ignored when setting up the scenario range. This, the SPM acknowledges at least insofar as sea-level rise projections are concerned:

> Larger values cannot be excluded, but understanding of these effects is too limited to assess their likelihood or provide a best estimate or an upper bound for sea-level rise. (p. 15)

But is it admissible to disregard potentially severe consequences and effects merely on the grounds that they are poorly understood? Hardly so, in any case not if the precautionary principle is applied as prescribed by the United Nations Framework Convention on Climate Change (UNFCCC, Article 3, paragraph 1). These ideas thus translate into a first argument against modal verificationism:[13]

(1) Applying modal verificationism in climate policy advice (in order to set up the range of future scenarios) has the effect that potentially severe consequences are disregarded merely on the grounds that they are poorly understood.
(2) Any methodology which has the effect that potentially severe consequences are disregarded merely on the grounds that they are poorly understood is at odds with the precautionary principle.
(3) Thus: Applying modal verificationism in climate policy advice is at odds with the precautionary principle.
(4) The UNFCCC embraces the precautionary principle.
(5) Any methodology used in climate policy advice has to be in line with provisions of the UNFCCC.

(6) Thus: Modal verificationism must not be used in climate policy advice.

A second argument against modal verificationism arises independently of the precautionary principle in the light of observations we have made about that methodology so far. The IPCC admits in the SPM:

> Assessed upper ranges for temperature projections are larger than in the TAR [...] mainly because the broader range of models now available suggests stronger climate-carbon cycle feedbacks. (IPCC, 2007a, p. 14)

So by applying modal verificationism in the year 2007 we learn (ex post) that this very methodology was too restrictive in 2001. This is not just an unproblematic sort of self-correction since the diagnosis given for the overly restrictive range in 2001 applies in 2007 as well: the carbon cycle is still not fully understood. So can't we infer that, given the diagnosis of past failures, the methodology of modal verificationism is too restrictive and thence inappropriate today, too? Here is the argument:

(1) According to modal verificationism, i. e. by its own standards, modal verificationism underestimated the climate scenario range in 2001 because the carbon cycle and its interaction with the climate system was poorly understood.
(2) The carbon cycle, its interaction with the climate system as well as further climate processes are still poorly understood.
(3) Thus: It is likely that modal verificationism applied today will underestimate the climate scenario range (by its own standards), too.
(4) It is inappropriate to apply a methodology in climate policy advice which is likely to underestimate the future scenario range by its own standards.
(5) Thus: It is inappropriate to apply modal verificationism in climate policy advice today.

In sum, the methodology of modal verificationism – both in its pure as well as its weak version – cause a systematic underestimation of the uncertainties and the risks we face. Climate policy advice, when based on this methodology, is biased and tends to play down the gravity of our situation. It is definitely a relevant policy information that it might get even worse than the diagrams set up with modal verificationism suggest. For how bad the worst case really is will affect decisions relating to our major climate policy questions:

Mitigation. How drastic should our efforts to prevent further climate

change be? Specifically, to what extent should we reduce our GHG emissions?

Adaptation. How far reaching should our planning for future climate change be? Specifically, what kind of adaptation measures do we have to initiate today?

Assistance. How generous should our assistance to third countries be? Specifically, what kind of help to mitigate and adapt do developing countries need?

So modal verificationism gets a major policy information wrong. It should, therefore, not be applied in climate policy advice. But what are the alternatives? And, are they viable?

4. Modal falsificationism

The methodological principle of modal falsificationism represents an alternative to modal verificationism. It states:

Modal falsificationism. It is scientifically shown that some hypothetical statement about the future is possibly true as long as it is not shown that this statement is incompatible with our relevant background knowledge, i.e. as long as the hypothetical possibility statement is not falsified.

In other words, an arbitrary storyline (a possibilistic hypothesis) is considered as scientifically possible unless it has been shown to be inconsistent with what we know (i.e. falsified).

Let me briefly pinpoint the loose analogy to classical falsificationism, explaining the name chosen. According to this analogy, scenarios correspond to hypotheses and background knowledge corresponds to empirical data. Unlike modal verificationism, which claims that only positively verified scenarios are to be taken seriously, modal falsificationism, in analogy to classical falsificationism, stipulates to invent arbitrary future scenarios (instead of theories), in a first step, before testing them, in a second step, systematically against the background knowledge. All those creatively constructed future scenarios that have not been falsified shall be accepted (as possible).

The difference between modal verificationism and modal falsificationism can also be thought of as a difference regarding the allocation of burdens of proof. Assume it is contentious whether some scenario *S* should be considered as possible in policy deliberations or not. Then, accord-

ing to modal verificationism, the burden of proof falls on the party that says *S* is possible whereas, according to modal falsificationism, the party denying that *S* is possible carries the burden of proof.

Modal falsificationism, in contrast to modal verificationism, does not conflict with the precautionary principle. Full scientific understanding or the ability to model a certain process or effect is not a precondition anymore for considering that process or effect in policy deliberations. Modal falsificationism is the more cautious and risk averse methodology.

Applying modal falsificationism in scientific climate policy advice has far reaching methodological consequences. Before the next section explores how a positive implementation of that principle might look like, here is a major negative consequence: According to modal falsification, GCMs have no role to play when setting up the range of future climate scenarios. The reason is simple: GCMs describe at most possible features, properties, and dynamics of our climate system. There is no GCM which could be labelled as the „best and correct model" of our climate system.[14] Whatever is deduced from a GCM has thence to be interpreted as a possibility-statement, too. But then it becomes logically impossible to infer from a GCM (together with observational data) that some statement about the future is inconsistent with our background knowledge, i.e. impossible. All one can infer is that the statement is possibly impossible – *if* the GCM were correct, and *if* we knew it, the respective statement would be inconsistent with the relevant background knowledge – which does not suffice to exclude it from the scenario range according to modal falsificationism. The irrelevance of GCMs under modal falsificationism thence illustrates that any application of modal falsificationism must rely on robust models which do not merely represent possible worlds.

5. Applying modal falsificationism

If modal falsificationism, in order to be applicable, requires robust models and if GCMs are not robust, the question arises whether it has any chance of being implemented. Can we exclude any scenarios from the future scenario range at all? How are we supposed to demonstrate that a scenario is inconsistent with our background knowledge? These questions cannot be answered in the abstract. In order to address them in this

section, I will discuss on a case-by-case basis how modal falsification might work and what obstacles have to be overcome. I will consider sea-level rise scenarios first before turning to temperature projections.

1. Sea-level rise scenarios

Can we falsify some sea-level rise scenarios, i.e. demonstrate that they are impossible? The short answer is: yes. If the entire West Antarctic and Greenland ice sheets collapse, all ice caps and glaciers disappear and earth warms by 10 K until 2100, the sum of the individual contributions to sea-level rise would not exceed 6m from West Antarctic ice sheet (IPCC, 2001, p. 678) + 8 m from Greenland ice sheet (IPCC, 2001, p. 648) + 1 m from glaciers and ice caps (IPCC, 2001, p. 648) + 5 m from thermal expansion[15] = 20 m. Any scenario projecting higher sea-level rise can be excluded as inconsistent with our background knowledge: there is simply no source where further rise could stem from.

This, of course, is a ‚conservative estimate'. It lies one to two orders of magnitude above the IPCC's upper bound. So can we do better? Can we, say, exclude a sea-level rise of more than 2 m for this century?

Recently, Stefan Rahmstorf estimated the range of future sea-level rise (Rahmstorf, 2007). His reasoning, in spite of its simplicity, is remarkable because it does not rely on GCMs. So is Rahmstorf's method robust and does it represent an exemplification of modal falsificationism? We shall have a closer look at his argument which proceeds in five steps:

1. The initial, as opposed to the long-run, equilibrium response of sea-level to a given temperature increase is assumed to be proportional to that increase.
2. Sea-level rise in the 21st century is supposed to represent the short-term, initial response to global warming and can thus be approximated by the linear relationship.
3. The linearity hypothesis is tested for 1880–2001 data – and confirmed.
4. The proportionality factor is estimated given the data.
5. The simple linear model is used to generate sea-level rise scenarios on the basis of temperature rise projections. The scenario range obtained is significantly more elevated than the IPCC range (figure 4).

As a first point to note, Rahmstorf's study shows that the GCM range is underestimating the future scenario range and thence represents a further argument against modal verificationism. But does it exemplify modal fal-

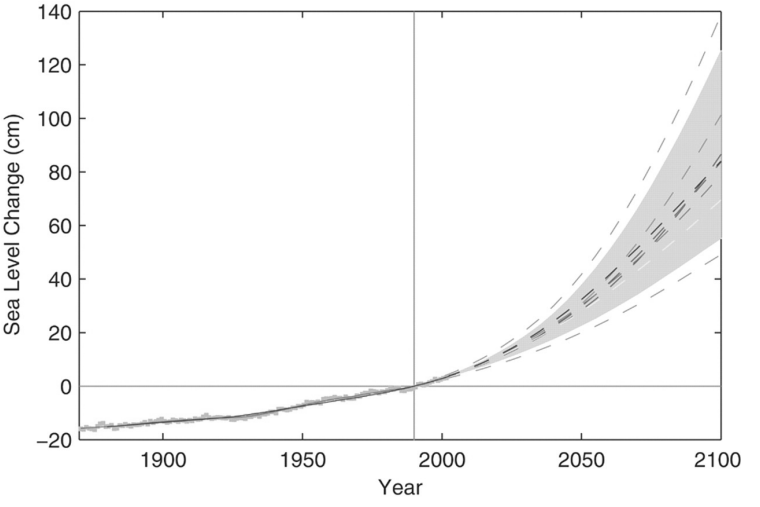


Figure 4: Future sea-level rise scenarios obtained by a semi-empirical approach, based on a linear approximation. The boundary conditions are the same as in the IPCC projections. Source: Rahmstorf (2007).

sificationism, i. e. can we exclude sea-level rise higher than 1.4 m based on this reasoning? Rahmstorf himself is cautious and does not claim so:

> Given the dynamical response of ice sheets observed in recent decades and their growing contribution to overall sea-level rise, this approximation may not be robust. The ice sheets may respond more strongly to temperature in the 21st century than would be suggested by a linear fit to the 20th century data, if time-lagged positive feedbacks come into play (for example, bed lubrication, loss of buttressing ice shelves, and ocean warming at the grounding line of ice streams).

So by assuming linearity, Rahmstorf excludes certain processes and feedbacks on an ad hoc basis. This contradicts modal falsificationism. Moreover, any attempt to provide a tight upper limit for sea-level rise requires that extreme temperature change scenarios can be excluded. So, can we estimate an upper bound for temperature rise in line with modal falsificationism?

Global mean temperature scenarios

As explained in section 2.2, climate sensitivity is the main parameter of our climate system which determines future global warming given a cer-

tain increase in GHG concentrations. Climate sensitivity, however, does not capture all uncertainties related to future temperature anomaly (e.g. uncertainties regarding carbon cycle feedbacks). Still, this section will focus on the question whether we can estimate an upper bound for climate sensitivity because if that were not the case, estimating an upper bound of future temperature rise would be equally impossible. I will present two promising methods to set up a climate sensitivity range in line with modal falsificationism and explain why they ultimately fail. Then, I explore whether there is a general reason for this failure which proofs modal falsificationism inapplicable once and for all – which is not the case. Finally, I discuss the consequences we have to draw from our inability to establish upper bounds for climate sensitivity according to modal falsificationism.

1st method: palaeo-climate data. One way to estimate climate sensitivity without using GCMs is by relying on palaeo-climate data obtained from ice-cores drilled in Antarctica or Greenland (e.g. Lorius et al., 1990). These studies use energy balance models of the climate system that identify on a highly aggregated level the causally relevant factors for global temperature change, namely incoming solar radiation, aerosols, albedo (ratio of reflected solar radiation), GHG concentration. Based on the palaeo-data, a statistical analysis can then be used to estimate the causal strengths („amplitudes") of the different factors, or to reject certain hypotheses about the causal strength of GHGs as inconsistent with the robust model and the data.

Such an analysis assumes that the causal structure of the climate system today is basically the same as during the ice-age cycles. But isn't it possible that the statistical relationships have broken down? Specifically, cloud formation and cloud feedback could have operated quite differently in glacial and inter-glacial periods several ten-thousand years ago as compared to the 21st century with its unprecedented CO_2-concentration. In other words: Nothing seems to guarantee that climate sensitivity is a constant whose current value can be measured given palaeo-data.

To show, by GCM simulation, that the causal structure of the climate system during the ice-age cycles might have been different from today's climate is consistent with our interpretation of GCMs as only possibly correct. That is, basically, what Crucifix (2006) has done by demonstrating that very different sets of assumptions, plugged into a GCM, yield simulation results which are consistent with the available palaeo-data.[16]

Yet another use of palaeo-data is made by Alley (2003) who claims that climate sensitivity might be higher than estimated by the range of current GCMs, arguing that abrupt and relatively rapid climate change is well documented by palaeo-data without being fully captured by GCMs. Although this is a further relevant piece of evidence piling up against modal verificationism, it is not to be confused with an implementation of modal falsificationism.

In sum, palaeo-data, in conjunction with robust climate models, has, so far, not been used successfully to exclude future climate scenarios as impossible.

2nd method: annual variation. The model used for this second method as implemented by Tsushima et al. (2005) is even more aggregated and idealised than the energy-balance model (including aerosols, albedo, etc.) mentioned above. In this model, the earth, as seen from space, is conceived as a blackbody whose temperature adjusts to a given external radiative forcing such that outgoing and incoming energy re-balance. To which extent an increase in earth's surface temperature increases the radiation at the top of the atmosphere is determined by some unspecified feedback process which can be described by a single parameter:

> The sensitivity of the climate is essentially controlled by the so-called feedback parameter, which is the rate of radiative damping of the unit anomaly of the global mean surface temperature due to the outgoing radiation from the top of the atmosphere (TOA). By dividing the radiative forcing of climate by the feedback parameter, one gets the radiatively forced, equilibrium response of global surface temperature. This implies that the stronger is the rate of the radiative damping, the smaller is its equilibrium response to a given radiative forcing. (Tsushima et al., 2005)

In other words, the smaller the feedback-parameter, i. e. the less radiation at the TOA increases with a given surface warming, the more the surface temperature has to increase in order to counter-balance a given radiative forcing and to restore equilibrium of incoming and outgoing radiation at the top of the atmosphere. Tsushima et al. (2005) use the annual variation of radiative forcing in order to estimate the feedback-parameter, and infer that the thus measured parameter describes the response of the climate system to the perturbation by GHGs.

A closer look reveals that the argument is basically an argument by analogy: Its first premiss assumes that the response of the climate system to the annual variation (in radiative forcing) is similar to its response

to a radiative forcing due to an increase in GHG concentration. This implies, essentially, that the feedback parameter is a universal constant of the system. Secondly, based on (i) measurement of incoming solar radiation, (ii) satellite observation of earth's outgoing radiation, and (iii) surface temperature data, we can estimate a confidence interval for the feedback parameter, or exclude certain extreme values. As climate sensitivity is proportional to the reciprocal value of the feedback parameter, this yields a demonstration that certain values of climate sensitivity can be rejected as inconsistent with our current background-knowledge (the aggregate model and the data).

This said, the similarity between the two approaches is obvious. Both rely on highly aggregated models, both assume that climate sensitivity/ the feedback-parameter is a constant, both measure that parameter in a specific situation (ice age cycles/annual variation) and both infer from that measurement the response of the climate system to a future increase in GHG concentration. Like the first method, the second method hinges on the crucial assumption of structural similarity and like for method 1, GCMs show that this similarity might actually not hold:

> Since the pattern of the annual variation of surface temperature [...] differs greatly from that of the global warming simulated by a model, it is quite likely that the rate of the radiative damping of the global mean surface temperature anomaly is significantly different between the two. (Tsushima et al., 2005)

The cloud and albedo feedbacks, in particular, might differ significantly between annual variations of solar radiation and long-run variation of GHG concentration.[17]

So the second approach fails to provide a robust exclusion of future scenarios in line with modal falsificationism, too.

3. Consequences of inability to falsify extreme scenarios

The similarity between the two approaches presented above suggests a general underlying reason for the failure of modal falsificationism. It seems as if aggregated models are inappropriate to falsify possibilistic hypothesis:

(1) Future global mean temperature change causally depends on many detailed feedback processes in the climate system.
(2) Thus: Whether a certain future global mean temperature scenario is possible inferentially depends on how these feedback processes operate.

(3) Thus: We have to know how these feedback processes operate in order to know whether some future global mean temperature scenario is (im)possible.

Only climate models that capture the feedback processes can be used to robustly falsify future climate scenarios. But that means that we are back at the GCMs which, as we have shown, are currently inappropriate in the methodological framework of modal falsificationism. We seem to face a dilemma: With GCMs, we cannot falsify future scenarios; but without GCMs, we can't, either.

Yet, the above argument, in spite of rightly diagnosing the problem which caused the failure of the two methods discussed, is not valid. The inference step from (1) to (2) relies on the following general principle: If the evolution of some variable x causally depends on processes of type *F*, then whether a certain future development of *x* is (im)possible inferentially depends on how *F*-processes operate. This principle is invalidated by counterexamples like the following one. The evolution of global sea-level depends on many complicated feedback processes, in particular dynamical processes in the West Antarctic and Greenland ice sheets. To rule out a sea-level rise of 50 m within the next 100 years as impossible is, nevertheless, inferentially independent of the actual way those ice dynamical processes work; no matter how they operate, the 50 m scenario can definitely be excluded.

Thus, there is no a priori reason to think that we cannot limit the range of future scenarios in accordance with modal falsificationism. As a matter of fact, however, it is currently difficult to robustly falsify extreme temperature change scenarios and, as a consequence, sea-level rise scenarios.[18]

Although Frame et al. (2006) equate excluding future scenarios with falsifying model versions – operating cognitively in the framework of modal verificationism –, they correctly summarise our situation by stressing that, for the time being,

> [the] implication is that the risks associated with any stabilization or equilibrium-based [emission-] scenario remain critically dependent on subjective prior assumptions because of our inability to find directly observable quantities that scale with [climate] sensitivity. Thus any attempt to use [climate] sensitivity as a scientific and policy variable suffers from an inability to find a robust upper bound.

But where does this leave us? I argued that modal falsificationism is the correct methodology for preparing scientific climate policy advice.

And now we learn that it is currently impossible to significantly limit the range of future climate scenarios according to this methodology. Yet, would a scientist's testimony which boils down to saying that we don't know how much temperature might rise in the future count as a scientific policy advice at all? Or is modal falsificationism itself invalidated (by providing useless results)? Frame et al. (2006) propose an interesting argument related to these questions:

> The basic problem with sensitivity is that high sensitivities take a long time to come into equilibrium. This requires that forcings remain constant (or very nearly so) for a long time, possibly hundreds of years. This seems environmentally and economically unlikely.

Moreover:

> Fortunately, stabilization scenarios are not the only options available: indeed, the chances of future generations maintaining a specified concentration of CO_2 indefinitely are probably very small. Other properties of the climate system are much better constrained by observations: for example, the normalized transient climate response (NTCR) which we define as the rate of warming in degrees per year divided by the fractional rate of CO_2 increase per year. […]
>
> NTCR turns out to be much more relevant than sensitivity for a scenario in which forcing reaches the equivalent of 550ppmv CO_2 (double pre-industrial, though the same point holds for higher and lower peak forcings) and then declines by 10% over the ensuing 40 years, continuing to decline thereafter. […]
>
> Given the difficulties in determining climate sensitivity, we suggest that the climate policy community might begin to consider whether the UNFCCC should be interpreted a little differently. Since the ultimate objective of the Convention is to avoid dangerous climate change, and the commitment to stabilize future greenhouse gas concentrations does not preclude stabilization at near pre-industrial levels, ‚stabilize' could be interpreted simply as ‚bring back down' rather than as ‚freeze' at a given elevated concentration in perpetuity […].

This argument might be reconstructed as follows:

(1) We are currently not able to state a robust upper bound for climate sensitivity and long-term temperature projections.
(2) We can robustly estimate the range of short-term, transient climate response to CO_2 increase.
(3) If climate policy aims at reducing GHG concentration to pre-industrial levels in the medium term instead of stabilising it at some higher value, transient climate response instead of climate sensitivity becomes the major relevant climate parameter.

(4) If we cannot robustly state an upper bound for parameter *x* but we can robustly assess parameter *y*, and if *y* (instead of *x*) becomes the major relevant parameter provided policies aim at goal *B* (instead of goal *A*), then policies should aim at goal *B* (instead of goal *A*).
(5) Thus: International climate policy should aim at reducing GHG concentration to pre-industrial levels in the medium term.

Given our epistemic situation, this is a plausible conclusion. Yet, the argument that backs the conclusion is highly problematic. Premiss (4), in particular, has absurd consequences for it makes not only our policies but the goals we shall pursue dependent on our scientific knowledge. The mere fact that it is difficult to acquire a certain type of information is by no means a sufficient condition for changing (possibly radically) our normative goals.

Here comes a more charitable interpretation of the line of reasoning, factoring in the precautionary approach[19] as a normative principle which drives the readjustment of policy measures in the light of uncertainty:

(1) We are currently not able to state a robust upper bound for climate sensitivity.
(2) Modal falsificationism: Scenarios have to be taken into account in policy deliberation unless they are falsified.
(3) Thus: Arbitrarily extreme scenarios regarding climate sensitivity have to be taken into account in climate policy deliberations.
(4) If the scenario range climate policy is based on includes arbitrarily extreme scenarios regarding climate sensitivity, then a climate policy which aims at reducing GHG concentrations to a pre-industrial level in the medium term has the comparatively best worst-case.
(5) Precautionary principle (maximin rule): The climate policy with the comparatively best worst-case should be implemented.
(6) Thus: International climate policy should aim at reducing GHG concentration to pre-industrial levels in the medium term.

Note that the reduction of GHG concentration to pre-industrial levels in the medium term has the comparatively best worst case because it does not allow the climate system to adjust fully to a new (possibly very much warmer) equilibrium state but re-establishes the pre-industrial equilibrium. (Premiss (4) thus assumes that we can robustly rule out worst cases triggered by a quick reduction of GHG concentration to pre-industrial levels.)

Coming back to the critical questions raised above, our inability to

robustly falsify extreme climate scenarios does not represent a falsification of modal falsificationism and is not a useless result at all; rather, it is itself a relevant policy information based on which, in the light of the precautionary principle, a rational climate policy decision can be taken. The corresponding policy recommendation, to reduce the GHG concentration to its pre-industrial level in the medium term, is significantly stronger and more far-reaching than the policy aims of recent reports such as the Stern Review (recommending stabilisation at 450–550 ppm CO_2 equivalent (Stern, 2007)) or a policy paper by the German Advisory Council on Global Change (recommending stabilisation at 450 ppm CO_2 equivalent (WBGU, 2007)).

Although the principle of modal falsificationism and the precautionary principle figure as premisses in one and the same argument above, both can be clearly separated, and do not entail each other. Modal falsificationism, again, is a methodology for setting up the future scenario range under uncertainty. The precautionary principle is a decision principle under uncertainty. The former can, yet need not be combined with the latter. The argument above has shown that it is not impossible to justify a climate policy decision based on the (unbounded) scenario range we currently obtain when applying modal falsificationism: it is possible when we base our decision on the precautionary principle. It might, however, turn out to be impossible when we, or rather the democratically legitimised decision makers, decide to base their climate policies not on the precautionary approach but on some other principle.

6. Concluding remarks

This paper has strictly focused on *global* climate scenarios. Many climate policy decisions, however, require projections of climate change on a regional level. Here, in terms of our ability to robustly falsify future scenarios, things are even worse. The argument that detailed knowledge about climate processes is necessary to falsify regional climate scenarios seems to me much more plausible than the corresponding argument for global scenarios. Moreover, even if the range of global future scenarios were limited, these limits would not easily translate into limits for regional climate change which could turn out to be much more extreme than the global average. So, if the range of future global climate scenarios

were significantly limited and if it were, for example, robustly shown that a stabilisation at 400 ppm would not trigger a global mean warming of more than 2 K, it could still be rational to aim at a GHG reduction to pre-industrial levels in the light of the uncertainty governing regional climate scenarios.

The main findings of our reasoning can now be summarised as follows:

– The way the IPCC currently sets up and reports ranges of future climate scenarios, i.e. in agreement with modal verificationism, is biased.
– Modal falsificationism represents a sound alternative to modal verificationism and should be adopted in climate policy advice.
– As of our current knowledge about the climate system, we are not in position to identify a single worst-case climate change scenario.
– This does not prevent one from taking rational policy decisions: On the background of this situation, the precautionary principle, e.g., calls for drastic curbs of our GHG emissions and a stabilisation of the CO_2 concentration at pre-industrial levels as soon as possible.

Notes

1 E.g. Oreskes (2004) or Rahmstorf and Schellnhuber (2006).
2 I.e. parts per million volume; accordingly, CO_2 made up 0.018–0.03 percent of the atmosphere.
3 For the way different kinds of foreknowledge can rationally fuel a policy-decision compare also part III in Betz (2006).
4 The so-called SRES (Special Report on Emission Scenarios) scenarios (IPCC, 2000).
5 Or, more specifically, with a simple climate model which was tuned to different general circulation models. (IPCC, 2001, p. 577)
6 These are sometimes, and more precisely, referred to as atmosphere-ocean general circulation models (AOGCMs), provided the models couple a three-dimensional atmosphere with a three-dimensional ocean module.
7 This assumption is, however, not unproblematic. By making use of so-called flux adjustments, some climate models violate principles of energy- and mass-conservation. Thus, it is not clear in how far these climate models are even physically possible. Because these models lack a track-record of empirical successes, such contrary-to-fact assumptions cannot be justified pragmatically – as proposed by Winsberg (2003, 2006) –, either.
8 I used to refer to this methodology as „modal inductivism" (cf. Betz 2009), but an anonymous reviewer has convinced me that „modal verificationism" is more appropriate and telling a name.
9 See for instance http://www.climateprediction.net.

10 The gray shading in 3b, which I have introduced, demarks the intervals that are indicated by the bars in 3a and obtained from expert judgement.
11 Compare, for instance, the study by Hegerl et al. (2006), see also IPCC (2007b, pp. 798 f.).
12 For a more detailed exposition of this argument see Betz (2007).
13 Generally, there is a risk of begging the question when arguing against modal verificationism, namely insofar other definitions of what is possible are used as a premiss; e.g. one cannot, without begging the question, assume in such an argument that dynamic changes in Greenland ice flow are possible (and infer that modal verificationism is inappropriate as it does not say so), because that very assumption presumes some other rule for what has to be considered as possible which conflicts with modal verificationism.
14 Cf. Parker (2006) for an analysis of this plurality in climate modelling.
15 This is a conservative upper boundary of a rather well-understood contributing factor, because (i) 21st century sea-level rise will only be a fraction of long-term sea-level rise, and (ii) no model-simulation, not even for extreme CO_2- scenarios, yields 5m sea leve rise in the long run (IPCC, 2001, p. 676).
16 Moreover, this example shows that GCMs do not become entirely worthless under modal falsificationism.
17 Accordingly, Tsushima et al. (2005) write that the „rate of radiative damping of local surface temperature anomaly is similar to the damping of the annual variation in global surface temperature under clear sky. Therefore, it is likely that the rate of the radiative damping of global surface temperature variation under clear sky is similar between the annual variation and global warming despite the difference in pattern. On the other hand, a similar statement may not be made for the albedo-, and cloud feedback."
18 Discussing the practicality of modal falsificationism, it is worth noting that it does not appear to be a primary aim of climate change research to robustly falsify extreme scenarios. In other words: modal falsificationism is currently hardly implemented. Instead, the further sophistication of GCMs seems to be a research priority. Thus, if climate research started to invest significant cognitive resources in implementing modal falsificationism, new methods for falsifying future scenarios might be developed. Only after modal falsificationism has been given a fair try, we will be in a position to make a final judgement regarding its practicality.
19 It is interpreted here in its core version (cf. Gardiner 2006); see for different interpretations of the precautionary principle Morris (2000).

References

Alley, R. B., 2003: Palaeoclimatic insights into future climate challenges. In: *Philosophical Transactions of the Royal Society of London Series A – Mathematical and Physical Engineering Sciences* 361(1810), pp. 1831–48.

Betz, G., 2006: *Prediction or Prophecy? The Boundaries of Economic Foreknowledge and Their Socio-Political Consequences*. DUV.

Betz, G., 2007: Probabilities in climate policy advice: A critical comment. In: *Climatic Change* 85(1-2), pp. 1–9.

Betz, G., 2009: Underdetermination, Model-ensembles and Surprises: On the Epistemology of Scenario-analysis in Climatology. In: *Journal for General Philosophy of Science* 40(1), pp. 3–21.

Crucifix, M., 2006: Does the last glacial maximum constrain climate sensitivity? In: *Geophysical Research Letters* 33(18), L18701.

Frame, D. J.; Stone, D. A.; Stott, P. A.; Allen, M. R., 2006: Alternatives to stabilization scenarios. In: *Geophysical Research Letters* 33(14), p. L14707.

Gardiner, S. M., 2006: A Core Precautionary Principle. In: *The Journal of Political Philosophy* 14(1), pp. 33–60.

Hegerl, G. C.; Crowley, Thomas J.; Hyde, William T.; Frame, David J., 2006: Climate sensitivity constrained by temperature reconstructions over the past seven centuries. In: *Nature* 440(7087), pp. 1029–32.

IPCC, 2007a: *Climate Change 2007. The Physical Science Basis. Summary for Policymakers*. In: IPCC (2007b).

IPCC, 2007b: Climate Change 2007. *The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

IPCC, 2000: *Emissions Scenarios. Special Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

IPCC, 2001: Climate Change 2001. *The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Knight, F., 1921: *Risk, Uncertainty and Profit*. Houghton Mifflin.

Lorius, C.; Jouzel, J.; Raynaud, D.; Hansen, J.; Le Treut, H., 1990: The ice-core record: climate sensitivity and future greenhouse warming. In: *Nature* 347(6289), pp. 139–45.

Morris, J., 2000: Defining the precautionary principle. In: *Rethinking Risk and the Precautionary Principle*. Butterworth Heinemann.

Oreskes, N., 2004: The scientific consensus on climate change. In: *Science* 306 (5702), p. 1686.

Parker, W. S., 2006: Understanding pluralism in climate modeling. In: *Foundations of Science* 11(4), pp. 349–68.

Rahmstorf, S., 2007: A semi-empirical approach to projecting future sea-level rise. In: *Science* 315(5810), pp. 368–70.

Rahmstorf, S.; Schellnhuber, H. J., 2006: *Der Klimawandel*. C.H. Beck.

Siegenthaler, U.; Stocker, T. F.; Monnin, E.; Lüthi, D.; Schwander, J.; Stauffer, B.; Raynaud, D.; Barnola, J.-M.; Fischer, H.; Masson-Delmotte, V.; Jouzel, J., 2005: Stable carbon cycle? Climate relationship during the late Pleistocene. In: *Science* 310 (5752), pp. 1313–17.

Stainforth, A.; Aina, T.; Christensen, C.; Collins, M.; Faull, N.; Frame, D. J.; Kettleborough, J. A.; Knight, S.; Martin, A.; Murphy, J. M.; Piani, C.; Sexton, D.; Smith, L. A.; Spicer, R. A.; Thorpe, A. J.; Allen, M. R., 2005: Uncertainty in predictions of the climate response to rising levels of greenhouse gases. In: *Nature* 433(7024), pp. 403–6.

Stern, N., 2007: *The Economics of Climate Change. The Stern Review*. Cambridge University Press.

Tsushima, Y.; Abe-Ouchi, A.; Manabe, S., 2005: Radiative damping of annual variation in global mean surface temperature: Comparison between observed and simulated feedback. In: *Climate Dynamics* 24(6), pp. 591–7.

WBGU, 2007: *New Impetus for Climate Policy: Making the Most of Germany's Dual Presidency. Policy paper 5*. In: *German Advisory Council on Global Change*. http://www.wbgu.de.

Winsberg, E. 2003: Simulated experiments: Methodology for a virtual world. In: *Philosophy of Science* 70(1), pp. 105–25.

Winsberg, E., 2006: Models of success vs. the success of models: Reliability without truth. In: *Synthese* 152(1), pp. 1–19.

Verzeichnis der Autoren

Jun.-Prof. Dr. Gregor Betz
Institut für Philosophie
Universität Stuttgart
Seidenstraße 36
70174 Stuttgart

Francisco Antonio Doria
Professor Emeritus
Production Engineering Program
and Advanced Studies Research
Group COPPE
Federal University
21945 Rio de Janeiro
Brasilien

Manuel Doria
Junior Researcher
Fuzzy Sets Laboratory
Production Engineering Program
COPPE
Federal University
21945 Rio de Janeiro
Brasilien

Prof. Dr. Dr. Brigitte Falkenburg
Institut für Philosophie
Technische Universität Dortmund
Emil-Figge-Str. 50
44221 Dortmund

Prof. Dr. Andreas Hüttemann
Philosophisches Seminar
Westfälische Wilhelms-Universität Münster
Domplatz 23
48143 Münster

Prof. Dr. Olaf L. Müller
Humboldt-Universität zu Berlin
Institut für Philosophie
Unter den Linden 6
10099 Berlin

Dr. Maria E. Kronfeldner
Abteilung Philosophie
Universität Bielefeld
Postfach 10 01 31
33501 Bielefeld

Matthias Rang
Phänomenologie und Didaktik
der Physik
Leuphana Universität Lüneburg
21335 Lüneburg

Prof. Dr. Manfred Stöckler
Institut für Philosophie
Universität Bremen
Postfach 330440
28334 Bremen

Prof. Dr. Sven Walter
Institute of Cognitive Science
Universität Osnabrück
Albrechtstraße 28
49069 Osnabrück

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. 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 und als WORD-File auf Diskette oder CD 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 3 Belegexemplare des jeweiligen Heftes.

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