

## Vorwort

Mein Interesse an philosophischen Gedankenexperimenten und ihrer Methodologie erwuchs in erster Linie aus Ulrich Kühnes Enthusiasmus für Gedankenexperimente in den Naturwissenschaften und den skeptischen Fragen meiner Studierenden in meiner ersten Vorlesung über personale Identität an der Universität Tartu.

Ulrich arbeitete damals an seiner Dissertation über Gedankenexperimente in den Naturwissenschaften. Seine Dissertation ist seit Kurzem im Buchhandel erhältlich (Kühne [176]). Ich habe von Ulrichs gründlicher historischer Arbeit an vielen Stellen profitiert, auch wenn ich letztlich in manchen Punkten zu anderen Schlussfolgerungen komme. Diese Unterschiede sollte man vielleicht nicht überbewerten, mir geht es hauptsächlich um Gedankenexperimente in der *Philosophie*.

Dass *dieses* Thema problematisch ist, wurde mir erst so richtig klar, als meine Studierenden in Tartu nach dem dritten Gedankenexperiment zu Gehirntausch und Teleportation nachbohrten, weshalb eine Argumentation, die auf ausgedachte und abwegige Situationen Bezug nimmt, überhaupt ernst genommen werden sollte. Ich wusste damals nur eine relativ unbefriedigende Antwort zu geben. Ich hoffe, dass die Antwort, die ich hier präsentiere, etwas besser ist.

Wenn es bisweilen Spaß machte, die Methodologie von Gedankenexperimenten zu untersuchen, dann weil es Kollegen gab, die das Interesse an diesen seltsamen Gebilden teilten und mit Rat, Preprints, Reprints und freundlichem Interesse zur Verfügung standen. Ich möchte an dieser Stelle davon zumindest denen explizit danken, die mir jetzt gerade einfallen:

David Atkinson, Stefan Bagusche, Volker Beeh, Dieter Birnbacher, Mike Bishop, Martin Boeltau, Sacha Bourgeois-Gironde, Manuel Bremer, Marc Breuer, Jim Brown, Axel Bühler, Filip Buekens, Ross Cameron, David Chalmers, Dan Lopez de Sa, Hans Dooremalen, Klaus-Jürgen Düsberg, Jerry Fodor, Dagfinn Føllesdal, Stacie Friend, Tamar Szabó Gendler, Güven Güzeldere, Johannes Haag, Sören Häggqvist, Bob Hale, Reese Heitner, Lloyd Humberstone, Carmen Kaminsky, Christoph Kann, Kathy Kannuck, Philipp Keller, Jens Kipper, Hartmut Kliemt, Ulrich Kühne, Jochen Lechner, Joseph Levine, Edouard Machery, Peter Menzies, Tim De Mey, Desiderio Murcho, Nancy Nersessian, Albert Newen, Martine Nida-Rümelin, John D. Norton, Jeanne Peijnenburg, John Perry, Oliver Petersen, Tom Polger, Lothar Ridder, Sven Rosenkrantz, Marcus Roßberg, Barry Smith, Harry Soodak, Thomas Spitzley, Frank Stäudner, Michael Stöltzner, Richard Swinburne, Alois Weber, Erik Weber, Markus Werning

Dieses Buch ist eine überarbeitete Fassung meiner Dissertation, die 2005 an der Philosophischen Fakultät der Heinrich-Heine Universität Düsseldorf angenommen wurde. Dieter Birnbacher danke ich für viele anregende Gespräche und seine konstruktive Kritik, wie auch für seine freundschaftliche Unterstützung. Ebenso möchte ich Christoph Kann für seine Unterstützung besonders danken. Bei der Drucklegung standen mir Jochen Lechner, Michael Preuss und Reinold Schmücker mit Rat und Tat zur Seite. Den Umschlag gestaltete Jörg Bennert. Für die anhaltende Unterstützung durch meine Eltern kann ich nicht genug danken. Ihnen ist diese Arbeit gewidmet.

# Inhalt

VORWORT.....	5
INHALT .....	7
1. EINLEITUNG .....	11
1.1 GEDANKENEXPERIMENTE IN DER PHILOSOPHIE .....	12
1.1.1 Philosophische Gedankenexperimente .....	14
1.1.2 Das Problem .....	15
1.1.3 Aufgabenstellung.....	16
1.2 GLIEDERUNG UND HAUPTERGEBNISSE .....	18
1.2.1 Das Gedankenexperiment in der Wissenschaftstheorie .....	19
1.2.2 Eine allgemeine Theorie des Gedankenexperiments.....	20
1.2.3 Gedankenexperimente in der Philosophie.....	22
1.2.4 Kritik an philosophischen Gedankenexperimenten im Allgemeinen .....	23
1.2.5 Der modale Status philosophischer Targetthesen.....	24
1.2.6 Kritik an philosophischen Gedankenexperimenten gegen Realdefinitionen.....	25
1.2.7 Vorstellbarkeit und Möglichkeit .....	27
1.2.8 Die Leistungsfähigkeit philosophischer Gedankenexperimente.....	29
2. DAS GEDANKENEXPERIMENT IN DER WISSENSCHAFTSTHEORIE DER NATURWISSENSCHAFTEN .....	31
2.1 ‚GEDANKENEXPERIMENT‘ BEI ERNST MACH.....	32
2.1.1 Ernst Mach und der Beginn der Begriffsgeschichte.....	32
2.1.2 Machs Naturalismus, Empirismus und das Prinzip der Denkökonomie.....	40
2.1.3 Newtons Eimerversuch .....	41
2.1.4 Gedankenexperiment nach Mach .....	47
2.2 ‚GEDANKENEXPERIMENT‘ BEI MACHS ZEITGENOSSEN, BEI HEMPEL UND POPPER.....	49
2.2.1 Die negative Rezension Machs.....	49
2.2.2 Die Ausblendung des Entdeckungszusammenhangs .....	51
2.2.3 Die Wiedergeburt des Themas aus der Entwicklung der Physik .....	53
2.3 ‚GEDANKENEXPERIMENT‘ BEI KUHN .....	61
2.3.1 Die Standardauffassung und das Paradox des Gedankenexperiments .....	64
2.3.2 Begriffswandel bei Kindern als Modell für wissenschaftlichen Begriffswandel ..	65
2.3.3 „In meinen Ohren klingt das wie ein schroffer Widerspruch. Was meint Ihr, Signore Simplicio?“ .....	65
2.3.4 Die Inadäquatheit der Standardauffassung.....	69
2.3.5 Kuhns Funktion für das Gedankenexperiment .....	70
2.3.6 Konzessionen an den Gegner?.....	72

3. EINE ALLGEMEINE THEORIE DES GEDANKENEXPERIMENTS .....	73
3.1 DREI TYPEN VON GEDANKENEXPERIMENTEN .....	74
3.1.1 Klärende Gedankenexperimente.....	75
3.1.2 Funktionale Gedankenexperimente.....	77
3.1.3 Gedankenexperimente zur Überzeugungsänderung.....	79
3.2 GEDANKENEXPERIMENTE ZUR ÜBERZEUGUNGSÄNDERUNG .....	80
3.2.1 Psychologische Modelle des „Laboratory of the Mind“ .....	80
3.2.2 Gedankenexperimente als Paradoxien .....	84
3.2.3 Gedankenexperimente als Argumente gegen „Targetthesen“ .....	89
3.2.4 Kuhn und Conflict Vagueness.....	92
3.2.5 Sind Gedankenexperimente Argumente?.....	93
3.2.6 Die Rechtfertigung von (physikalischen) Gedankenexperimenten durch die Evolutionäre Erkenntnistheorie.....	97
3.2.7 Substantielle und akzidentelle Eigenschaften von Gedankenexperimenten .....	100
3.3 DIE ELIMINIERBARKEIT VON GEDANKENEXPERIMENTEN .....	102
3.3.1 Die Norton-Brown Debatte und die Frage der Eliminierbarkeit.....	102
3.3.2 Nichteliminierbarkeit .....	105
3.3.3 Hintergrundannahmen.....	112
4. GEDANKENEXPERIMENTE IN DER PHILOSOPHIE .....	114
4.1 GEDANKENEXPERIMENTE IN DER THEORETISCHEN PHILOSOPHIE .....	116
4.1.1 Searles chinesisches Zimmer.....	116
4.1.2 Mary, die Neurophysiologin.....	119
4.1.3 Gettier-Fälle .....	120
4.1.4 Zwillingserde.....	122
4.1.5 Gehirne im Tank.....	123
4.1.6 Zombies .....	127
4.1.7 Das Reduplikationsargument .....	130
4.2 GEDANKENEXPERIMENTE IN DER PRAKTISCHEN PHILOSOPHIE.....	132
4.2.1 Thomsons Geiger.....	132
4.2.2 Funktionieren Gedankenexperimente in der Praktischen Philosophie anders? .....	135
4.3 WEITERE ARTEN VON GEDANKENEXPERIMENTEN IN DER PHILOSOPHIE.....	144
4.3.1 Eine weitere Art von Gedankenexperimenten in der Praktischen Philosophie?.....	145
5. KRITIK AN PHILOSOPHISCHEN GEDANKENEXPERIMENTEN IM ALLGEMEINEN .....	153
5.1 ES GIBT KEINE ERKLÄRUNG, WESHALB GEDANKENEXPERIMENTE IN DER PHILOSOPHIE FUNKTIONIEREN SOLLTEN .....	154
5.2 WIR WISSEN EINFACH NICHT, WAS WIR SAGEN WÜRDEN .....	156
5.2.1 Es gibt keine Tatsache in Bezug darauf, was wir sagen würden .....	156
5.2.2 Obschon es Tatsachen gibt, die festlegen, was wir in Bezug auf eine kontrafaktische Situation sagen würden, können wir sie durch die Methode des Gedankenexperiments nicht ergründen.....	159
5.3 PHILOSOPHISCHE GEDANKENEXPERIMENTE SIND ZIRKULÄR ODER WIDERSPRÜCHLICH.....	164
5.3.1 Zwei Indikatoren für schlechte Gedankenexperimente.....	165
5.3.2 Schlechte Gedankenexperimente in der Physik .....	168
5.3.3 Sind Gedankenexperimente in der Physik <i>so</i> schlecht? .....	170
5.3.4 Sind Gedankenexperimente in der Philosophie <i>so</i> schlecht? .....	171
5.3.5 Philosophische Gedankenexperimente können nicht empirisch entschieden werden.....	175

6. DER MODALE STATUS PHILOSOPHISCHER TARGETTHESEN .....	177
6.1 VERSCHIEDENE MODALITÄTEN .....	178
6.2 DEFINITIONEN UND IHRE MODALITÄTEN .....	182
6.2.1 Nominaldefinitionen .....	183
6.2.2 Realdefinitionen .....	184
6.2.3 Begriffsexplikationen .....	188
7. GEDANKENEXPERIMENTE GEGEN REALDEFINITIONEN .....	193
7.1 KRITIK AN GEDANKENEXPERIMENTEN GEGEN SACHANALYSEN .....	194
7.1.1 Gedankenexperimente sind unzuverlässig, wenn es um einen Begriff geht, der keine natürliche Art bezeichnet .....	194
7.1.2 Gedankenexperimente sind unzuverlässig, wenn es um einen Begriff geht, der eine natürliche Art bezeichnet .....	202
7.2 DAS APRIORI UND DAS METAPHYSISCHE .....	204
7.2.1 Modaler Rationalismus .....	206
7.2.2 Die Putnam Orthodoxie .....	229
7.2.3 Moderater Deskriptivismus .....	233
7.3 KRITIK AN GEDANKENEXPERIMENTEN GEGEN BEDEUTUNGSANALYSEN .....	237
7.3.1 Die Relativität von Intuitionen .....	237
7.3.2 Die Repräsentationsstruktur unserer Begriffe .....	251
8. VORSTELLBARKEIT UND MÖGLICHKEIT .....	263
8.1 DIE GRUNDLAGE UNSERES MODALEN WISSENS .....	265
8.1.1 Methodologie des Gedankenexperiments bei Mach, Popper und Kuhn .....	265
8.1.2 Methodologie und Erkenntnistheorie des Gedankenexperiments .....	266
8.2 VORSTELLBARKEIT ALS KRITIKGRUNDLAGE AN DER TARGETTHESE .....	272
8.2.1 Philosophische Vorstellbarkeit? .....	274
8.2.2 Der Konfusionsvorwurf und die Glaubbarkeit von p .....	276
8.2.3 Der petitio-Vorwurf .....	279
8.2.4 „Positive“ Vorstellbarkeit .....	281
8.2.5 Mögliche Kritik an einer solchen Charakterisierung positiver Vorstellbarkeit .....	283
8.3 KONSISTENTISMUS UND „NEGATIVE“ VORSTELLBARKEIT .....	285
8.3.1 Formale Logik ist unzureichend? .....	286
8.3.2 Bedeutungspostulate? .....	287
8.3.3 Welche Logik? .....	288
8.3.4 Fallibilität? .....	289
8.3.5 Verschiedene Ebenen der Betrachtung .....	291
8.3.6 Warum der Konsistentismus nicht funktionieren kann .....	293
9. DIE LEISTUNGSFÄHIGKEIT VON GEDANKENEXPERIMENTEN IN DER PHILOSOPHIE .....	319
9.1 PHILOSOPHIE VS. PSYCHOLOGIE .....	322
9.1.1 Das Gedankenexperiment als Einsicht in Naturzusammenhänge .....	322
9.1.2 Gedankenexperimente als Daten für die Psychologie .....	324
9.1.3 Gedankenexperimente zur Explikation theoretischer Zusammenhänge .....	325
9.2 ARBEITSTELUNG IN DER PHILOSOPHIE .....	331
9.3 ALTERNATIVE KONZEPTIONEN .....	335
10. LITERATURVERZEICHNIS .....	337



# 1. Einleitung

*A logical theory may be tested by its capacity for dealing with puzzles, and it is a wholesome plan, in thinking about logic, to stock the mind with as many puzzles as possible, since these serve much the same purpose as is served by experiments in physical science.*

Bertrand Russell: ‚On Denoting‘ (Russell [283])

Das Thema dieses Buches ist die methodologische Rolle von Gedankenexperimenten in der Philosophie. Es geht um folgende Fragen: Was verbindet/unterscheidet Gedankenexperimente in Philosophie und Naturwissenschaft? Was können philosophische Gedankenexperimente leisten? Lassen sich methodologische Richtlinien formulieren, die ein erfolgreiches Gedankenexperimentieren wahrscheinlich machen (oder sollte vom Gebrauch dieser Methode in der Philosophie generell abgeraten werden)?

Gedankenexperimente fehlen in keinem systematischen Teilbereich der modernen Philosophie. Wir kennen aus verschiedenen Spezialgebieten eine Reihe seltsamer Szenarien, die wir für<sup>1</sup> Gedankenexperimente verwenden, sei es das „Chinesische Zimmer“ von Searle, der „weltberühmte Geiger“ von Thomson, die „Zwillingerde“ und die „Gehirne im Tank“ von Putnam, verschiedene Teletransporterunfälle aus der Debatte um personale Identität und die Zombies aus der Philosophie des Geistes.

Obwohl diese Science Fiction Episoden in der Philosophie zur Steigerung des Unterhaltungswerts einiges beitragen, kann man sich natürlich fragen, was diese Fiktionen in der Philosophie verloren haben:

As a newcomer to philosophy, one is soon treated, both in class and textbooks, to a steady diet of strange, exotic, or downright bizarre examples. The average undergraduate student's reaction to these examples is, I think it fair to say, one of bewilderment. Why should Olga's relatives take such pleasure in making soap of her? That's unrealistic! How could somebody acquire my memories, wishes, thoughts by stepping into a machine? Surely that's not possible! How could it be that a group of people gather to decide on the principles of society without even knowing their own age, sex, physical or

---

<sup>1</sup> Wenn ich hier sage, dass wir Szenarien „für Gedankenexperimente verwenden“, will ich damit darauf hinweisen, dass zumindest ein unterschiedlicher Sprechakt vorliegt, wenn ich ein Gedankenexperiment vortrage und wenn ich dieselbe Episode als eine *Science-Fiction* Geschichte erzähle.

psychological characteristics? What is the point of these exercises? (Häggqvist [137], 11)

Die meisten dieser „Newcomer“ akzeptieren die Verwendungsweise irgendwann, ohne dass davon auszugehen ist, dass sie eine befriedigende Antwort auf ihre skeptischen Fragen erhalten haben. Bei anderen führt dieser Skeptizismus dazu, sich mit der methodologischen Rolle von Gedankenexperimenten genauer auseinanderzusetzen. Letzteres ist vermehrt in den letzten Jahren geschehen. Nicht selten kommen die Untersuchungen dabei zu dem Ergebnis, dass die Skepsis des „Newcomers“ mehr als berechtigt ist. So kommt beispielsweise Kathleen V. Wilkes zu folgendem Ergebnis: „In sum, we cannot extract philosophically interesting conclusions from fantastical thought experiments.“ (Wilkes [345], 46)

Die vorliegende Untersuchung strebt eine Neubewertung dieses skeptischen Standpunktes an. Obschon es stimmt, dass bestimmte Verwendungsweisen von Gedankenexperimenten in der Philosophie auf methodologisch zweifelhaften Fundamenten stehen, sind philosophische Gedankenexperimente insgesamt besser, als die Skeptiker uns glauben machen wollen.

Im Folgenden wird in 1.1 das Thema dieses Buches näher charakterisiert, in 1.2 die Gliederung erläutert und dabei (als Orientierungshilfe für den Leser) die Hauptergebnisse zusammengefasst.

## 1.1 GEDANKENEXPERIMENTE IN DER PHILOSOPHIE

Gedankenexperimente spielen sowohl in den empirischen Wissenschaften wie auch in der Philosophie eine wichtige Rolle. Betrachten wir zunächst die empirischen Wissenschaften, wobei wir uns auf die Naturwissenschaften beschränken. In der Geschichte der Naturwissenschaften ist nach verbreiteter Auffassung insbesondere zu Zeiten so genannter wissenschaftlicher „Krisen“<sup>2</sup>, also zu denjenigen Phasen der Wissenschaftsgeschichte, die einem Paradigmenwechsel vorangingen, in starkem Maße „gedankenexperimentiert“ worden. Das gelte sowohl für die Ablösung der Aristotelischen Physik durch Galileo Galilei, als auch für die Phasen, in denen sich die Newtonsche Physik, die spezielle und allgemeine Relativitätstheorie, sowie die Quantenmechanik etablieren konnten. Gedankenexperimente sollen in all diesen Fällen eine erkenntniserweiternde Rolle gespielt haben, die ent-

---

<sup>2</sup> Ich habe mich bemüht, eine weitgehend eindeutige Verwendungsweise von Anführungszeichen zu benutzen. Einfache Anführungszeichen markieren in erster Linie, dass ich über einen *Ausdruck* rede, ich verwende sie häufig auch, wenn ich über den Begriff rede, der von dem Ausdruck in Anführungszeichen ausgedrückt wird. Wenn es nötig ist, unterscheide ich Begriffe von Ausdrücken dadurch, dass ich die Begriffe durch den Ausdruck in Grossbuchstaben bezeichne. Doppelte Anführungszeichen signalisieren, dass ich zitiere oder mich einer metaphorischen Ausdrucksweise bediene (wird durch den Kontext jeweils klar). In Fällen, in denen ich in einen angeführten Ausdruck „hineinquantifiziere“, benutze ich manchmal auch französische Anführungszeichen, „*‘* und „*‘*“.

weder durch tatsächliches Experimentieren nicht erfüllt werden konnte, oder für deren Erfüllung tatsächliche Experimente nicht notwendig waren.

Ein berühmtes Beispiel für den erfolgreichen Einsatz von Gedankenexperimenten in der Physik ist Galileis so genanntes Pisa-Experiment.<sup>3</sup> Die Aristotelische Physik beinhaltet die Gesetzesaussage, dass die Fallgeschwindigkeit von Körpern direkt proportional zu ihrem Gewicht sei. Eine 8kg schwere Eisenkugel würde dementsprechend achtmal schneller fallen als eine bloß 1kg schwere Kugel. Diese Auffassung der Aristotelischen Physik führt Galilei mit einem Gedankenexperiment in seinem *Discorsi ad absurdum* und argumentiert dafür, dass große und kleine Körper desselben Materials mit der gleichen Geschwindigkeit fallen.

Man stelle sich vor, dass ein schwerer Körper (A) und ein leichter Körper (B) zusammengebunden und aus einiger Höhe fallen gelassen werden. Nach der Aristotelischen Theorie haben beide ihre natürliche Fallgeschwindigkeit, die linear proportional zu ihrem Gewicht ist. Der kleinere Körper hat entsprechend eine geringere natürliche Fallgeschwindigkeit (F) als der größere Körper,  $F(B) < F(A)$ . Entsprechend sollte der kleinere Körper den größeren Körper etwas abbremsen, der größere Körper den kleineren etwas beschleunigen, so dass die Geschwindigkeit des gesamten Systems (A+B) zwischen der natürlichen Geschwindigkeit des kleinen und des großen Körpers liegen sollte, auf jeden Fall also kleiner als die der Geschwindigkeit des großen Körpers alleine sein sollte,  $F(A+B) < F(A)$ .

Andererseits ist das Gewicht des Gesamtsystems größer als das des großen Körpers alleine, weshalb die Fallgeschwindigkeit des Gesamtsystems nach der Aristotelischen Theorie doch auf jeden Fall größer als die des großen Körpers alleine sein sollte,  $F(A+B) > F(A)$ .

Somit prognostiziert die Aristotelische Theorie sowohl, dass das Gesamtsystem schneller fällt als der große Körper alleine, wie auch dass das Gesamtsystem langsamer fällt – was offenbar ein Widerspruch ist. Galilei schließt daraus, dass die natürliche Geschwindigkeit fallender Körper unabhängig von ihrem Gewicht ist, verschieden große Körper desselben Materials also mit gleicher Geschwindigkeit fallen.

Dies ist ein Paradebeispiel für erfolgreiches Gedankenexperimentieren im *destruktiven* wie *konstruktiven* Sinn (vgl. Brown [38]). Eine empirische Gesetzesaussage wird widerlegt, eine konkurrierende Gesetzesaussage argumentativ gestützt, ohne dass der Lehnstuhl verlassen werden musste. Häufig sind physikalische Gedankenexperimente auch nur destruktiv, also nur als Kritik an einer empirischen Theorie intendiert (wie in den Fällen von Einstein/Podolsky/Rosen-Gedankenexperimenten gegen die Vollständigkeit der Quantenmechanik und von Maxwells Dämon-Gedankenexperiment gegen die strikte Notwendigkeit des zweiten Hauptsatzes der Thermodynamik), manchmal auch nur konstruktiv zur argumentativen Stützung einer bestimmten Theorie (wie im Fall von Stevins Schiefe-Ebene-Experiment („Kugelkranzbeweis“) zur Etablierung des Gesetzes, dass die

---

<sup>3</sup> Ich referiere hier die Standardinterpretation des Galileischen Gedankenexperiments, wie sie sich etwa bei Popper [260] oder Brown [38] findet. Zu den wissenschaftshistorischen Problemen mit dieser Interpretation, vgl. Kapitel 3.3.2.1.

Kraft, die benötigt wird, um einen Körper auf einer schiefen Ebene zu halten, umgekehrt proportional zur Länge der Ebene ist (bei gegebener Höhe)).

### 1.1.1 PHILOSOPHISCHE GEDANKENEXPERIMENTE

In der Philosophie wird in weit größerem Maße mit Gedankenexperimenten „geforscht“ als in den empirischen Wissenschaften. Mindestens seit Platon (z.B. „Ring des Gyges“-Gedankenexperiment aus *Politeia*) werden Gedankenexperimente erdacht, um philosophische Theorien zu Fall zu bringen oder zu stützen.<sup>4</sup> Dabei ist es in den letzten Jahrzehnten sogar zu einer Zunahme des Gedankenexperimentierens vor allem innerhalb der analytischen Philosophie gekommen.

Ein Paradebeispiel für ein philosophisches Gedankenexperiment ist im so genannten *Zombieargument* enthalten, das aus der Philosophie des Geistes stammt. Hierbei wird gegen den *a posteriori* Physikalismus, also die metaphysische Auffassung, dass jede mentale Eigenschaft bzw. jeder mentale Zustand mit einer physischen Eigenschaft bzw. einem physischen Zustand identisch ist, etwa folgendermaßen<sup>5</sup> argumentiert:

#### *Zombie-Argument*

- (1) Wenn der *a posteriori* Physikalismus wahr ist, dann ist er notwendig<sup>6</sup> wahr.
- (2) Wenn der Physikalismus notwendig wahr ist, dann gibt es keine mögliche Welt, in der eine physische Eigenschaft  $\Phi$ , die in der aktuellen Welt mit einer mentalen Eigenschaft  $\Psi$  korreliert ist, vorkommt, ohne dass in dieser Welt auch  $\Psi$  vorkommt.
- (3) Es ist vorstellbar, dass es eine mögliche Welt gibt (eine so genannte „Zombie-Welt“), die in allen physikalischen Eigenschaften mit der aktuellen Welt identisch ist, in der aber keine mentalen Eigenschaften vorkommen.
- (4) Wenn Zombie-Welten vorstellbar sind, dann sind sie auch möglich.
- (5) Wenn Zombie-Welten möglich sind, dann gibt es eine mögliche Welt, in der eine physische Eigenschaft  $\Phi$ , die in der aktuellen Welt mit einer men-

<sup>4</sup> Zu den historischen Wurzeln des philosophischen Gedankenexperiments in der Vorsokratik vgl. Rescher [278].

<sup>5</sup> Aus Platzgründen und zur besseren Verständlichkeit ist das folgende Argument eine starke Vereinfachung dessen, was in der modernen Philosophie des Geistes verhandelt wird. Dieser Physikalismus heißt „*a posteriori*“, da angenommen wird, dass wir die Identität von mentalen und physischen Zuständen nur *a posteriori* herausfinden können.

<sup>6</sup> Es wird dabei davon ausgegangen, dass der Physikalismus eine Identität zwischen Eigenschaften oder Zuständen behauptet, und Identität notwendig ist (es nicht möglich ist, dass etwas nicht mit sich selbst identisch ist).

talene Eigenschaft  $\Psi$  korreliert ist, vorkommt, ohne dass in dieser Welt auch  $\Psi$  vorkommt.

- (6) Folglich ist der *a posteriori* Physikalismus falsch.

Neben dem Zombie-Argument finden sich in der Philosophie noch viele weitere prominente Gedankenexperimente, so das „Knowledge Argument“ von Frank Jackson oder Searles Gedankenexperiment vom Chinesischen Zimmer, Putnams „Gehirne im Tank“-Gedankenexperiment oder sein Gedankenexperiment mit der Zwillingserde, Nagels Fledermaus, Goodmans und Kripkes zerrüttete Prädikate, etc.

Im Zombie-Argument (wie auch verwandten Argumenten in der Philosophie) wird von der Vorstellbarkeit einer bestimmten Situation (dem „Gedankenexperiment“ in (3)) auf die Möglichkeit einer solchen Situation geschlossen (in (4)). Dass die Möglichkeit einer Situation oder Welt dabei überhaupt von Interesse ist, liegt daran, dass die jeweils angegriffene philosophische These als eine Notwendigkeitsbehauptung aufgefasst wird. So wird beispielsweise der *a posteriori*-Physikalismus seit Kripkes *Naming and Necessity* (Kripke [172]) gemeinhin für eine These gehalten, die, wenn sie wahr ist, „metaphysisch notwendig“ wahr ist. Wenn es dann gelingt zu zeigen, dass es metaphysisch möglich ist, dass es eine Zombie-Welt gibt, ist der Physikalismus widerlegt.

### 1.1.2 DAS PROBLEM

Viele<sup>7</sup> meinen, dass die Philosophie bei ihrem Gedankenexperimentieren aber weit weniger erfolgreich sei, obwohl hierbei dieselbe Methode wie in der Naturwissenschaft zur Anwendung komme. Während die Naturwissenschaftler durch Gedankenexperimente zu akzeptierten Resultaten kämen, seien die philosophischen Gedankenexperimente nur selten durchschlagend.<sup>8</sup> Das Zombie-Argument ist beispielsweise höchst kontrovers, während das Gedankenexperiment Galileis nur selten Zweifel hervorruft. Der Unterschied im Erfolg bei der Anwendung derselben Methode wird dann zumeist dadurch erklärt, dass Philosophen entweder beim Gedankenexperimentieren nicht genügend Umsicht walten lassen, oder dass formale und inhaltliche Besonderheiten philosophischer „Theorien“ das Gedankenexperimentieren scheitern ließen.

---

<sup>7</sup> Wie wir sehen werden, sieht das nahezu jeder so, der sich zu diesem Thema geäußert hat. Hierunter fallen auch praktisch all diejenigen Autoren, die zunächst versichern, dass sie „im Gegensatz zu vielen anderen“ auch einige naturwissenschaftliche Gedankenexperimente kritisch sehen (wie etwa Peijnenburg und Atkinson [252] oder Bunzl [46]).

<sup>8</sup> Vgl. Sorensen [304], 12; Peijnenburg und Atkinson [252].

### 1.1.3 AUFGABENSTELLUNG

Um die Kritik an den philosophischen Gedankenexperimenten besser beurteilen zu können, soll in dieser Arbeit die Funktion von philosophischen Gedankenexperimenten genauer untersucht werden. Die auf diese Weise gewonnenen Ergebnisse sollen zur Entwicklung einer Methodologie des Gedankenexperiments herangezogen werden. D.h., es soll untersucht werden, ob erfolgreiches philosophisches Gedankenexperimentieren möglich ist, und falls ja, ob sich hierzu methodologische Vorschläge formulieren lassen.

Da in der Philosophie Erkenntnis- und Geltungsansprüche kontrovers sind, können sinnvolle methodologische Vorschläge nur *relative* Aussagen machen, wenn man den Gegenstandsbereich der Philosophie nicht einschränken möchte und methodologische Überlegungen die inhaltlichen Problemlösungen nicht bereits voraussetzen sollen. Methodologische Vorschläge in der Philosophie sind daher zunächst relativ zu den *Zielen*, die mit Gedankenexperimenten verfolgt werden. Gegeben bestimmte Ziele ist die Rechtfertigung für die Möglichkeit sinnvollen philosophischen Gedankenexperimentierens außerdem relativ zu den philosophischen *Hintergrundüberzeugungen*, die die an einer (mit Gedankenexperimenten geführten) Debatte beteiligten Philosophen außerdem haben. Es wird also darauf ankommen, diese Ziele zu unterscheiden und eventuelle methodologische Vorschläge unter Rücksichtnahme auf verschiedene mögliche Hintergrundüberzeugungen zu formulieren. Zu diesen Hintergrundüberzeugungen gehört der *Geltungsanspruch*, der mit philosophischen Thesen verbunden wird, wie auch der *Erkenntnisanspruch*, den man erhebt.

Der „Geltungsanspruch“ bezieht sich (hier) auf den intendierten Geltungsbereich, der mit einer Aussage verbunden ist. Dieser kann modal definiert sein und als Geltungsbereich Mengen von möglichen Welten festlegen, aber auch andere Bereichsbeschränkungen entlang anderer Dimensionen erhalten. So kann eine „philosophische“ These – wie wir im Detail noch genauer sehen werden – manchmal als eine Behauptung verstanden werden, die nur für die aktuelle Welt einen Wahrheitsanspruch erhebt, aber manchmal auch als begriffliche oder metaphysische Notwendigkeit, die entsprechend auch den Anspruch erhebt, in anderen möglichen Welten als der aktuellen Welt wahr zu sein. Je nachdem in welchen möglichen Welten eine Behauptung wahr sein soll, kann man vom „modalen Geltungsbereich“ einer Aussage reden. Der Geltungsbereich einer Aussage kann aber auch entlang anderer Dimensionen *innerhalb* von möglichen Welten beschränkt sein. So ist eine Bedeutungsanalyse – wie wir ebenfalls noch sehen werden – zwar in aller Regel eine Behauptung, die in allen begrifflich möglichen Welten wahr sein soll, sie kann aber innerhalb dieser Welten durchaus auf eine bestimmte Sprechergemeinschaft beschränkt sein (beispielsweise auf bestimmte Experten oder besonders „kompetente“ Sprecher).

Der „Erkenntnisanspruch“ bezieht sich (hier) darauf, was man behauptet, von einer bestimmten Informationsgrundlage ausgehend wissen bzw. gerechtfertigt glauben zu können. So wird uns in dieser Untersuchung häufig der Anspruch begegnen, bestimmte Aussagen a priori oder „intuitiv“ auf der Grundlage einer be-

stimmt Informationsbasis beurteilen zu können. Was man a priori wissen kann (und ob man überhaupt irgendetwas a priori wissen kann) ist aber umstritten. Ähnlich umstritten ist es, wann man eine Erkenntnismethode für gerechtfertigt halten darf. Würde sich diese Untersuchung von vornherein auf bestimmte Geltungsbereiche philosophischer Thesen und eine bestimmte Erkenntnistheorie festlegen, wäre sie nur für denjenigen Teil der Philosophie von Interesse, der diese Hintergrundannahmen und Ansprüche teilt. Diese Untersuchung sollte daher in dieser Hinsicht möglichst voraussetzungsarm sein und entweder bezüglich solcher Kontroversen neutral bleiben, oder den in Bezug auf eine solche Kontroverse eingenommenen Standpunkt argumentativ entwickeln.

Es ist also die Aufgabe dieser Untersuchung, eine Charakterisierung und Beurteilung der methodologischen Rolle(n) von Gedankenexperimenten in der Philosophie zu leisten, die möglichst<sup>9</sup> unabhängig von spezifischen Geltungs- und Erkenntnisansprüchen philosophischer Einzelpositionen nachvollziehbar ist.

Trotz dieser Relativierungen kann es aber dennoch sein, dass wir zu einem allgemeinen Ergebnis kommen. Wenn es sich – beispielsweise – erweist, dass Gedankenexperimente im Lichte jeder nachvollziehbaren Kombination von Erkenntnis- und Geltungsansprüchen eine sinnvolle kritische Funktion innerhalb der philosophischen Methodologie erfüllen können, dann sollte man Gedankenexperimente *allgemein* als Instrumente der Kritik ernst nehmen (weil man sinnvolle Instrumente der Kritik nicht ignorieren darf).

---

<sup>9</sup> Natürlich können hierbei nicht alle tatsächlichen, geschweige denn alle möglichen Auffassungen berücksichtigt werden. Auffassungen, die außerhalb der so genannten „analytischen Philosophie“ liegen, wie auch „global“ skeptische Auffassungen (sowie besonders exotische, abwegige oder inkonsistente Auffassungen) werden nicht berücksichtigt. Vollständige Neutralität wäre außerdem mindestens so uninteressant wie die ausschließliche Berücksichtigung einer einzigen Menge von Auffassungen.

## 1.2 GLIEDERUNG UND HAUPTERGEBNISSE

Diese Untersuchung besteht aus insgesamt neun Teilen, deren erster Teil diese Einleitung darstellt. Im ersten großen thematischen Abschnitt der Arbeit, bestehend aus Teil 2 bis Teil 4, wird eine allgemeine Theorie des Gedankenexperiments rekonstruiert, die auf den methodologischen Einsichten der Wissenschaftstheorie der Naturwissenschaften basiert.

Eine Untersuchung zu Gedankenexperimenten in der Philosophie mit einer Untersuchung der Rolle von Gedankenexperimenten in den Naturwissenschaften zu beginnen, ist durch folgende Überlegungen motiviert:

(a) Das Wort ‚Gedankenexperiment‘ wird heutzutage für eine Unmenge verschiedenster Dinge verwendet. Manche bezeichnen damit kontrafaktische Überlegungen im Allgemeinen<sup>10</sup>, Computersimulationen<sup>11</sup>, Textaufgaben im Physik- oder Mathematikunterricht<sup>12</sup>, Erinnerungen an in der Vergangenheit beobachtete Ereignisse<sup>13</sup>, Computerkunst<sup>14</sup>, etc. Dies gilt schon für die Verwendungsweise dieses Ausdrucks innerhalb der Philosophie, aber in weitaus größerem Maße für seine umgangssprachliche Verwendung. Versuchte man alle diese Verwendungsweisen mit einer gemeinsamen Definition einzufangen, stünde man mit einem Begriff da, der so ziemlich alles zu bezeichnen vermag. Der allgemeine Sprachgebrauch erweist sich damit als ungeeigneter Ausgangspunkt für unsere Untersuchung.

Nun ist es so, dass das Wort ‚Gedankenexperiment‘ kein Wort ist, das in der Umgangssprache vorhanden war und dann in der Wissenschaftstheorie adaptiert wurde, um dort eine begrenztere Rolle mit einer schärferen und präziseren Bedeutung zu spielen, sondern ein Wort, das aus der Wissenschaftstheorie stammt und dessen Anwendung später auf andere Bereiche ausgeweitet wurde.

Die paradigmatischen Fälle aus der Geschichte der Naturwissenschaften, die in der Wissenschaftstheorie der Naturwissenschaften zur Einführung dieses Begriffes geführt haben, sollten daher charakteristische Gemeinsamkeiten aufweisen, die eine Explikation von ‚Gedankenexperiment‘ erlauben.

(b) Die meisten einschlägigen Untersuchungen zu Gedankenexperimenten in der Philosophie gehen davon aus, dass der Begriff ‚Gedankenexperiment‘ innerhalb der Methodologie der Naturwissenschaften mit dem Begriff ‚Gedankenexperiment‘ innerhalb der Methodologie der Philosophie in engem Zusammenhang steht. Die Kriterien für gute oder schlechte Gedankenexperimente, die sich an den naturwissenschaftlichen Fällen formulieren lassen, sind nach dieser Auffassung auf die Fälle in der Philosophie übertragbar. Falls dem so ist, ist ein Blick in die Geschichte der Naturwissenschaften und ihrer Wissenschaftstheorie offensichtlich

---

<sup>10</sup> Rescher [278].

<sup>11</sup> Stäudner [314], Humphreys [156].

<sup>12</sup> Brendel [35].

<sup>13</sup> Mach [193].

<sup>14</sup> Dietrich [88].

lohnend, will man erfahren, wie sich die Methodologie des Gedankenexperimentierens in der Philosophie am Besten charakterisieren lässt. *Ob* dem so ist, wird man aber auch nur dann herausfinden können, wenn man in die Geschichte der Naturwissenschaften nachgesehen hat, was dort – unter der wissenschaftstheoretischen Bezeichnung ‚Gedankenexperiment‘ – eigentlich getan wird und es damit vergleicht, was Philosophen tun, wenn man ihre Vorgehensweise als ‚Gedankenexperimentieren‘ bezeichnet.

### 1.2.1 DAS GEDANKENEXPERIMENT IN DER WISSENSCHAFTSTHEORIE

In Teil 2, *Das Gedankenexperiment in der Wissenschaftstheorie der Naturwissenschaften*, wird zunächst untersucht, wie der Begriff ‚Gedankenexperiment‘ in die Wissenschaftstheorie eingeführt wurde. Auf diese Weise soll festgestellt werden, was aus Sicht der Wissenschaftstheorie zu den paradigmatischen Episoden der Wissenschaftsgeschichte gehört, die als „Gedankenexperimente“ behandelt werden, außerdem soll untersucht werden, warum diese Episoden für die Wissenschaftstheorie von Interesse sind. Für einen Vergleich von naturwissenschaftlichen und philosophischen Gedankenexperimenten müssen wir einen Begriff vom naturwissenschaftlichen Gedankenexperiment präzisieren, der zumindest auf jene Episoden der Wissenschaftsgeschichte zutrifft, die nachweislich von besonderem wissenschaftstheoretischem Interesse sind und deshalb mit ‚Gedankenexperiment‘ bezeichnet wurden. Ansonsten wäre ein solcher Vergleich uninteressant.

„Gedankenexperimente“ der Naturwissenschaften mit denen der Philosophie zu vergleichen ist müßig, wenn man bei der Präzisierung, was man mit ‚Gedankenexperiment‘ bezeichnen möchte, keine solchen Restriktionen beachtet hat. Gegeben die kaum restringierte umgangssprachliche Verwendungsweise des Ausdrucks kann man ‚Gedankenexperiment‘ beliebig präzisieren. Eine solche Vorgehensweise könnte einem aber den Vorwurf einbringen, dass die Begriffsexplikation von ‚Gedankenexperiment‘ schon die Antworten auf alle weiteren Fragen *per definitionem* festgelegt hat. Am Anfang steht also eine Resümierung der Karriere des Themas ‚Gedankenexperiment‘ in der Wissenschaftstheorie. Dabei wird Folgendes in den einzelnen Unterkapitel festgestellt:

2.1: Die „eigentliche“ Begriffsgeschichte des Gedankenexperiments beginnt mit den Versuchen Ernst Machs, Argumentationen, die er in der Geschichte der Naturwissenschaften bei z.B. Galilei oder Stevin vorfindet, in eine empiristisch konzipierte Methodologie der Naturwissenschaften einzuordnen. Was Mach dabei in *Die Mechanik in ihrer Entwicklung* oder in *Erkenntnis und Irrtum* als Gedankenexperiment ausweist, gehört bis heute der Klasse paradigmatischer Gedankenexperimente an (welche heute freilich mindestens um die Gedankenexperimente im Zusammenhang mit den Relativitätstheorien und der Quantenmechanik erweitert ist). Machs Erörterungen bleiben dabei eher im Bereich der

„Phänomenologie der Psychologie“<sup>15</sup>, auch wenn er gewisse Mutmaßungen darüber anstellt, warum diese Methode funktioniert, wenn sie es tut. In diesen Mutmaßungen formuliert Mach Ansätze zu einer evolutionären *Erkenntnistheorie* des Gedankenexperiments, da nach seiner Auffassung Gedankenexperimente zu gerechtfertigten Überzeugungsänderungen führen können.

2.2: Diese letzte Einschätzung Machs wird von seinen Zeitgenossen (Meinong, Russell, Duhem) und unmittelbaren historischen Nachfolgern (z.B. Hempel) nicht geteilt. Da sich Gedankenexperimente nicht ohne weiteres in die vorausgesetzte Erkenntnistheorie einpassen lassen, werden sie als Heuristiken behandelt und damit zunächst aus dem Gegenstandsbereich der Wissenschaftstheorie ausgesondert. Eine Ausnahme hierzu bildet Karl Popper, der die Rolle von Gedankenexperimenten in Argumentationen besonders betont und dafür plädiert, die Dialektik<sup>16</sup> der Argumentationssituation bei der Bewertung von Gedankenexperimenten hinzuzuziehen.

2.3: Durch die zunehmende Relevanz, die wissenschaftshistorischen Untersuchungen für die Wissenschaftstheorie im letzten Drittel des letzten Jahrhunderts eingeräumt wird, erleben auch die Gedankenexperimente (vor allem durch den Wissenschaftshistoriker Thomas S. Kuhn) eine Neubewertung. Gedankenexperimente bekommen in Phasen wissenschaftlicher Revolutionen, in denen ein Paradigma durch ein neues abgelöst wird, eine wichtige Vermittlungsfunktion zugestanden. Dieser *vermittelnde* Charakter besteht dabei *nicht* darin, der Gegenseite (also den Vertretern des „alten“ Paradigmas) einen internen Widerspruch nachzuweisen (wie Popper dies vermutet hatte).

### 1.2.2 EINE ALLGEMEINE THEORIE DES GEDANKENEXPERIMENTS

Auf der Grundlage der Ergebnisse des zweiten Teils wird im dritten Teil, *Eine allgemeine Theorie des Gedankenexperiments*, eine Präzisierung des Begriffs ‚Gedankenexperiment‘ vorgenommen. Dabei werden in Kapitel 3.1 „klärende Gedankenexperiment“, „funktionale Gedankenexperimente“ sowie „Gedankenexperimente zur Überzeugungsrevision“ (auch ‚kritische Gedankenexperimente‘ genannt) unterschieden. Letztere sind diejenigen Gedankenexperimente, die das Hauptinteresse der Wissenschaftstheorie auf sich gezogen haben. Sie lassen sich folgendermaßen charakterisieren:<sup>17</sup>

---

<sup>15</sup> Kühne [175], 174.

<sup>16</sup> Zur „Dialektik“ einer Argumentationssituation (oder eines „argumentativen Kontexts“) gehören für mich die relativ schlichten Fragen danach, wer unter welchen Voraussetzungen mit wem über (bzw. gegen) was argumentiert. Aus der Dialektik der Argumentationssituation (oder des „Diskurses“) können Aussagen darüber gemacht werden, wer für eine Annahme die Beweislast trägt, wann jemand in seiner Argumentation eine *petitio principii* begeht, etc.

<sup>17</sup> In leichter Abwandlung von Gähde [113].

- (i) Das Ziel eines Gedankenexperiments besteht darin, begründete Überzeugungsänderungen beim Adressaten zu bewirken.
- (ii) Um dieses Ziel zu erreichen, werden entweder Experimente beschrieben, oder es werden bestimmte Sachverhalte geschildert, die denk- oder vorstellbar sein sollen.
- (iii) Das Ziel der Überzeugungsänderung soll erreicht werden, ohne dass die geschilderten Experimente *de facto* ausgeführt oder die entsprechenden Sachverhalte als real angenommen werden müssten.

Für den Rest unserer Untersuchung wollen wir uns auf diese Funktion von Gedankenexperimenten besonders konzentrieren.

In Kapitel 3.2 wird ein Argumentschema entwickelt, das dazu geeignet ist, Gedankenexperimente zur Überzeugungsrevision in ihrem argumentativen Kontext zu rekonstruieren. Es wird hierzu gezeigt, dass diese Analyse wichtige Eigenschaften naturwissenschaftlicher Gedankenexperimente *erklären* kann. Gedankenexperimente zur Überzeugungsrevision kommen demnach typischerweise in *modalen* Argumenten vor, in denen eine bestimmte Überzeugung, die als (in irgendeinem Sinne) notwendig wahr angesehen wird, durch ein Gegenbeispiel falsifiziert werden soll, das eine mit der fraglichen Notwendigkeit inkompatible Möglichkeit präsentiert. Die Frage, ob ein solches Argument gültig ist, ist (unter anderem) eine Frage danach, ob die in der gedankenexperimentellen Situation repräsentierte Möglichkeit mit der Notwendigkeit der Targetthese tatsächlich inkompatibel *ist* (was, falls es verschiedene Grade oder Arten von Möglichkeit und Notwendigkeit gibt, jeweils zu prüfen wäre).

So behaupten naturwissenschaftliche Hypothesen eine *Naturnotwendigkeit* und ein relevantes Gegenbeispiel müsste demnach (mindestens) eine *Naturmöglichkeit* präsentieren. Ein solches (gültiges) gedankenexperimentelles Argument *rechtfertigt* nun eine Überzeugungsänderung in Bezug auf die kritisierte These zu dem Maße, zu dem man gerechtfertigt ist, anzunehmen, dass die Prämissen des gedankenexperimentellen Arguments *wahr* sind. Insbesondere in dem Maße, zu dem man gerechtfertigt ist, anzunehmen, dass die präsentierte gedankenexperimentelle Situation tatsächlich *naturmöglich* ist.

Die Frage, wie Gedankenexperimente zu gerechtfertigten Überzeugungsänderungen führen können, kann für manche naturwissenschaftliche Gedankenexperimente vermutlich nach dem Muster der (in der Tradition Machs stehenden) Spekulationen zu einer evolutionären Erkenntnistheorie erklärt werden. Demnach besitzen wir angeborenes (intuitives) Wissen von den relevanten Naturmöglichkeiten.<sup>18</sup> Dies scheint aber nicht für alle naturwissenschaftlichen Gedankenexperimente der Fall zu sein.

---

<sup>18</sup> ... beispielsweise von solchen Möglichkeiten, die in alltäglichen ((ehemals) überlebenswichtigen) Entscheidungsprozessen als relevante Möglichkeiten in Betracht gezogen werden müssen. Vgl. die Beispiele in Cooper [78] und Maffie [198].

Zur genaueren Klärung der Frage, inwiefern die intuitive Beurteilung physikalischer Zusammenhänge beim Gedankenexperiment eine entscheidende Rolle spielt, wird in Kapitel 3.3 untersucht, ob Gedankenexperimente eliminierbar sind, d.h., ob jegliche Konklusion, die durch ein Gedankenexperiment erreicht wurde, genauso (überzeugend) durch ein Argument hätte etabliert werden können, in dessen Prämissen kein Bezug auf kontrafaktische oder hypothetische Gegebenheiten und Einzeldinge vorkommt. Sollten (zumindest manche) Gedankenexperimente nicht eliminierbar sein, wird die Frage, wie Gedankenexperimente zu gerechtfertigten Überzeugungsänderungen führen konnten, brisant.

Gedankenexperimente wären in einer Argumentation *verzichtbar*, wenn die Bezugnahme auf einen kontrafaktischen Fall (der für Gedankenexperimente ja charakteristisch ist) sich durch eine Menge von universalen Prämissen ersetzen ließe.

Dass eine rationale Rekonstruktion von Gedankenexperimenten, die Gedankenexperimente durch nicht-gedankenexperimentelle Argumente ersetzt, zumindest manchmal nicht möglich ist, liegt nach der hier vorgestellten Analyse daran, dass versteckte Hintergrundannahmen in die Darstellungen eingehen, die zum Zeitpunkt der Durchführung des Gedankenexperiments nicht gerechtfertigt werden konnten, von denen die jeweiligen Autoren aber annehmen durften, dass sie spezialisiert auf den kontrafaktischen Fall nicht problematisch sind. Dass in diesen Fällen das nun brisante Rechtfertigungsproblem gelöst werden kann, liegt daran, dass die physikalischen „Intuitionen“, die hier eine Rolle spielen, *tatsächlich* unproblematisch sind.

### 1.2.3 GEDANKENEXPERIMENTE IN DER PHILOSOPHIE

Im vierten Teil wenden wir uns den Gedankenexperimenten in der Philosophie zu. In Kapitel 4.1 werden zunächst paradigmatische Gedankenexperimente aus der theoretischen Philosophie vorgeführt, die für die restlichen Erläuterungen als Beispielmateriale dienen. Hierzu gehört aus der Philosophie des Geistes Searles Chinesisches Zimmer, Jacksons Mary, sowie das Zombie-Argument, aus der Erkenntnistheorie die Familie der Gettier-Fälle, aus der Sprachphilosophie das Argument von der Zwillingserde, aus der Metaphysik das Argument von den Gehirnen im Tank und aus der Debatte um personale Identität das Reduplikationsargument. In Kapitel 4.2 werden dann Gedankenexperimente aus der Praktischen Philosophie diskutiert, insbesondere der „weltberühmte Geiger“ von Thomson. Von Jonathan Dancy ist argumentiert worden, dass ethische Gedankenexperimente einer ganz bestimmten Logik folgen, die nur dann sinnvoll rekonstruiert werden kann, wenn man einen Partikularismus in Bezug auf die ethische Beurteilung von Sachverhalten voraussetzt. Wir werden sehen, dass die rationale Rekonstruktion von Gedankenexperimenten in der Praktischen Philosophie unabhängig von einer solchen Annahme ist. Der Großteil der bis dahin betrachteten Gedankenexperimente besitzt nach unserer Analyse die charakteristischen Eigenschaften kritischer Gedankenexperimente, wie sie in Teil 3 herausgearbeitet wur-

den. Wie wir in Kapitel 4.3 sehen, befinden sich daneben auch in der Philosophie klärende Gedankenexperimente, die nur illustrative Funktionen haben. Eine dritte Gruppe philosophischer „Gedankenexperimente“, die kein Gegenstück in der naturwissenschaftlichen Methodologie zu haben scheinen, sind Modelle zur Sicherstellung umfassender Interessenberücksichtigung in der Ethik. Obwohl sich für diese Verfahrensweisen gewisse methodologische Probleme stellen, sind diese jedoch weder gravierend, noch in deutlicher Weise mit den methodologischen Problemen kritischer Gedankenexperimente verwandt. Im Folgenden soll daher die Funktionsweise kritischer Gedankenexperimente in der Philosophie im Zentrum stehen.

#### 1.2.4 KRITIK AN PHILOSOPHISCHEN GEDANKENEXPERIMENTEN IM ALLGEMEINEN

Mit Teil 5, *Kritik an philosophischen Gedankenexperimenten im Allgemeinen*, beginnt der zweite große thematische Abschnitt dieser Arbeit: eine Diskussion der bisher vorgebrachten Kritik an philosophischen Gedankenexperimenten in systematischer Ordnung. Der allgemeinste Einwand, der dabei gegen das Gedankenexperimentieren in der Philosophie vorgebracht wurde, besteht darin, darauf hinzuweisen, dass wir keine positive Theorie zu ihrer Rechtfertigung besitzen. Dieser Einwand ist jedoch als wenig überzeugend zurückzuweisen (5.1). Ein Defizit in der Erkenntnis- oder Wissenschaftstheorie kann alleine noch keine Methode in Misskredit bringen. Der Einwand wird erst interessant, wenn man dazu behaupten möchte, dass eine solche Rechtfertigung prinzipiell ausgeschlossen ist, beispielsweise weil sich zeigen lässt, dass (A) die Methode unzuverlässig ist, oder (B) die Methode aus prinzipiellen Gründen nicht evaluierbar ist. Für (B) hat man z.B. ins Feld geführt, dass in Gedankenexperimenten die prinzipiell unüberprüfbare Annahme vorausgesetzt werde, dass wir zuverlässig prognostizieren können, wie wir selbst in kontrafaktischen Fällen sprachlich handeln würden (z.B. von Fodor [106]).

In Wittgensteinscher bzw. Quinescher Nachfolge wird der Einwand in der Stoßrichtung von (A) vorgebracht, dass Gedankenexperimente deshalb unzuverlässige Mittel zur Erforschung (beispielsweise) der Bedeutung unserer Begriffe darstellen, weil unsere Sprache nur für die aktuell auftretenden Situationen gemacht ist und außerhalb dieser nicht mehr oder nicht mehr zuverlässig prognostizierbar funktioniert. Dieser Einwand wird in 5.2 diskutiert.

In Kapitel 5.3 gehen wir dem ebenfalls sehr allgemeinen Einwand nach, dass philosophische Gedankenexperimente häufig bestimmte typische Defekte aufweisen: sie seien zirkulär und/oder widersprüchlich. Dieser Einwand wird nicht gegen philosophische Gedankenexperimente im Allgemeinen erhoben, sondern nur gegen bestimmte Gedankenexperimente, allerdings mit der Hintergrundannahme, dass die Philosophie keine Ressourcen besitzt, mit dieser Problematik umzugehen, wohingegen die Naturwissenschaften, bei denen diese Probleme beim Gedanken-

experimentieren ebenfalls auftreten können, solche Ressourcen haben. Wir werden sehen, dass dieser Vorwurf hauptsächlich auf der Unkenntnis der philosophischen Diskussion beruht. Die Philosophie besitzt sehr wohl Ressourcen, um in Streitfällen eine Entscheidung über die Aussagekraft eines Gedankenexperiments an eine andere Subdisziplin zu delegieren. Auf diese Weise kann die systematische Philosophie es vermeiden, dauerhaft auf der Stelle zu treten.

#### 1.2.5 DER MODALE STATUS PHILOSOPHISCHER TARGETTHESEN

Weitere Einwände in der Stoßrichtung von (A) zielen eher auf inhaltliche Spezifika philosophischer Gedankenexperimente bzw. der von ihnen attackierten Theorien ab, den sogenannten „Targetthesen“ eines Gedankenexperiments. Gedankenexperimente zur Überzeugungsänderung können in der Philosophie funktional noch insofern weiter differenziert werden, als sie verschiedene Überzeugungstypen (also verschiedene Typen von Targetthesen) zum Gegenstand haben. Manche Überzeugungen, für die in der Philosophie argumentiert wird, haben den Status einer metaphysischen bzw. empirischen These, manche Überzeugungen eher den Status einer Bedeutungsanalyse oder den eines normativen Vorschlags, eine bestimmte Praxis auf bestimmte Weise zu regeln (sei es nun die Sprachpraxis im Fall einer Begriffsexplikation oder unser anderes soziales Handeln im Fall eines ethischen Normvorschlags). In jedem dieser Fälle sind der Sinn bzw. die Adäquatheit von Gedankenexperimenten gesondert zu motivieren und sind auch aus unterschiedlichen Gründen kritisiert worden.

Seit Kapitel 3.2 wissen wir, dass Gedankenexperimente zur Überzeugungsrevision gegen modale Thesen vorgebracht werden. Es lassen sich aber verschiedene Sinne unterscheiden, in denen eine These modaler Natur sein kann. Seit Kapitel 3.2 wissen wir auch, dass sich der Grad, zu dem wir gerechtfertigt sind, auf der Grundlage eines gedankenexperimentellen Arguments eine Überzeugung zu revidieren, (unter anderem) davon abhängt, zu welchem Grad wir gerechtfertigt sind anzunehmen, dass es sich bei der in der gedankenexperimentellen Situation *behaupteten* relevanten Möglichkeit tatsächlich um eine relevante Möglichkeit handelt.

Wenn nun philosophische Theorien von unterschiedlichem modalen Status sein können (also – beispielsweise – Notwendigkeitsbehauptungen unterschiedlicher Stärke darstellen können), ist anzunehmen, dass für die als Gegenbeispiel in Frage kommenden Möglichkeiten unterschiedliche Gründe vorgebracht werden müssen, aus denen man gerechtfertigt sein soll, zu glauben, dass die behauptete Möglichkeit tatsächlich besteht. Entsprechend gibt es auch unterschiedliche Kritiken an philosophischen Gedankenexperimenten, die nachzuweisen versuchen, dass man in Bezug auf eine bestimmte Art von Möglichkeit *prinzipiell* nicht berechtigt ist, zu glauben, dass es sich bei einer behaupteten Möglichkeit um eine tatsächliche Möglichkeit der relevanten Art handelt.

Um eine systematische Behandlung dieser Kritiken zu ermöglichen, wird in Teil 6 zunächst erläutert, welche Arten von Möglichkeiten und Notwendigkeiten unterschieden werden können (6.1). Sodann wird in Kapitel 6.2 gezeigt, welche Typen von Definitionen, die als Paradigmen für philosophische Theoriebildung herangezogen werden, welche Modalitäten involvieren.

Es kann dabei zwischen drei Typen von Definitionen unterschieden werden, die als Gegenstand philosophischer Kritik in Frage kommen: (i.) „Sachanalysen“ (empirische bzw. „metaphysische“ Analysen, die einen notwendigen Zusammenhang zwischen den Dingen behaupten, auf die sich die Ausdrücke in Definiens und Definiendum beziehen), (ii.) Bedeutungsanalysen (die einen notwendigen Zusammenhang zwischen den Ausdrücken aus Definiens und Definiendum behaupten) und (iii.) Explikationen (die die Bedeutung eines bereits in einer Sprache existierenden Ausdrucks für die Verwendung in wissenschaftlichen Theorien normativ einschränken). In Bezug auf jeden dieser Typen ist argumentiert worden, dass Gedankenexperimente für die Kritik einer solchen Definition kein sinnvolles Mittel sind. Wir werden diese Kritiken dementsprechend im Folgenden sortieren und der Reihe nach diskutieren.

#### 1.2.6 KRITIK AN PHILOSOPHISCHEN GEDANKENEXPERIMENTEN GEGEN REALDEFINITIONEN

Der siebte Teil dieser Arbeit verteidigt philosophische Gedankenexperimente, die gegen Realdefinitionen gerichtet sind, gegen mögliche und tatsächliche Kritiken und erläutert dabei ihre Funktionsweise, ihre erkenntnistheoretischen Probleme und ihre Adäquatheitsbedingungen.

Dieser Teil der Arbeit orientiert sich dabei an der in Teil 6 skizzierten funktionalen Ausdifferenzierung des Gedankenexperiments zur Überzeugungsänderung und untersucht die einzelnen an der Methode des Gedankenexperiments vorgetragenen Einwände in Bezug auf ihre Relevanz für diese zu unterscheidenden Funktionen.

In Kapitel 7.1 werden zunächst solche Kritiken an der Methode des Gedankenexperiments diskutiert, die Gedankenexperimente betreffen, die gegen Sachanalysen gerichtet sind. Wie wir sehen werden, orientieren sich die Kritiker an der Methode des Gedankenexperiments insbesondere an der Frage, ob die vom Gedankenexperiment attackierte Überzeugung *natürliche Arten* betrifft oder nicht.

So ist in der Debatte um die Zuverlässigkeit von Gedankenexperimenten zum Thema „personale Identität“ eingewendet worden, dass der Begriff der Person kein natürlicher Artbegriff sei (z.B. von Wilkes [345]).

Die hierzu komplementäre Auffassung, dass Gedankenexperimente in der Debatte um personale Identität inakzeptabel sind, weil der Personenbegriff ein natürlicher Artbegriff *ist*, ist ebenfalls vertreten worden (z.B. von Wiggins [343]).

Zwischen diesen beiden Positionen besteht allerdings nur scheinbar eine Spannung. Gemeinsam ist beiden Einwänden gegen die Methode des Gedankenexpe-

rimentierens eine bestimmte Bedeutungstheorie, insbesondere eine kausale Referenztheorie. Die Grundvorstellung dabei ist, dass durch bestimmte Tatsachen der Begriffseinführung und Begriffsverwendung ein natürlicher Artausdruck „rigide“ eine bestimmte Art bezeichnet. Was aber für die Mitglieder einer solchen Art möglich oder notwendig ist, fällt dann nicht mehr automatisch mit unseren *Überzeugungen* darüber zusammen, was für die Mitglieder dieser Art möglich oder notwendig ist. Wir beziehen uns auch dann de facto (über eine kausale Referenzkette) auf diese Arten, wenn wir radikal falsche Überzeugungen über die Mitglieder dieser Art haben. Wenn die Philosophie aber daran interessiert ist, herauszubekommen, was eine Person, Wissen oder Bewusstsein ist, und nicht primär an unserer *Meinung* darüber, was ‚Person‘, ‚Wissen‘ oder ‚Bewusstsein‘ bezeichnen, wie kann sie dann hoffen, dies über Gedankenexperimente herauszufinden?

Wie wir sagten, setzt dieser Einwand eine kausale Referenztheorie voraus. Um diesen Einwand beurteilen zu können, muss zunächst genauer geklärt werden, was eine kausale Auffassung der Referenz eigentlich besagt. Wie wir sehen werden, ist dieses Thema innerhalb der Philosophie hoch kontrovers, weshalb wir in unserer Verteidigung der Methode des Gedankenexperiments (gemäß unserer Aufgabenstellung) die kontroversen Standpunkte berücksichtigen müssen. Dies geschieht in Kapitel 7.2. Es werden dabei drei verschiedene Auffassungen in Bezug auf natürliche-Art-Begriffe unterschieden, die die Methode des Gedankenexperiments auf unterschiedliche Weise rekonstruieren. Der „modale Rationalismus“ ist dabei eine Position, die zwar im Sinne der kausalen Referenztheorie zwischen „epistemischen“ und „metaphysischen“ Möglichkeiten *unterscheidet*, aber den Bereich des Apriori nicht vom Bereich des metaphysisch Notwendigen *trennt*. Der modale Rationalismus behauptet vielmehr, dass es eine systematische Beziehung zwischen beiden Bereichen gibt. Diese Position könnte den Einwand gegen das Gedankenexperimentieren auch *mit* kausaler Referenztheorie abweisen. Allerdings ist die Position des modalen Rationalismus nicht unproblematisch.

Eine alternative Konzeption der kausalen Referenztheorie (die „Putnam-Orthodoxie“) *trennt* den Bereich des Apriori von dem des metaphysisch Notwendigen völlig ab und leugnet die systematischen Beziehungen, die der modale Rationalismus voraussetzt. Wäre auch unter einer so radikalen Auffassung Gedankenexperimentieren sinnvoll? Es wäre zunächst vermutlich mindestens so sinnvoll wie in den Naturwissenschaften, was aber nur bedingt beruhigend ist, da wir ja in Kapitel 3.2 gesehen haben, dass Gedankenexperimente in den Naturwissenschaften ebenfalls mit Rechtfertigungsproblemen zu kämpfen haben. Es könnte aber auch dann noch sinnvoll sein, wenn man die kritische Rolle des Gedankenexperiments auf die Bedeutungsanalyse beschränkt und die Kritik an Sachanalysen als indirekte Kritik (nicht an der Wahrheit, sondern an der Relevanz) dieser Analysen versteht.

Die Bedeutungsanalyse – so könnte man in der vermittelnden Position des „moderaten Deskriptivismus“ argumentieren – klärt uns ja immer noch darüber auf, was wir mit bestimmten Ausdrücken meinen, und dies kann dabei helfen, Erklärungsprobleme besser zu verstehen und Erklärungslücken zu entdecken, die aus einer Diskrepanz zwischen unserem Vorverständnis und der gelieferten wissenschaftlichen Erklärung resultieren können. Indem Gedankenexperimente in Be-

zug auf Bedeutungsanalysen solche Erklärungslücken aufdecken, dienen sie indirekt zur Kritik an Sachanalysen, da sie die Irrelevanz dieser Analysen für bestimmte Fragestellungen nachweisen.

Doch auch die Position, dass Gedankenexperimente hauptsächlich in der Bedeutungsanalyse eine Rolle spielen, ist nicht unproblematisch, wie wir in Kapitel 7.3 sehen werden. So lautet ein aktueller Einwand gegen Gedankenexperimente zur Kritik an Bedeutungsanalysen, dass die Intuitionen, auf deren Grundlage wir die Bedeutung eines Ausdrucks rekonstruieren, nicht kulturinvariant sind. Bedeutet das nicht, dass eine Analyse – beispielsweise – des Wissensbegriffs, die auf unseren intuitiven Reaktionen auf Gedankenexperimente aufgebaut ist, dann nur kulturrelativ von Relevanz sein kann? Wir werden argumentieren, dass diese Kulturrelativität unproblematisch ist, wenn man die Bedeutungsanalyse nicht als Universalienlinguistik missversteht. Wir klären Begriffe, weil die Art, wie wir sie verwenden, Anlass zu philosophischen Problemen gibt. Dass in anderen Kulturen „diese“ Begriffe anders verwendet werden, ist dabei unerheblich. Wir diskutieren dieses Thema im Detail in Kapitel 7.3.1.

In Kapitel 7.3.2 gehen wir einem Problem nach, das sich ebenfalls für Gedankenexperimente als Instrumente zur Kritik an Bedeutungsanalysen stellt. Bei der Anwendung der Methode des Gedankenexperiments scheinen wir implizit davon auszugehen, dass sich unsere Begriffe in hinreichende und notwendige Bedingungen analysieren lassen. Hat die moderne Psychologie nicht gezeigt, dass dies ein falsches Bild von Begriffen ist? Wie wir sehen werden, ist die Behauptung, die moderne Psychologie habe etabliert, dass Begriffe sich nicht mit hinreichenden und notwendigen Bedingungen angeben lassen, überzogen. Es *kann* aber sehr wohl sein, dass unsere Begriffsverwendung zu unsystematisch ist, um eine Rekonstruktion zuzulassen. Es gibt zwar keine guten Gründe, dies von irgendwelchen philosophisch interessanten Begriffen anzunehmen, aber falls dem so sein sollte, wäre das tatsächlich zunächst ein Problem für die Bedeutungsanalyse. Was zu solchen Fällen zu sagen ist, soll uns aber erst in Teil 9 beschäftigen.

### 1.2.7 VORSTELLBARKEIT UND MÖGLICHKEIT

In Teil 8 diskutieren wir einen Einwand gegen Gedankenexperimente, der unabhängig von unserer Unterscheidung zwischen Gedankenexperimenten gegen Sachanalysen und Gedankenexperimenten gegen Bedeutungsanalysen vorgetragen werden kann. In beiden Fällen wird ja davon ausgegangen, dass wir ein Verfahren besitzen, zu begründeten Urteilen darüber zu gelangen, was (im relevanten Sinne) möglich sein soll. Worin aber soll dieses Verfahren bestehen? Von einer Erkenntnismethode, die Teil einer Methodologie sein soll, scheint man doch fordern zu können, dass (a) sie hinreichend klar beschrieben werden kann, und (b) (auf irgendeiner Beschreibungsebene) angegeben werden kann, warum man dem so beschriebenen Verfahren trauen können soll.

So kann man in Bezug auf die Wahrnehmung (in gewissem Rahmen) positiv angeben, welche Umstände man möglichst herbeiführen sollte (günstige Lichtverhältnisse, optische Hilfsmittel, etc.), bzw. möglichst vermeiden sollte (Einfluss von Drogen und Alkohol, bunte Brillengläser in der falschen Stärke, etc.), will man zu zuverlässigen Urteilen über die Umgebung kommen. Außerdem lässt sich in Bezug auf unseren Wahrnehmungsapparat und die Wahrnehmungseigenschaften der Dinge in unserer Umgebung prinzipiell angeben, *warum* wir Wahrnehmungen trauen können, wenn sie nach dem angegebenen Verfahren zustande gekommen sind.

Bei Urteilen über Möglichkeiten stellen sich hier aber zwei Schwierigkeiten ein: Erstens ist es fraglich, ob sich tatsächlich eine hinreichend klare Beschreibung desjenigen Verfahrens finden lässt, auf dessen Grundlage wir zu Urteilen über Möglichkeiten kommen. Wie wir sehen werden, ist das Verfahren, mit dem Philosophen zu Urteilen über Möglichkeiten gelangen, das Verfahren des „Vorstellbarkeitstests“. Allerdings ist ‚es ist vorstellbar, dass p‘ extrem vieldeutig. Unter diesen Bedeutungen diejenige herauszufinden, von der man annehmen darf, dass Philosophen diese meinen, wenn sie von der Vorstellbarkeit auf die Möglichkeit schließen, ist kein besonders leichtes Unterfangen. Stephen Yablo hat eine (vergleichsweise klare) Explikation vorgeschlagen, von der es aber immer noch fraglich ist, ob sie für die Zwecke einer Methodologie hinreichend verständlich ist.

Die zweite Schwierigkeit besteht darin, dass, selbst wenn man eine nachvollziehbare Charakterisierung des fraglichen Verfahrens hätte, es trotzdem prinzipielle Schwierigkeiten gäbe, zu erklären, *warum* es funktioniert.

Möglichkeiten sind – vereinfacht gesagt – diejenigen Sachverhalte, die mit einer bestimmten Menge von Beschränkungen (den Notwendigkeiten) kompatibel sind. In der Bedeutungsanalyse (wie auch in der Metaphysik) wird – nach allem, was wir bisher gesagt haben – versucht, unser implizites Wissen von diesen Beschränkungen explizit zu machen. Wir sind davon ausgegangen, dass die relevanten Beschränkungen intern (aber unbewusst) repräsentiert sind und die Grundlage unseres intuitiven Urteilens über Möglichkeiten bilden. Für die Erklärung eines beliebigen (!) Verfahrens (Vorstellbarkeitstest oder was auch immer) auf Sachverhalte zu schließen, die mit den relevanten Beschränkungen kompatibel sind, sollte es also auf der Ebene der rationalen Rekonstruktion eine Schlussweise geben, die unter Idealbedingungen „auf die richtige Weise“ von den Beschränkungen zu den Kompatibilitätssurteilen führt. ‚Auf die richtige Weise‘ bezeichnet hierbei einen *gerechtfertigten* Übergang von Überzeugungen (bestimmter Art) auf andere Überzeugungen (bestimmter Art). Setzt man die Standards für gerechtfertigte Übergänge sehr hoch – etwa auf die Höhe der klassischen deduktiven Logik – und lässt nur wahrheitserhaltende Übergänge als gerechtfertigt zu, dann lässt sich zeigen, dass es selbst unter Idealbedingungen keinen gerechtfertigten Übergang von den Beschränkungen zu den mit ihnen kompatiblen Sachverhalten gibt, wenn die Beschränkungen eine gewisse (minimale) Komplexität aufweisen.

Überzeugte Deduktivisten, die letztendlich nur *wahrheitserhaltende* Übergänge als gerechtfertigt akzeptieren, könnten ein solches Verfahren wie den Vorstellbarkeitstest demnach *prinzipiell* nicht als ein Verfahren zur Beurteilung der Kompati-

bilität von Sachverhalten mit (implizit bekannten) Beschränkungen akzeptieren. Erst außerhalb der klassischen Logik (wie beispielsweise in der „Adaptiven Logik“) kann ein solches Verfahren als „logischer“ Schluss rational rekonstruiert werden.

Wollen wir an der klassischen Logik und einem strengen Deduktivismus festhalten, könnten wir noch die Annahme aufgeben, dass unser Bedeutungswissen (bzw. unser Wissen über metaphysische Zusammenhänge) in Form von Notwendigkeiten repräsentiert ist, und statt dessen von der (nicht besonders plausiblen) Annahme ausgehen, dass sich unser implizites Wissen über die Bedeutung von Ausdrücken oder die Metaphysik aus Wissen über mögliche Sachverhalte zusammensetzt. (In diesem Fall macht ein Vorstellungstest nur bewusst, was wir unbewusst wissen, ohne dass dabei von Überzeugungen bestimmter Art auf andere Überzeugungen geschlossen wird). Angesichts unserer endlichen Kapazitäten und der damit einhergehenden Limitierung hinreichend viele solcher Möglichkeiten für die „metaphysische“ Analyse oder die Begriffsanalyse zur Verfügung zu haben, macht diese Alternative beide zu einem eher konstruktiven als rekonstruktiven Unternehmen. Damit verlassen wir offenbar den Bereich der Analyse und betreten den Fall der Explikation, der den Schluss unserer Untersuchung bildet.

#### 1.2.8 DIE LEISTUNGSFÄHIGKEIT PHILOSOPHISCHER GEDANKENEXPERIMENTE

Selbst innerhalb der analytischen Philosophie gibt es kaum eine Auffassung, die wirklich universell geteilt wird und dazu spürbar über den Grad einer Trivialität hinauskommt. Auch die Logik ist kein einheitliches Werkzeug zur philosophischen Forschung. Die besondere Qualität der analytischen Philosophie (die sie vor ihren „Konkurrenten“ auszeichnet) liegt unseres Erachtens aber auch nicht darin, dass sie eine gut fundierte methodische Basis hätte, sondern darin, dass sie die Methoden, die sie anwendet, mit kritischem Bewusstsein anwendet. Wenn in der Philosophie der Logik diskutiert wird, ob die Prädikatenlogik erster Stufe den Begriff der logischen Folgerung richtig expliziert, oder ob eine nicht-klassische Logik das nicht besser leistet, ist dies kein Hinweis auf eine Grundlagenkrise der Philosophie, sondern zeigt, dass alles in den richtigen Bahnen läuft. Die Uneinigkeit darüber, welche Werkzeuge für die philosophische Forschung legitim sind, sollte in der Forschungspraxis aber nicht dazu führen, den Gebrauch von bestimmten Werkzeugen immer dann, wenn es gerade passt, für illegitim zu erklären.

Wenn man eine philosophische Position mit einer *reductio ad absurdum* zur Strecke gebracht zu haben scheint, hat man nicht nur ein Schulterzucken verdient und den Hinweis, dass *reductio*-Beweise ja gar nicht immer so unproblematisch sein sollen. Wer einer *reductio* aus dem Weg gehen möchte, sollte in der Lage sein, anzugeben, weshalb er dieser Meinung ist und sich auch im Klaren darüber sein, welche Probleme eine solche Position möglicherweise mit sich bringt. Dies sollte auch für die Methode des Gedankenexperiments gelten. Wenn man sie ablehnt, braucht man Gründe.

Wie wir in Teil 9 sehen werden, reichen die Gründe, die bisher gegen das Gedankenexperiment vorgetragen wurden hierzu nicht aus. Selbst eine naturalistische Position, die sich mit Begriffsexplikationen begnügen möchte, also weder besonderes Vertrauen zur Umgangssprache (wie es der modale Rationalismus hat), noch besonderes Interesse an der Umgangssprache (wie es die Bedeutungsanalyse hat) besitzt, muss zum Zweck der Explikation normativ Adäquatheitskontexte festlegen, die das Explicatum erfüllen muss. Wie wir sehen werden, ist das Verfahren, mit dem solche Kontexte festgelegt werden, dasselbe Verfahren wie das, mit dem modale Rationalisten über metaphysische Möglichkeiten forschen. D.h., auch wenn man die weitergehenden Geltungs- und/oder Erkenntnisansprüche eines modalen Rationalismus ablehnt, sind kritisch gemeinte Gedankenexperimente dennoch zu prüfen, da sie in der Methodologie der Begriffsexplikation als relevante Adäquatheitskontexte uminterpretiert werden können. Dieser Fall ist aber praktisch ein *worst case*-Szenario für die Verteidigung des Gedankenexperimentierens. Wenn tatsächlich argumentiert werden kann, dass Gedankenexperimente auch unter diesen Annahmen sinnvolle methodische Werkzeuge sind, ist diese Verfahrensweise rehabilitiert.

## 2. Das Gedankenexperiment in der Wissenschaftstheorie der Naturwissenschaften

Dieser Teil widmet sich der Geschichte des Ausdrucks ‚Gedankenexperiment‘ in der *Wissenschaftstheorie* der Naturwissenschaften. Ausgehend von der Frage, wie und wozu dieser Begriff in die Wissenschaftstheorie eingeführt wurde, soll dann in Kapitel 3 zusammenfassend dargestellt werden, zu welchen Einsichten die Wissenschaftstheorie der Naturwissenschaften in Bezug auf die Methodologie des Gedankenexperiments gelangt ist.

Will man eine gehaltvolle Charakterisierung von Gedankenexperimenten zum Zweck einer eventuellen Übertragung methodologischer Einsichten von den Naturwissenschaften auf die Philosophie, dann sollte man sich ansehen, welche Episoden in der Wissenschaftsgeschichte der Naturwissenschaften aus welchen Gründen von Wissenschaftstheoretikern als „Gedankenexperimente“ bezeichnet worden sind, und untersuchen, welche (vermeintlichen<sup>19</sup>) funktionalen Gemeinsamkeiten und Unterschiede unter diesen paradigmatischen Fällen ausgemacht werden können. Auf diese Weise kann man einen eingeschränkten Begriff des Gedankenexperiments entwickeln, für den man hoffen darf, dass er möglichst fruchtbar auf die Philosophie angewendet werden kann. Dies soll in diesem und dem nächsten Kapitel geschehen.<sup>20</sup>

---

<sup>19</sup> Da wir ja den Absichten der Wissenschaftstheoretiker folgen, wenn wir uns auf die Suche nach einer Definition von ‚Gedankenexperiment‘ begeben, kann es natürlich sein, dass in der *Wissenschaftsgeschichte* sich *de facto* nichts findet, was diese Definition erfüllt. Wir werden an einigen wissenschaftshistorischen Beispielen in Kapitel 3 sehen, dass dem glücklicherweise nicht so ist.

<sup>20</sup> Dieser Weg erscheint mir aus verschiedenen Gründen als der sinnvollste. So spricht zunächst einmal Vieles dafür, dass man Gedankenexperimente nur über ihre funktionale Rolle charakterisieren kann: *Erstens* kann man angesichts der heterogenen Verwendungsweise des Ausdrucks ‚Gedankenexperiment‘ Gedankenexperimente vermutlich nicht funktionsunabhängig charakterisieren und dann in einem zweiten Schritt eine sinnvolle methodologische Funktion dafür suchen. Eine solche Charakterisierung wäre nahezu leer. So wenig, wie die Tatsache, dass ein Chemiker beim Kaffekochen aus Schusseligkeit jeden Morgen die Anzahl der Kaffeelöffel variiert, als „Experiment“ bezeichnet werden kann, ist jeder Gedanke an ein Experiment durch einen Wissenschaftler automatisch ein Gedankenexperiment. *Zweitens* ist es so, dass selbst wenn man Gedankenexperimente gehaltvoll charakterisieren könnte, ohne funktionale Terminologie zu benutzen, dies freilich auch kein Grund wäre, Gedankenexperimente *nicht* funktional zu charakterisieren. Zum einen ist es bestenfalls ein Streit um Worte, zum anderen geht es in dieser Untersuchung um *methodologische* Fragen. Methodologische Fragen stellen sich überhaupt nur in Bezug auf *Funktionen* und *Zwecke*.

## 2.1 ‚GEDANKENEXPERIMENT‘ BEI ERNST MACH

Es wird häufig darauf hingewiesen<sup>21</sup>, dass Hans Christian Ørsted den Begriff des Gedankenexperiments erfunden hat. In der Tat scheint er der erste zu sein, der von ‚Gedankenexperimenten‘ oder ‚Gedankenversuchen‘ gesprochen hat und damit eine Methode bezeichnen wollte, die sich sowohl in den Naturwissenschaften, wie auch in der Mathematik finde, und für die Immanuel Kants Naturlehre aus *Metaphysische Anfangsgründe der Naturwissenschaft* die „schönsten Beyspiele“<sup>22</sup> liefere.

Obzwar er damit ein neues Wort gebildet hatte, kann man sich fragen, ob es Ørsted gelungen ist, dieses auch mit einem für unsere Zwecke relevanten Inhalt zu füllen. Zunächst kann man feststellen, dass ihm sehr wahrscheinlich nicht gelungen ist, ein neues Thema in die Wissenschaftsphilosophie zu bringen. Ørsted blieb nach allem, was wir wissen, ohne Wirkung und die Begriffsgeschichte des Gedankenexperiments beginnt erst mit Ernst Mach.<sup>23</sup>

Darüber hinaus kann man zeigen, dass Ørsted den Ausdruck außerdem im Kern für Methoden verwendet wissen wollte, die für alle späteren Autoren der Wissenschaftstheorie bestenfalls zu einer Nebenbedeutung von ‚Gedankenexperiment‘ gehören. Wie ich an anderer Stelle (vgl. Cohnitz [74]) ausführlich argumentiert habe, ging es Ørsted im Kern wohl darum, eine Lanze für gewisse umstrittene mathematische Betrachtungsweisen in der Infinitesimalrechnung zu brechen, ohne diese Methode aber in deutlicher Weise in den später für den Begriff des Gedankenexperiments paradigmatisch werdenden Episoden der Wissenschaftsgeschichte ebenfalls am Werk zu sehen.

### 2.1.1 ERNST MACH UND DER BEGINN DER BEGRIFFSGESCHICHTE

Ernst Mach (1838-1916) beschäftigt sich mit dem Gedankenexperiment zum ersten Mal in seiner 1883 verfassten Wissenschaftsgeschichte *Die Mechanik in ihrer Entwicklung – Historisch-kritisch dargestellt*<sup>24</sup>.

<sup>21</sup> Zum ersten Mal bei Witt-Hansen [350], aber auch in Pulte [262], Kühne [174], Kühne [175], Kühne [176].

<sup>22</sup> Ørsted [246], 482f.

<sup>23</sup> Zu Ørsted vgl. Kühne [176] und Cohnitz [74].

<sup>24</sup> Dieses Werk wurde von Mach mehrfach überarbeitet. Der hier verwendete Text (Mach [193]) entspricht der siebten Auflage von 1912 und gibt vermutlich kein getreues Bild der Ansichten Machs um 1883. Die genaue Genese der Machschen Überlegungen zum Thema Gedankenexperiment ist hier aber auch nur von nebensächlicher Bedeutung. Vgl. Kühne [175], 132. Ob Mach Ørsteds Arbeit zum Gedankenexperiment kannte, ist unklar. Falls dem so ist, hat Mach sich von Ørsted den Ausdruck geborgt und ihn mit neuem Inhalt gefüllt. Vgl. Cohnitz [74].

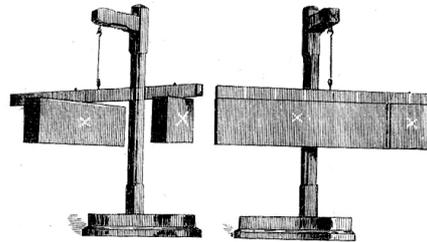


Abbildung 2.1-1

Als Erläuterung sei hier kurz eine „Ableitung“ des Hebelgesetzes betrachtet (nach Kühne [175], 321-313, obige Abbildung stammt aus Mach [193], 40). In der obigen Abbildung zeigt die rechte Seite zunächst eine symmetrische Balkenwaage. Dies entspricht dem „speziellen Fall“, der durch die Prämisse (a.) ausgedrückt und vorausgesetzt wird. Wir zerlegen nun den homogenen Balken asymmetrisch in zwei kleinere Balken. Beide Balken sind wieder jeweils an ihrem Mittelpunkt aufgehängt und sind damit Anwendungsfälle des „speziellen“ Falles, nämlich des symmetrischen Hebelgesetzes. In obiger Abbildung sind die Mittelpunkte dieser Balken mit ‚X‘ kenntlich gemacht.

„Die bloße Iteration der Anwendung des symmetrischen Hebelgesetzes erzwingt scheinbar die Antiproportionalität des allgemeinen Hebelgesetzes: Offensichtlich wird bei jeder Zerlegung der Balkenwaage in zwei kleinere das Gewicht einer Teil-Balkenwaage linear mit ihrer Länge wachsen und der Abstand ihres Schwerpunkts (X) von der Aufhängung der Gesamt-Balkenwaage linear schrumpfen. Also wird das Produkt aus Gewicht und Abstand vom Aufhängungspunkt der rechten Teil-Balkenwaage immer gleich dem der linken sein.“ (Kühne [175], 313)

Der empirische Gehalt der in dieser Pseudoableitung „eingeschmuggelten“ Prämisse wird durch den linken Teil der Abbildung wiedergegeben. Dass die Gesamtbalkenwaage auch im Gleichgewicht bleibt, wenn man die Teil-Balkenwaagen um ihre Aufhängungspunkte dreht, ist empirisch so, aber nicht als *logische Folge* des symmetrischen Hebelgesetzes.

Machs erklärte Absicht, die er mit dieser Schrift verfolgt, ist eine „aufklärende“ (im Sinne von „antimetaphysische“) Darstellung des „naturwissenschaftlichen“ Inhalts der Mechanik zu geben, bei der die Quellen dieses Wissensbestandes, wie auch der Grad an Gewissheit, den wir ihm zuschreiben dürfen, offen gelegt werden soll. Machs Darstellung deckt dabei die Entwicklung der Mechanik von der Entdeckung des Hebelgesetzes bis zu seiner Gegenwart.

Ohne sie als ‚Gedankenexperiment‘ zu bezeichnen, behandelt Mach als erstes die Ableitung des allgemeinen Hebelgesetzes durch Archimedes und seine historischen Nachfolger. Worauf Mach in seiner Darstellung insbesondere abhebt, ist, dass sowohl bei Archimedes, wie bei Galilei, Stevin, Lagrange und Huygens in dieser „Ableitung“ gepfuscht wird:

So überraschend uns nun auf den ersten Blick die Leistung von Archimedes und seinen Nachfolgern erscheint, so steigen uns bei genauer Betrachtung doch Zweifel an der Richtigkeit derselben auf. Aus der bloßen Annahme des Gleichgewichts gleicher Gewichte wird die verkehrte Proportion zwischen Gewicht und Hebelarm abgeleitet! Wie ist das möglich? (Mach [193], 38)

Die Voraussetzungen des Archimedischen Beweises sind zunächst die Folgenden:

- a. Gleichschwere Größen, in gleicher Entfernung (vom Unterstützungspunkt) wirkend, sind im Gleichgewicht.
- b. Gleichschwere Größen, in ungleicher Entfernung (vom Unterstützungspunkt) wirkend, sind nicht im Gleichgewicht.

Hieraus leitet Archimedes (und nach ihm Galilei, Stevin, Lagrange und Huygens) das spezielle Hebelgesetz ab:

- c. Kommensurable Größen sind im Gleichgewicht, wenn sie ihrer Entfernung (vom Unterstützungspunkt) umgekehrt proportioniert sind.

Mach weist zunächst darauf hin, dass bereits die Prämissen (a.) und (b.) dieses Beweises empirische Annahmen machen (darüber, dass außer dem Gewicht und der Entfernung vom Unterstützungspunkt nichts sonst eine Rolle spielt, wie beispielsweise die Farbe der Hebelarme oder die Position eines Beobachters). Die darin enthaltene Erfahrung sei zwar „unwillkürlich“ und „instinktiv“, und die Prämissen daher durchaus evident, aber eben aus der Erfahrung.

Wenn die Prämissen sich aber bereits als voraussetzungsreich entpuppen, wie viel mehr verwundert es, dass das allgemeine Hebelgesetz aus diesen Annahmen einfach folgt:

Wenn wir schon die bloße Abhängigkeit des Gleichgewichts vom Gewicht und Abstand überhaupt nicht aus uns *herausphilosophieren* konnten, um wieviel weniger werden wir die Form dieser Abhängigkeit, die Proportionalität, auf spekulativem Wege finden können. (Mach [193], 38)

In der Tat lässt sich Archimedes und seinen Nachfolgern auch nachweisen, dass zusätzliches Erfahrungswissen in diese „Ableitungen“ eingeschmuggelt wird (vgl. Abbildung 2.1-1). Damit soll freilich keineswegs das Hebelgesetz in irgendeiner Form diskreditiert werden, aber dieser Nachweis dient dem Machschen Ziel, die Aposteriorizität unseres physikalischen Wissens aufzuweisen:

Daß ein solcher Vorgang [der Ableitung des allgemeinen Hebelgesetzes aus den Spezialfällen und zusätzlichem empirischen Wissen] möglich und zulässig, auf einer gewissen Stufe der Forschung sogar sehr fruchtbar, vielleicht der einzig richtige ist, habe ich nirgends bestritten, im Gegenteil, durch die Art, wie ich die nach dem Muster des Archimedes angelegten Ableitungen Stevins und Galileis dargestellt habe, ausdrücklich anerkannt. Mein ganzes Buch verfolgt aber das Ziel, den Leser zu überzeugen, daß man *Eigenschaften*

der Natur nicht mit Hilfe selbstverständlicher Annahmen aus den Fingern saugen kann, sondern daß diese aus der *Erfahrung* entnommen werden müssen. (Mach [193], 44)

Nach dem Hebelgesetz kommt Mach darauf zu sprechen, dass das *Prinzip der schiefen Ebene* (auf schiefen Ebenen von gleicher Höhe wirken gleiche Gewichte im umgekehrten Verhältnis der Längen der schiefen Ebenen) durch eine von den Überlegungen zum Hebelgesetz völlig unabhängige Entwicklung gefunden wurde, obzwar Galilei bereits wusste, dass durch das Hebelgesetz alleine das Prinzip der schiefen Ebene abgeleitet werden kann.<sup>25</sup>

Simon Stevin habe das Prinzip auf „ganz originelle Weise“ untersucht. Mach schildert daraufhin Stevins Gedankenexperiment, als dessen Resultat die Verallgemeinerung, dass auf schiefen Ebenen von gleicher Höhe gleiche Gewichte im umgekehrten Verhältnis der Längen der schiefen Ebenen wirken, wie selbstverständ-

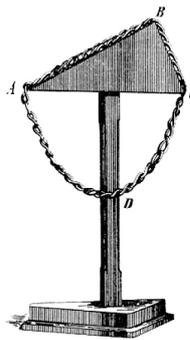


Abbildung 2.1-2

„Stevin geht etwa folgender Art vor. Er denkt sich ein dreiseitiges Prisma mit horizontalen Kanten, [...] *ABC* [...]. Hierbei soll beispielsweise  $AB = 2 BC$  und *AC* horizontal sein. Um dieses Prisma legt Stevin eine in sich zurücklaufende Schnur mit 14 gleich schweren, gleich weit abstehenden Kugeln. Wir können dieselbe mit Vorteil durch eine geschlossene gleichmäßige Kette oder Schnur ersetzen. Die Kette wird entweder im Gleichgewicht sein oder nicht. Nehmen wir das letztere an, so muß die Kette, weil sich bei ihrer Bewegung die Verhältnisse nicht ändern, wenn sie einmal in Bewegung ist, fortwährend in Bewegung bleiben, also ein Perpetuum mobile darstellen, was Stevin absurd erscheint. Demnach ist nur der erste Fall denkbar. Die Kette bleibt im Gleichgewicht. Dann kann der symmetrische Kettenteil *ADC* ohne Störung des Gleichgewichts entfernt werden. Es hält also das Kettenstück *AB* dem Kettenstück *BC* das Gleichgewicht. Auf schiefen Ebenen von gleicher Höhe wirken demnach gleiche Gewichte im umgekehrten Verhältnis der Längen der schiefen Ebene.“ (Mach [193], 49-50)

lich folgt. Wir wollen uns dieses Gedankenexperiment im Detail im nächsten Teil dieser Untersuchung ansehen, Abbildung 2.1-2 gibt aber hier schon Machs Rekonstruktion als Verständnishilfe.

Nach dem, was Mach wenige Seiten zuvor über das Ziel seines ganzen Buches gesagt hat, nämlich „den Leser zu überzeugen, daß man Eigenschaften der Natur nicht mit Hilfe selbstverständlicher Annahmen aus den Fingern saugen kann“, erwartet man natürlich eine fulminante Demaskierung des Etikettenschwindlers Stevin. Schließlich gibt dieser – zumindest laut dem, was Mach uns berichtet – keinerlei Hinweis auf empirische Untersuchungen. Was als hauptsächliche Prämisse in Stevins Überlegung einzugehen scheint, ist die Aussage, dass ein Perpetuum mobile einer bestimmten Art „Stevin absurd erscheint“, nicht aber, dass diese Überzeugung durch Blut, Schweiß und Tränen des Experimentalphysikers gewonnen wurde.

Diese Erwartung wird von Mach jedoch gehörig enttäuscht. Ganz im Gegenteil stellt sich Mach vor Stevin und verteidigt dessen Argumentationsweise:

In der Annahme, von welcher Stevin ausgeht, daß die geschlossene Kette sich nicht bewegt, liegt ohne Frage nur eine *ganz instinktive* Erkenntnis. Er fühlt sofort, und wir mit ihm, daß wir etwas einer derartigen Bewegung Ähnliches nie beobachtet, nie gesehen haben, daß dergleichen nicht vorkommt. Diese Überzeugung hat eine solche *logische Gewalt*, daß wir die hieraus gezogene Folgerung über das Gleichgewichtsgesetz der schiefen Ebene ohne Widerrede annehmen, während uns das Gesetz als bloßes Ergebnis des Versuchs oder auf eine andere Art dargelegt zweifelhaft erscheinen würde. (Mach [193], 50)

Bei empirischen Versuchen habe man schließlich immer mit störenden Einflüssen, wie z.B. Reibung zu kämpfen, die unter Umständen am Resultat Zweifel nähren könnten. Stevin versuche daher gar nicht erst mit empirischen Belegen zu überzeugen, sondern mit Einsichten, denen eine „höhere Autorität“ zukommt:

Woher kommt diese höhere Autorität? Erinnern wir uns, daß der wissenschaftliche Beweis, die ganze wissenschaftliche Kritik nur aus der Erkenntnis der eigenen Fehlbarkeit der Forscher hervorgegangen sein kann, so liegt die Aufklärung nicht weit. Wir fühlen deutlich, daß wir selbst zu dem Zustandekommen einer instinktiven Erkenntnis nichts beigetragen, daß wir nichts willkürlich hineingelegt haben, sondern daß sie ganz ohne unser Zutun da ist. Das Mißtrauen gegen unsere eigene subjektive Auffassung des Beobachteten fällt also weg. (Mach [193], 50)

Was soll das heißen? Warum ist bei Stevin so eine Argumentation plötzlich in Ordnung, bei Archimedes, Galilei, Lagrange und Huygens aber nicht?

Wie Mach rasch betont, sind solche instinktiven Erkenntnisse nicht unfehlbar. Obzwar sie eine „große logische Kraft“ besitzen, kann Erfahrung sie offenbar erschüttern. Als Beispiel nennt Mach genau das Phänomen, was den bereits erwähn-

---

<sup>25</sup> Vgl. Mach [193], 47.

ten Ørsted als Naturforscher über die Grenzen Dänemarks hinaus berühmt gemacht hat, die senkrechte Ausrichtung einer Magnetnadel zu einem Strom durchflossenen Leiter. Dieses Phänomen steht im Gegensatz zur „instinktiven“ Auffassung, dass Abstoßungs- und Anziehungskräfte immer nur entlang einer Linie wirken.<sup>26</sup>

Dass aber unser instinktives Wissen fehlbar ist, macht die Sache nicht unbedingt besser. Was eigentlich interessieren sollte, ist, warum Stevin seiner instinktiven Erkenntnis plötzlich vertrauen darf, warum Stevins Vorgang „kein Fehler“ ist.

Die erste Hälfte von Machs Antwort besteht darin, die Quelle unserer instinktiven Erkenntnis anzugeben. Wir besitzen einen unanalysierten, zum Teil unterbewussten Erfahrungsschatz, der uns offenbar durch die Evolutionsgeschichte mitgegeben ist. Dieser Erfahrungsschatz sei hauptsächlich negativ und bestehe in Wissen darüber, was unmöglich ist (wie in Stevins Fall in dem Wissen, dass ein Perpetuum mobile unmöglich ist). Wenn wir wissen wollen, was es erklärt, dass wir solch instinktives Wissen von Prinzip X haben, sei häufig die Erklärung dafür Prinzip X:

Unter welchen Bedingungen konnte die gegebene instinktive Erkenntnis entstehen? Gewöhnlich finden wir dann, daß *dasselbe* Prinzip, zu dessen Begründung wir die instinktive Erkenntnis herangezogen haben, wieder die *Grundbedingung* für das Entstehen dieser Erkenntnis bildet. Das ist auch ganz unverfänglich. Die instinktive Erkenntnis leitet uns zu dem Prinzip, welches sie selbst erklärt und welches durch deren Vorhandensein, das ja eine Tatsache für sich ist, wieder gestützt wird. (Mach [193], 52)

Auch hier ist nicht ganz klar, was Mach meint. Soll das Prinzip der schiefen Ebene erklären können, warum wir instinktiv glauben, dass eine Kette und ein Prisma kein Perpetuum mobile bilden? Das „erklärt“ vielleicht die *Wahrheit* dieses „Prinzips“, aber doch noch nicht die „Entstehung dieser Erkenntnis“. Oder geht es darum, dass die Entstehung einer „Erkenntnis“ partiell durch ihre Wahrheit erklärt wird?<sup>27</sup> Letzteres erklärt uns aber noch nicht, warum Stevin eine empirische Vermutung als Prinzip in einem Beweis verwenden darf, Archimedes et al. aber nicht.

Mach holt an dieser Stelle weiter aus. Die Argumentation Stevins erscheine uns so „geistreich“, weil sein Resultat seine Voraussetzungen zu transzendieren scheine. Stevin hätte sein Resultat aber auch aus anderen bereits etablierten Prinzipien beweisen können. Stevin hätte außerdem – wie auch Galilei im Fall des Hebelgesetz-

---

<sup>26</sup> Diese „instinktive“ Auffassung hat Kant als synthetisches Apriori in den *Metaphysischen Anfangsgründen* angeführt. Vgl. Cohnitz [74] und Kühne [176].

<sup>27</sup> Weil Erkenntnisse etwa *wahre* Überzeugungen sind. Was Mach vermutlich meint, ist dass die Wahrheit des Prinzips der schiefen Ebene Teil der Erklärung sein wird, warum unsere Intuitionen durch die Evolutionsgeschichte so gebildet wurden, dass sie zu diesem Prinzip konvergieren. Jeder Umstand, der kausal Einfluss auf die Ausbildung unserer Intuitionen hatte, war ein Umstand in dem das Prinzip galt. Die Wahrheit des fraglichen Prinzips ist allerdings weder notwendig (wie Mach hier selbst zugibt, vgl. auch Nozick [244]), noch hinreichend (vgl. Maffie [198]).

zes – das verwendete Prinzip aus einer Reihe von Experimenten finden können, welche sich aber erübrigten, da „der Erfolg nicht zweifelhaft“ gewesen sei.

Bis zu dieser Stelle hat Mach nicht viel mehr getan als Stevins Gedankengang zu loben. In Anbetracht der Tatsache, dass Stevin in der Tat ein wahres Prinzip gefunden hat, gäbe es ja auch zunächst gar keinen Grund zur Beschwerde. Aber genau dies müssten Archimedes et al. doch auch für sich reklamieren können.

Es wird sich der Mühe lohnen, noch einmal die Übereinstimmung und den Unterschied in dem Gedankengang von Stevin und Archimedes zu betrachten. Beide gehen vom Instinktiven aus. Stevin hat aber die sehr allgemeine Einsicht gewonnen, daß eine leichtbewegliche, schwere, geschlossene Kette von beliebiger Form in Ruhe bleibt. Er kann hieraus ohne Schwierigkeit quantitativ leicht übersehbare spezielle Fälle ziehen. Der Fall, von welchem Archimedes ausgeht, ist hingegen der denkbar *speziellste*. Aus demselben kann er unmöglich in einwandfreier Weise das Verhalten unter *allgemeineren* Bedingungen ableiten. Wenn es ihm scheinbar gelingt, so liegt dies daran, daß er den Fall schon kennt, während Stevin das Gesuchte ohne Zweifel annähernd auch schon kennt, aber auf dem eingeschlagenen Wege auch direkt hätte finden können. (Mach [193], 54)

Archimedes und Stevin schmuggeln zwar beide eine Erfahrungserkenntnis in ihre Ableitung, die Erfahrungserkenntnis Stevins ist aber (zumindest nach Mach) allgemeiner Art und besteht nicht wie bei Archimedes et al. einfach in der Voraussetzung des zu Beweisenden.

Soweit scheint die Verwendung instinktiver Erkenntnis, wie sie im Gedankenexperiment Stevins zutage gefördert wird, als grundsätzlich legitim, ja sogar vorteilhaft, da im Gedankenexperiment durch Idealisierungen von störenden Umständen abstrahiert werden kann, die eine empirische Bestätigung immer anzweifelbar erscheinen lassen. Ausnahmen bilden Fälle, in denen das eigentliche Beweisziel als instinktive Erkenntnis eingeschmuggelt wird. Es gibt aber noch andere Gründe dafür, dass der Bezug auf instinktive Erkenntnis einer Pseudoableitung manchmal nicht überzeugen kann.

Machs nächstes negatives Beispiel dafür, „daß man Eigenschaften der Natur nicht mit Hilfe selbstverständlicher Annahmen aus den Fingern saugen kann“, ist Daniel Bernoullis geometrischer „Beweis“ des Satzes des Kräfteparallelogramms:

Der Satz des Kräfteparallelogramms war Bernoulli als ein Erfahrungssatz bereits bekannt. Was Bernoulli tut, besteht also darin, daß er sich vor sich selbst *unwissend stellt* und den Satz aus möglichst wenigen Voraussetzungen herauszuphilosophieren sucht. Diese Arbeit ist keineswegs sinnlos und zwecklos. Im Gegenteil, man findet durch dieses Verfahren, wie wenige und wie unscheinbare *Erfahrungen* den Satz schon geben. (Mach [193], 65)

Bis dahin macht Bernoulli also noch nichts Verwerfliches (trotz der leicht negativen Darstellung Machs durch die pejorative Verwendung von ‚herausphilosophieren‘). Stevin ist in seinem Gedankenexperiment schließlich auch nicht anders vorgegangen.

Diesmal besteht Bernoullis „Vergehen“ vielmehr darin, dass er zur Ableitung des Satzes vom Kräfteparallelogramm eine Erfahrungserkenntnis hinzuziehen muss, die allgemeiner ist als das Parallelogrammprinzip und *außerdem* schwerer zu beobachten (nämlich die Erfahrungserkenntnis, dass jedes an einem Punkt angreifende Kraftsystem durch eine Kraft von bestimmter Größe und Richtung ersetzbar ist). Bernoullis Rückführung des leichter Beobachtbaren auf das schwerer Beobachtbare wird von Mach als „Verstoß gegen die Ökonomie“ der Wissenschaft betrachtet und gegen Bernoullis Ableitung vorgebracht.

Diese Diagnose steht etwas in Spannung mit dem zuvor Gesagten. Wir erinnern uns, dass Stevins Gedankenexperiment im Gegensatz zur Archimedischen „Ableitung“ deswegen „kein Fehler“ war, weil eine allgemeinere Erkenntnis zur Ableitung einer spezielleren verwendet wurde. Bernoulli scheint nun aber gerade dies vorgeworfen zu werden. Die Auflösung liegt wohl darin, dass die allgemeinere Erkenntnis Stevins, dass ein Perpetuum mobile einer bestimmten Art unmöglich ist, zumindest nicht schwerer durch Beobachtung gewonnen werden kann, wie das Prinzip der schiefen Ebene selbst, Stevin das Ökonomieprinzip also deshalb nicht verletzt.

Es lohnt vermutlich nicht, in exegetischer Absicht die von Mach vermuteten Unterschiede im Detail nachzurekonstruieren. In seinem ‚Rückblick auf die Entwicklung der Statik‘ ebnet Mach ohnehin die an den einzelnen Pseudoableitungen monierten Unterschiede vom Stevinschen Fall kurzerhand ein, zumal sie sich in einer reifen Wissenschaft solche Pseudoableitungen als überflüssig erweisen.<sup>28</sup>

Wird ein Prinzip neu entdeckt, sei es „ganz in der Ordnung“, dass man versucht, die neue Entdeckung auf irgendeine Art zu beweisen. Diese Versuche würden aber überflüssig, wenn sich das Prinzip bewährt hat. Von da ab betrachte man es als empirisch genauso gut belegt, wie diejenigen Prinzipien, die man zuvor zu seiner Ableitung herangezogen hat. Die instinktiven Erkenntnisse hätten bei diesem Vorgang eine besondere Stellung, da sie eine besondere epistemische Qualität besäßen:

Es ist schon besprochen worden, daß die instinktiven Erkenntnisse ein ganz besonderes Vertrauen genießen. Wir wissen nicht mehr, wie wir sie erworben haben, und können daher an der Art der Erwerbung nichts mehr be-

---

<sup>28</sup> Vgl. Mach [193], 97: „Archimedes *beweist* in der angedeuteten Art sein Hebelgesetz, Stevin sein Gesetz des schiefen Druckes, Daniel Bernoulli das Kräfteparallelogramm, Lagrange das Prinzip der virtuellen Verschiebungen. [...] Wir können heute das Hebelprinzip, die statischen Momente, das Prinzip der schiefen Ebene, das Prinzip der virtuellen Verschiebung, das Kräfteparallelogramm als durch *gleichwertige* Beobachtung gefunden ansehen.“ In seiner späteren Monographie ‚Über Gedankenexperimente‘ ebnet Mach den in der *Mechanik* betonten Unterschied zwischen Galileis Argumentation zum Hebelgesetz und Stevins Argumentation zum Prinzip der schiefen Ebene endgültig ein: „Umstände, die man in Bezug auf einen gewissen Erfolg als einflußlos erkannt hat, kann man in Gedanken beliebig variieren, ohne diesen Erfolg zu ändern. Man gelangt aber durch geschickte Handhabung dieses Verfahrens zu Fällen, welche auf den ersten Blick von dem Ausgangsfall wesentlich verschieden scheinen, also zur Verallgemeinerung der Auffassung. Stevin und Galilei üben dieses Verfahren meisterhaft bei ihrer Behandlung des Hebels und der schiefen Ebene.“ Mach [191], 185-186.

mängeln. Wir haben nichts zu ihrer Entstehung beigetragen. Sie treten uns mit einer Macht entgegen, welche dem Ergebnis einer willkürlichen reflektierten Erfahrung, bei welcher wir immer unser Eingreifen fühlen, niemals zukommt. Sie erscheinen uns als etwas von Subjektivität Freies, Fremdes, das wir aber doch stets zu Hand haben und das uns näher liegt als die einzelnen Naturtatsachen. (Mach [193], 98)

Doch auch instinktive Erkenntnis ist fehlbar und ihre Autorität reicht nur so weit, wie genaue Experimente sie nicht widerlegen können oder diese instinktiven Erkenntnisse auf neue „Erfahrungsgebiete“ ausgeweitet werden.

### 2.1.2 MACHS NATURALISMUS, EMPIRISMUS UND DAS PRINZIP DER DENKÖKONOMIE

In Anbetracht der Machschen Zielsetzung, den durch und durch empirischen Charakter der Naturwissenschaft nachzuweisen, verwundert es vielleicht, dass Mach sich überhaupt auf so detaillierte Weise mit Gedankenexperimenten und instinktiven Erkenntnissen auseinandersetzt.

Verständlich wird dies, wenn man bedenkt, dass Machs Erkenntnistheorie nicht im modernen Sinne zwischen Entdeckungs- und Begründungszusammenhang zu unterscheiden versucht. Wissenschaftstheorie stellt für den Naturalisten Mach die Frage danach, welche Gesetze die „Anpassung der Gedanken an die Tatsachen und die Anpassung der Gedanken aneinander“ bestimmen.<sup>29</sup> Aus diesem Grunde untersucht er „historisch-kritisch“ die Entwicklung der Mechanik, Wärmelehre und Optik. Dass dabei die erwähnten Pseudoableitungen ins Spiel kommen liegt zunächst einfach daran, dass sie eben Bestandteil der Wissenschaftsgeschichte sind und bei der Anpassung der Gedanken an die Tatsachen und der Anpassung der Gedanken aneinander offenbar eine Rolle gespielt haben. Machs naturalistische Wissenschaftstheorie kann solche Episoden nicht durch den Hinweis auf den Unterschied von Entdeckungs- und Rechtfertigungszusammenhang ausklammern.

Nichtsdestotrotz ist auch Mach in seiner Vorgehensweise *historisch-kritisch*, d.h., dass er nicht bloß alle in der Wissenschaftsgeschichte feststellbaren Begründungsversuche als nur zu konstatierendes Faktum behandelt, sondern im Namen des Ökonomieprinzips auch zu normativen Urteilen kommt: Archimedes oder Daniel Bernoulli machen einen *Fehler*. Machs Argumentation ist hier nicht – wie man vielleicht vermuten würde – darauf gerichtet, dass Archimedes oder Bernoulli Pseudoableitungen vortragen, die dem Empiristen Mach übel aufstoßen (denn das macht Stevin schließlich auch), sie ist darauf gerichtet, dass Archimedes und Bernoulli mit ihrem Vortrag keinen *Erfolg* haben, sie können nicht *überzeugen*.

---

<sup>29</sup> Vgl. Stadler [310], 132-167. Zur Rolle der Denkökonomie in Machs Philosophie, vgl. Goeres [125].

Archimedes' Beweis ist nicht das letzte Wort in der Geschichte der Mechanik, eine stattliche Anzahl von Nachfolgern versucht dasselbe Hebelgesetz immer wieder zu „beweisen“. Stevins „Beweis“ hingegen ist überzeugend, noch der heutige Leser „fühlt mit ihm“ die Überzeugungskraft seiner instinktiven Erkenntnis.

Das sind zunächst psychologische Tatsachen. Manchmal gelingt durch ein Verfahren die Anpassung der Gedanken an die Tatsachen und der Gedanken aneinander, manchmal eben nicht. Darin besteht Machs eigentliches Erklärungsproblem, das durch die Gesetze der Denkökonomie<sup>30</sup> gelöst werden soll. Sind diese Gesetze gefunden, so lassen sie sich normativ verwenden. Wenn wir den Grund kennen, warum Archimedes oder Bernoulli nicht überzeugen konnten, können wir jemandem, der in einem anderen Bereich dasselbe versucht, gleich seinen zukünftigen Misserfolg prognostizieren.

Diese Vorgehensweise steht im Gegensatz zu einer Auffassung von Wissenschaftstheorie, nach der zunächst *a priori* ausgetüfelt wird, nach welchem Schema sich Wissensansprüche rechtfertigen lassen, und dann die Wissenschaftsgeschichte „rational rekonstruiert“ wird, wobei alle Episoden als irrelevantes historisches Beiwerk herausfallen, die nicht zum Schema passen. Mach versucht ein möglichst konsequentes naturalistisches Programm zu verfolgen.

### 2.1.3 NEWTONS EIMERVERSUCH

Die wohl bekannteste, weil umstrittene normative Verwendung der naturalistisch gefundenen Resultate zum Thema Gedankenexperiment ist Machs Kritik an Newtons „Eimerversuch“.<sup>31</sup> Umstritten ist diese Kritik deshalb, weil Mach hier Gedankenexperimente nur deshalb als illegitim zurückzuweisen scheint, weil sie seiner Auffassung nicht entsprechen, er ansonsten Gedankenexperimente für seine Position aber in Ordnung findet.<sup>32</sup>

Wie wir bereits gesehen haben, gewinnt Mach aus seiner Analyse der Geschichte der Mechanik zunächst folgende verallgemeinerungsfähigen Erkenntnisse:

- (1) Eine Ableitung, von denen sich mindestens eine Prämisse auf eine „instinktive Erkenntnis“ bezieht, kann nicht überzeugen, wenn die instinktive Erkenntnis die Wahrheit der Konklusion voraussetzt (Beispiel Archimedes).
- (2) Eine Ableitung, von denen sich mindestens eine Prämisse auf eine „instinktive Erkenntnis“ bezieht, kann nicht überzeugen, wenn die instinktive Erkenntnis durch Beobachtung schwerer zu bestätigen ist als die Konklusion (Beispiel Bernoulli).

<sup>30</sup> Vgl. hierzu Mach [191], Mach [192], Mach [194].

<sup>31</sup> Zum „Eimerversuch“ vgl. Laymon [179].

<sup>32</sup> Vgl. hierzu Kühne [175], 149-157; denselben Vorwurf erheben Peijnenburg und Atkinson [252] gegen Einsteins Umgang mit Newtons Experiment.

Der Erfolg der instinktiven Erkenntnis wird von Mach wie gesagt *evolutionstheoretisch* erklärt. Einzelne instinktive Überzeugungen lassen sich mit Bezug auf ihre Genese kaum beurteilen, weil wir nicht mehr wissen, wie wir zu ihnen gelangt sind. Da wir diese Überzeugungen aber durch die Evolutionsgeschichte mitbekommen haben, können sie nur dann durch einen zuverlässigen Prozess entstanden sein, wenn wir davon ausgehen können, dass die Umstände, auf die sich die instinktiven Überzeugungen beziehen, in der Menschheitsgeschichte schon vorgekommen sind (und von Relevanz für den Selektionsprozess waren). Dies liefert uns noch

- (3) Eine Ableitung, die als Prämisse auf eine instinktive Erkenntnis bezieht, kann nicht überzeugen, wenn die „instinktive Erkenntnis“ sich auf Umstände bezieht, die uns *nicht* vertraut<sup>33</sup> sind.

In ‚Über Gedankenexperimente‘ ist die relative Nähe des Gedankenexperiments zu den Umständen in der Wirklichkeit ein Kennzeichen<sup>34</sup> für wissenschaftliches Gedankenexperimentieren:

Außer dem physikalischen Gedankenexperiment gibt es noch ein anderes, welches auf höherer intellektueller Stufe in ausgedehntem Maße geübt wird – das Gedankenexperiment. Der Projektentwerfer, der Erbauer von Luftschlössern, der Romanschreiber, der Dichter sozialer oder technischer Utopien experimentiert in Gedanken. Aber auch der solide Kaufmann, der ernste Erfinder oder Forscher tun dasselbe. Alle stellen sich Umstände vor, und knüpfen an diese Vorstellung die Erwartung, Vermutung gewisser Folgen; sie machen eine Gedankenerfahrung. Während aber die Ersteren in der Phantasie Umstände kombinieren, die in Wirklichkeit nicht zusammentreffen, oder diese Umstände von Folgen begleitet denken, welche nicht an dieselben gebunden sind, werden letztere, deren Vorstellungen gute Abbilder der Tatsachen sind, in ihrem Denken der Wirklichkeit sehr nahe bleiben. Auf der mehr oder weniger genauen unwillkürlichen Abbildung der Tatsachen in unseren Vorstellungen beruht ja die Möglichkeit der Gedankenexperimente. (Mach [191], 183-184)

Diese Überlegungen zur Überzeugungskraft von Gedankenexperimenten finden auch in Machs Auseinandersetzung mit Newtons Auffassung vom absoluten Raum ihre Anwendung.

In den *Principia Mathematica* argumentiert Newton (vgl. Newton [229]), dass sich relative von absoluten Bewegungen empirisch unterscheiden lassen, da Erstere keine Fliehkräfte von der Achse der Bewegung verursachen, Letztere aber schon

<sup>33</sup> In seiner Interpretation des Machschen Beitrags zum Thema Gedankenexperiment stellt Sorensen heraus, dass Mach bereits richtigerweise Vertrautheit gefordert hätte, Vertrautheit und Realismus aber zwei verschiedene Dinge seien. (3) stelle demnach eine sinnvolle Verallgemeinerung dar, es wäre aber keine, würde der letzte Halbsatz lauten: ‚die nicht *in der Wirklichkeit* vorkommen‘. Wie das nächste Zitat aus Mach [191] belegt, hat Mach eine solche Unterscheidung nicht im Sinn gehabt. Vgl. Sorensen [304], 66.

<sup>34</sup> Wie (1) und (2) stammt also auch (3) aus der Beobachtung, welche „instinktiven Erkenntnisse“ wissenschaftlich akzeptiert werden und welche bloße „Luftschlösser“ sind.

und dass dies für die Annahme eines absoluten Raumes spricht.<sup>35</sup> Ein Blick auf diese Argumentation lohnt sich, weil wir so einerseits einen genaueren Einblick erhalten, wie Mach die Methode des Gedankenexperiments angewendet sehen wollte, und andererseits dieses Beispiel in einem späteren Kapitel in einem anderen Zusammenhang erneut auftauchen wird.

Newtons Argument gehört eindeutig zu den paradigmatischen Gedankenexperimenten, obwohl es zunächst als Experimentbericht bei Newton vorkommt:

Man hänge z.B. ein Gefäß an einem sehr langen Faden auf, drehe dasselbe beständig im Kreise herum, bis der Faden durch die Drehung sehr steif wird; hierauf fülle man es mit Wasser und halte es zugleich mit dem letzteren in Ruhe. Wird es nun durch eine plötzlich wirkende Kraft in entgegengesetzte Kreisbewegung versetzt und hält diese, während der Faden sich ablöst, längere Zeit an, so wird die Oberfläche des Wassers anfangs eben sein, wie vor der Bewegung des Gefäßes, hierauf, wenn die Kraft allmählig auf das Wasser einwirkt, bewirkt das Gefäß, dass dieses (das Wasser) merklich sich umzudrehen anfängt. Es entfernt sich nach und nach von der Mitte und steigt an den Wänden des Gefäßes in die Höhe, indem es eine hohle Form annimmt. (Diesen Versuch habe ich selbst gemacht). Durch eine immer stärkere Bewegung steigt es mehr und mehr an, bis es in gleichen Zeiträumen mit dem Gefäße sich umdreht und relativ in demselben ruhet. Dieses Ansteigen deutet auf ein Bestreben, sich von der Axe der Bewegung zu entfernen, und durch einen solchen Versuch wird die wahre und absolute kreisförmige Bewegung des Wassers, welche der relativen hier ganz entgegengesetzt ist, erkannt und gemessen. Im Anfange, als die relative Bewegung des Wassers im Gefäße am grössten war, verursachte dieselbe kein Bestreben, sich von der Axe zu entfernen. Das Wasser suchte nicht, sich dem Umfange zu nähern, indem es an den Wänden emporstieg, sondern blieb eben, und die wahre kreisförmige Bewegung hatte daher noch nicht begonnen. Nachher aber, als die relative Bewegung des Wassers abnahm, deutete sein Aufsteigen an den Wänden des Gefäßes das Bestreben an, von der Axe zurückzuweichen, und dieses Bestreben zeigte die stets wachsende wahre Kreisbewegung des Wassers an, bis diese endlich am grössten wurde, wenn das Wasser selbst relativ im Gefäß ruhte. Jenes Streben hängt nicht von der Uebertragung des Wassers in Bezug auf die umgebenden Körper ab, und deshalb kann die wahre Kreisbewegung nicht durch eine solche Uebertragung erklärt werden. (Newton [229], 30)

Newton argumentiert hier offenbar an einem beobachtbaren Phänomen, weist sogar darauf hin, er habe den Versuch selbst gemacht. Worauf Newton hinaus will, ist dass die Veränderung der Oberflächenform des Wassers im Eimer sich nicht durch Bezug auf die sich relativ zum Wasser bewegenden Eimerwände erklären lässt. Was aber möglich ist, ist die Oberflächenform des Wassers in Bezug zu einem absoluten Raum zu erklären. Dieser wäre somit eine kausal wirksame, physi-

---

<sup>35</sup> Mach scheint das zumindest so aufzufassen, Einstein später ebenfalls (vgl. Einstein [99]). Ob Newton das auch so gemeint hat, steht auf einem anderen Blatt. Robert Disalle argumentiert überzeugend dafür, dass es Newton nur um den Nachweis der Messbarkeit einer theoretischen Größe (absolute Rotation) ging, ohne dass daraus irgendwelche metaphysischen Existenzannahmen folgen sollten. Vgl.: Disalle [89]. Ich danke Harry Soodak für diesen Hinweis.

kalische Größe.<sup>36</sup> Mach weist diese Argumentation zurück, indem er – laut einiger Interpreten – selbst eine „hypothetische Beobachtung“<sup>37</sup> vorbringt:

Wenn noch immer moderne Autoren durch die Newtonsche, vom Wassergefäß hergenommenen Argumente sich verleiten lassen, zwischen relativer und absoluter Bewegung zu unterscheiden, so bedenken sie nicht, daß das Weltsystem uns nur einmal gegeben, die ptolemäische oder kopernikanische Auffassung aber unsere Interpretationen, aber beide gleich wirklich sind. Man versuche, das Newtonsche Wassergefäß festzuhalten, den Fixsternhimmel dagegen zu rotieren und das Fehlen der Fliehkräfte nun nachzuweisen. (Mach [193], 252)

Es ist allerdings ein völliges Missverständnis des Machschen Arguments, wenn man meint, Mach würde hier mit einem Gedankenexperiment gegen ein reales Experiment vorgehen. Worauf Mach an dieser Stelle nur hinweist, ist, dass wir in Bezug auf den absoluten Raum nicht dazu in der Lage sind, ein *experimentum crucis* durchzuführen. Wir können eben nicht den Fixsternhimmel rotieren lassen und überprüfen, ob die Oberflächenform des Wassers sich auch dann ändert. Daher ist die Hypothese vom absoluten Raum nur dann die beste Erklärung, wie man keine empirisch äquivalente Theorie finden kann, die ökonomischer ist, weil sie eben keine neue Entität postuliert. Eine solche Theorie würde beispielsweise die Oberflächenform des Wassers mit Bezug auf die Fixsternsphäre erklären. Mach sagt also nur, dass das Newtonsche Argument nicht die Existenz eines absoluten Raumes etablieren kann, weil sich dieses Phänomen durch eine sparsamere empirisch äquivalente Theorie erklären lässt. Weder wird Newtons Argument als schlechtes Gedankenexperiment abgekanzelt, noch wird ein anderes Gedankenexperiment dagegen gestellt. An dieser Stelle geht es nicht um Gedankenexperimente. Dasselbe gilt auch für die etwas spätere Stelle:

Der Versuch Newtons mit dem rotierenden Wassergefäß lehrt nur, daß die Relativedrehung des Wassers gegen die Gefäßwände keine merklichen Zentrifugalkräfte weckt, daß dieselben aber durch die Relativedrehung gegen die Masse der Erde und die übrigen Himmelskörper geweckt werden. Niemand kann sagen, wie der Versuch qualitativ und qualitativ verlaufen würde, wenn die Gefäßwände immer dicker und massiver, zuletzt mehrere Meilen dick würden. Es liegt nur der eine Versuch vor, und wir haben denselben mit den übrigen uns bekannten Tatsachen, nicht aber mit unseren willkürlichen Dichtungen in Einklang zu bringen. (Mach [193], 256)

Auch an dieser Stelle argumentiert Mach nur, dass das Newtonsche Experiment kein eindeutiger Nachweis für einen absoluten Raum darstellen kann, weil es ja ein beobachtbares System von Massen gibt, relativ zu dem sich die Oberflächenform erklären lässt. Mach schließt dabei nicht einmal aus, dass man ein *experimentum crucis* entwickeln könnte, dies ist aber eben noch nicht passiert und auf der

<sup>36</sup> Vgl. Kühne [175], 150.

<sup>37</sup> Kühne [175], 150. Für die unzutreffende Auffassung, dass Mach hier selbst ein Gedankenexperiment vorträgt, vgl. auch Pulte [262].

Grundlage des bisher Beobachteten ist die Theorie, die auf den absoluten Raum verzichtet, vorzuziehen. Machs Argumentation scheint gerade darin zu bestehen, dass die Newtonsche Theorie selbst auf die Postulierung des absoluten Raumes verzichten kann, weil gerade ihr entscheidendes Corollar V keinen Bezug auf den absoluten Raum enthält, diese Postulierung also selbst innerhalb der Newtonschen Theorie überflüssig ist.<sup>38</sup>

Anders ist dies bei Machs Auseinandersetzung mit einem Argument von Carl Neumann (vgl. Neumann [227]). Hier findet tatsächlich eine Auseinandersetzung mit einem Gedankenexperiment statt.

Neumann argumentiert in *Die Prinzipien der Galilei-Newtonschen Theorie* ebenfalls für die Existenz eines absoluten Raumes. Sein Gedankenexperiment läuft so, dass man sich einen Rotationskörper vorstellt, dessen Rotationsbewegung durch die Wirkung der Fliehkräfte feststellbar ist: seine Pole sind abgeflacht. Nun räumt man das Universum leer, entfernt also alle anderen Körper bis auf den einen. Wäre die Bewegung des Rotationskörpers bloß relativ, sollte sich sein Zustand nun ändern. Da nichts mehr da ist, relativ zu dem er sich bewegt, ist er offenbar in Ruhe. Neumann suggeriert, dass sich der Rotationskörper in einer solchen Situation *nicht* ändern würde, seine Pole blieben abgeplattet. Folglich sei die Erklärung für die Abplattung der Pole nur in der kausalen Wirkung des absoluten Raumes zu finden.<sup>39</sup> Mach hat gegen dieses Argument zwei Einwände:

Es scheint mir kein Gewinn, wenn zur Vermeidung eines Widerspruchs eine an sich sinnlose Annahme gemacht wird. Ferner scheint mir der berühmte Mathematiker von der gewiß sehr fruchtbaren Methode des Gedankenexperiments hier einen gar zu freien Gebrauch zu machen. Man darf im Gedankenexperiment unwesentliche Umstände modifizieren, um an einem Fall neue Seiten hervortreten zu lassen. Daß aber die Welt einflußlos ist, darf nicht von vornherein angenommen werden. (Mach [193], 300)

Machs erster Vorwurf besteht in einem Hinweis auf die „unökonomische“ Schlussweise in Neumanns Argument. Wenn bestimmte Annahmen zu einem Widerspruch führen, sollte man versuchen, eine möglichst konservative Revision seines Überzeugungssystems vorzunehmen. In diesem Fall sollte man die Annahme aufgeben, dass sich der Rotationskörper in einem leeren Universum weiter-

<sup>38</sup>Mach [193], 256: „Um nun ein allgemein gültiges Bezugssystem zu haben, wagte Newton das Corollar V der Prinzipien [„Körper, welche in einen gegebenen Raum eingeschlossen sind, haben dieselbe Bewegung unter sich; dieser Raum mag ruhen oder sich gleichförmig und geradlinig, nicht aber im Kreise fortbewegen.“ Newton [229], 38]. Er denkt sich ein momentanes irdisches Koordinatensystem, für welches das Trägheitsgesetz gilt, im Raum, ohne Drehung des Fixsternhimmel, festgehalten. Ja er kann diesem System auch noch eine beliebige Anfangslage und gleichförmige Translation gegen das erwähnte momentane irdische System erteilen, ohne seine Brauchbarkeit zu verlieren. Die Kraftgesetze Newtons werden dadurch nicht alteriert; [...]. Durch diese Fassung hat Newton den Sinn seiner *hypothetischen* Erweiterung des Galileischen *Trägheitsgesetzes* *genau* angegeben, Man sieht auch, daß die Reduktion auf den absoluten Raum keineswegs nötig war, indem sich das Bezugssystem ebenso relativ bestimmt wie in jedem anderen Fall.“

<sup>39</sup>Vgl. Neumann [227]; Mach [193], 300; Kühne [175], 151.

hin drehen würde, nicht aber, dass zusätzlich zu den beobachtbaren Entitäten des Universums noch weitere unbeobachtbare kausalen Einfluss haben. Dies ist eine Kritik am Neumannschen Gedankenexperiment, die sich aber zunächst auf die Konklusion alleine bezieht, nicht auf das Gedankenexperiment als Ganzes.

Der zweite Einwand besteht dann darin, dass Neumanns Gedankenexperiment in anderer Hinsicht zu freizügig ist. Im Gedankenexperiment können Faktoren modifiziert werden, die man mit Grund für einflusslos halten kann.<sup>40</sup> Bei Neumanns Gedankenexperiment geht es aber ja nur um die Frage, ob das Universum einflusslos ist auf die Rotationsbewegung, die Variation betrifft also genau das, was man nicht mit Grund für einflusslos halten kann. Neumanns Gedankenexperiment macht also den gleichen Fehler wie die „Ableitung“ des Archimedes, sie kann auf der Grundlage von (1) kritisiert werden.<sup>41</sup>

Wie Ulrich Kühne<sup>42</sup> argumentiert, kann sie auch auf der Grundlage von (3) – dem inhaltlichen Kriterium – kritisiert werden. Ein Universum, das bis auf einen Körper leer ist, war bisher auch nicht angenähert Gegenstand der Erfahrung in der Menschheitsgeschichte. Es gibt keinen Grund anzunehmen, dass sich unsere Instinkte verlässlich zeigen, wenn wir sie zu Bereichen außerhalb der Fixsternsphäre befragen.

Mach gibt keinen Hinweis darauf, dass er diesen Vorwurf vorbringen möchte, es stimmt aber, dass seine Überlegungen zum Gedankenexperiment dazu die Ressourcen bereitstellen. Was nicht stimmt, ist, dass Mach an das Neumannsche Gedankenexperiment einen Maßstab anlegt (Szenario zu weit entfernt von den beobachtbaren Tatsachen), der für sein Gedankenexperiment (Rotation des Fixsternhimmels) nicht gilt. Wir haben gesehen, dass Machs Bemerkung zur Rotation des Fixsternhimmels kein Gedankenexperiment darstellt (in dem Sinne, dass Mach eine instinktive Erkenntnis in Bezug auf einen empirischen Sachverhalt behaupten würde), sondern nur in dem Hinweis besteht, dass zwischen zwei bestimmten rivalisierenden Theorien kein *experimentum crucis* durchgeführt werden kann.

Sofern man hier von einem „Gedankenexperiment“ im Machschen Sinne sprechen darf, handelt es sich um ein Gedankenexperiment, dessen Ausgang „schwankend und unbestimmt“ ist und das auf ein physisches Experiment „als seiner natürlichen Fortsetzung“ drängt<sup>43</sup>, nur dass Mach nicht sagen kann, wie ein entsprechendes physisches Experiment aussehen könnte.

---

<sup>40</sup> ... wie in Machs Rekonstruktion des Stevinschen Gedankenexperiments die Reibung. Man kann sich eine Reihe von Experimenten vorstellen, in denen die Reibungsfreiheit immer weiter angenähert wird.

<sup>41</sup> Vgl. hierzu auch Sorensen [305], 146-147.

<sup>42</sup> Kühne [175], 152.

<sup>43</sup> Vgl. Mach [191], 185.

#### 2.1.4 GEDANKENEXPERIMENT NACH MACH

Nach Mach sind Gedankenexperimente also in erster Linie ein Mittel, physikalisches Wissen, das durch die Evolutionsgeschichte in unseren Instinkten repräsentiert ist oder durch eigene unwillkürliche Erfahrung, die wir im Umgang mit physikalischen Gegenständen gemacht haben, in unserer Erinnerung gespeichert wurde, zur Bestätigung oder Widerlegung physikalischer Prinzipien heranzuziehen.

Die instinktiven Überzeugungen, die dabei zutage treten, erscheinen uns evident, weshalb Theorien, die in enger Anlehnung an solche Überzeugungen gebildet sind, leichter angewendet und verstanden werden. Ein Überzeugungssystem, das mit unseren instinktiven Überzeugungen kohärent ist, dessen allgemeine Prinzipien aus solchen Überzeugungen gegebenenfalls sogar abgeleitet sind, hat also gewisse instrumentelle Vorzüge.

Ihrem Gehalt nach scheinen diese instinktiven Überzeugungen auch genereller oder abstrakter zu sein als Überzeugungen, die man über bloße Beobachtungen bildet (sie sind bereits Resultate eines „logisch-ökonomischen Läuterungsprozesses“). So können sich instinktive Überzeugungen beispielsweise auf Situationen beziehen, die gar nicht beobachtbar sind (Reibungsfreiheit), oder nur in einer Folge von Einzelbeobachtungen angenähert werden können. Auch dies ist ein instrumenteller Vorzug dieser Überzeugungen.

Diese instinktiven Überzeugungen sind allerdings fallibel. Es kann sein, dass uns die Evolutionsgeschichte nicht mit wahren Überzeugungen ausgestattet hat, sondern bloß mit überlebensförderlichen Überzeugungen, oder zumindest nicht mit überlebenshinderlichen. Mach scheint aber in dieser Beziehung relativ optimistisch zu sein und unser instinktives Wissen für verlässlich zu halten, zumindest wenn die Umstände im Gedankenexperiment uns vertraut sind oder der Wirklichkeit sehr nahe bleiben.

Die Aktivierung dieser Ressource scheint im Gedankenexperiment dadurch zu erfolgen, dass man sich eine Situation (offenbar bildlich) vorstellt, in der wie in einem echten Experiment bestimmte Faktoren, deren Einfluss oder Einflusslosigkeit man überprüfen will, variiert gedacht werden. Dann wird registriert, welche Folgen man bei einer solchen Variation instinktiv erwarten würde. Falls man – durch Introspektion – zu der Auffassung gelangt, dass das „Resultat“ eines solchen Experimentablaufs in Gedanken, ohne einen intentionalen Akt, also automatisch hervorgebracht wurde, ist dies ein Indikator, dem Ergebnis Glauben zu schenken.

Gedankenexperimente spielen nach Mach bei der Findung und Etablierung von Naturprinzipien eine wichtige Rolle. Diese Rolle lässt aber mit der Zeit nach, wenn das gefundene Naturprinzip sich durch viele empirische Beobachtungen gut bewährt hat. Die empirische Erfolgsquote seit der Etablierung eines Naturprinzips ist irgendwann so hoch, dass sie überzeugender ist als die Evidenz der instinktiven Überzeugungen, auf denen es beruht.

Unter welchen Umständen Gedankenexperimente nach Mach weniger überzeugungskräftig sein können, haben wir oben bereits diskutiert. Eine positive Anleitung, wie man ein erfolgreiches Gedankenexperiment entwickelt, gibt Mach nicht.<sup>44</sup>

Neben seiner wichtigen Rolle innerhalb der Theoriendynamik dient das Gedankenexperiment aber auch der Vorbereitung realer Experimente und der Vermittlung physikalischen Wissens. Letzteres steht in engem Zusammenhang mit Machs didaktischer Auffassung, dass man Naturwissenschaften am Besten dadurch lernt, dass man nachvollzieht, welche Gründe in der Wissenschaftsgeschichte für die Akzeptanz bestimmter Prinzipien oder Gesetzmäßigkeiten ausschlaggebend waren. Außerdem spielt das Gedankenexperiment eine wichtige *heuristische* Rolle in der Mathematik.

---

<sup>44</sup> Mach stellt sich zwar diese Frage (Mach [191], 192: „Wie wird [das Gedankenexperiment] eingeleitet? Wie kann es sich zu einer mit Absicht, Bewußtsein und Verständnis gebrauchten Methode entwickeln?“), beantwortet sie aber mit der kryptischen Bemerkung, dass das Gedankenexperiment durch das „Paradoxe“ am natürlichsten zur „bleibenden Gepflogenheit“ werde.

## 2.2 ‚GEDANKENEXPERIMENT‘ BEI MACHS ZEITGENOSSEN, BEI HEMPEL UND POPPER

Machs Behandlung des Gedankenexperiments stellt zwar den Anfang der wissenschaftstheoretischen Beschäftigung mit diesem Thema dar, allerdings dauerte es eine ganze Weile bis dieser Faden in der Wissenschaftstheorie wieder aufgenommen wurde. Die Ursache dafür liegt einerseits in der unmittelbaren Kritik, die Machs Auffassung durch seine Zeitgenossen erfuhr, zum anderen daran, dass für die Wissenschaftstheorie, die nach Mach durch den „Wiener Kreis“ eigentlich erst entstand, zunächst andere Fragen auf der Tagesordnung standen.

### 2.2.1 DIE NEGATIVE REZENSION MACHS

Unmittelbare Kritik erfuhr die Machsche Behandlung des Themas ‚Gedankenexperiment‘ durch Pierre Duhem (1861-1921). Duhem ist zunächst wie Mach der Auffassung, dass sich die Entwicklung der Wissenschaft nicht allein in Begriffen logisch begründeter Induktion beschreiben lässt, sondern psychologische, phylogenetische, historische Faktoren hinzugezogen werden müssen.<sup>45</sup> Im Gegensatz zu Mach ist Duhem aber nicht der Meinung, dass die nicht-induktiven Teile der Wissenschaftsgeschichte für die Begründung, Erklärung oder Vermittlung der wissenschaftlichen Erkenntnisse eine Rolle spielen. Physikalische Prinzipien sind zunächst bloße Setzungen, die nicht durch „instinktive Erkenntnis“ begründet werden, sondern schlicht dadurch, dass sie in ein theoretisches System passen. Erst der weitere Verlauf der Wissenschaftsgeschichte, in dem sich das „Postulat“ in der empirischen Anwendung bewährt, verleiht dieser Setzung einen Sinn.

In fingierten Experimenten (seien sie „nicht realisiert“, „unrealisierbar“ oder „absurd“) eine irgendwie geartete Begründung für ein Postulat zu sehen und so etwas gegebenenfalls im Physikunterricht zu vermitteln, ist schlichtweg Betrug am Studenten.<sup>46</sup>

Unter den Mängeln, die eine [der induktiven Methode folgende Darstellung der Wissenschaftsgeschichte] kennzeichnen, ist der häufigste und gleichzeitig wegen der falschen Ideen, die er dem Verstand der Schüler einpflanzt, auch der schwerste, das fingierte Experiment. Der Physiker, der genötigt ist, sich auf ein Prinzip zu berufen, das in Wirklichkeit keineswegs aus den Tatsachen abgeleitet wurde, das keineswegs durch Induktion entstand, und dem es dabei widerstrebt, dieses Prinzip für das auszugeben, was es ist, d.h. für ein Postulat, ersinnt ein Experiment, das, wenn es ausgeführt und gelungen wäre, zu dem Prinzip führen könnte, dessen Berechtigung

---

<sup>45</sup> Vgl. Kühne [175], 161.

<sup>46</sup> Vgl. Rehder [275]; Kühne [175], 158-167.

dargetan werden soll. [...] Man rechtfertigt somit ein Prinzip nicht mit Hilfe der beobachteten Tatsachen, sondern mit Hilfe solcher, deren Realisierung man vorausagt. Und diese Voraussagung hat kein anderes Fundament, als den Glauben an das Prinzip, zu dessen Stütze man sich auf eben dieses Experiment beruft. Ein derartiges Beweisverfahren führt zu einem *Circulus vitiosus*, und derjenige, der es vorbringt, ohne hervorzuheben, daß das angegebene Experiment nicht ausgeführt wurde, begeht eine Unredlichkeit. (Duhem [94], 269)

Duhems Abkanzelung des Themas ‚Gedankenexperiment‘ weist schon in die Richtung, die die Wissenschaftstheorie von da ab zunächst verfolgen wird: rationale Rekonstruktion des Forschungsprozesses unter Absehung vom Entdeckungszusammenhang.

Alexis Meinong war der Nächste, der der Machschen Konzeption vom Gedankenexperiment einen schweren Schlag versetzen sollte. In seinem Buch *Über die Stellung der Gegenstandstheorie im System der Wissenschaften* (Meinong [207]) widmet Meinong sich unter anderem der Frage, ob mathematische Sätze *a priori* gültig sind oder von der Erfahrung abhängen. In diesem Zusammenhang stößt er auf Machs Begriff vom Gedankenexperiment. Vielleicht verbirgt sich hier ein Ersatz für Erfahrung und Experiment in der Mathematik? Dies ist zunächst keine völlige Fehlinterpretation Machs, hatte dieser doch in ‚Über Gedankenexperimente‘ selbst betont, dass das Gedankenexperiment in der Mathematik entstanden sei und von da aus erst in die Naturwissenschaften gelangte.<sup>47</sup>

Meinong wendet sich also dem Begriff ‚Gedankenexperiment‘ zu und kommt auf dem Wege der Begriffsanalyse zur Auffassung, dass ‚Gedankenexperiment‘ nichts anderes sinnvoll bezeichnen kann als ‚Gedanke an ein Experiment‘, was die Methode des Gedankenexperimentierens auf die Trivialität reduziert, dass man beim empirischen Forschen ab und zu nachdenken muss:

In der Intention also, einer Überschätzung des Experiments und seines Anteils am Wissen zu begegnen, fasse ich das Dargelegte zu der Behauptung zusammen, dass das an sich schon seiner Missverständlichkeit wegen bedenkliche Wort „Gedankenexperiment“ in dem einzig klaren Sinne, der ihm gegeben werden kann, nämlich als „gedachtes Experiment“, richtiger „Gedanke an ein Experiment“ doch höchstens dazu dienen kann, die Selbstverständlichkeit auszusprechen, dass man nichts erforschen, nichts in der Forschung verwerken kann, ohne daran zu denken, dass aber diesem „Denken“ ohne weitere Bestimmung, das also sowohl Annehmen als Urteilen, und im Falle des Urteilens ebensowohl evidenzloses als evidentens Urteilen sein kann, noch keinerlei charakteristische Erkenntnisbedeutung, am wenigsten eine von der Art des wirklichen Experimentes zukommt. (Meinong [207], 67-77)

In seiner Rezension des Meinongschen Buchs kommt Russell zu dem Urteil, dass dies ein „utterly destructive criticism of Mach’s ‚Gedankenexperiment‘“ darstellt.<sup>48</sup>

<sup>47</sup> Derselben Auffassung ist Mach aber auch in Bezug auf das physische Experiment. Vgl. Mach [191], 196.

<sup>48</sup> Russell [284]. Vgl. zu Meinong auch Rehder [275].

## 2.2.2 DIE AUSBLENDUNG DES ENTDECKUNGSZUSAMMENHANGS

In den Augen seiner Zeitgenossen, insbesondere jener, die der Machschen Philosophie am nächsten standen, war die Auseinandersetzung mit dem Gedankenexperiment also ein Fehlgriff. Doch auch eine methodologische Umorientierung in der nun eigentlich erst beginnenden Wissenschaftstheorie sorgte dafür, dass es um das Gedankenexperiment erst einmal wieder still wurde. Der dabei wichtigste Aspekt ist wohl die Trennung zwischen *Entdeckungs-* und *Rechtfertigungszusammenhang*, der dazu führte, Wissenschaftstheorie als „rationale Nachrekonstruktion“ des historischen Forschungsprozesses zu betreiben, bei der nur die wirklich rechtfertigenden Überlegungen historischer Wissenschaftler von Interesse bleiben (sofern überhaupt auf die Wissenschaftsgeschichte Bezug genommen wird).

Ein frühes Beispiel hierfür ist Rudolf Carnaps *Der logische Aufbau der Welt* (Carnap [54]). In dieser Untersuchung versucht Carnap das gesamte Begriffssystem der Wissenschaft als ein Konstitutionssystem zu rekonstruieren, als dessen Basis bloße Elementarerlebnisse fungieren. Diese Nachrekonstruktion verfolgt nach gängiger Auffassung zwei Ziele, zum einen den Nachweis der Einheit der Wissenschaft, zum anderen die Rechtfertigung wissenschaftlicher Aussagen durch Nachweis ihrer prinzipiellen Rückführbarkeit auf die Erfahrung.<sup>49</sup> Die psychologischen Mechanismen, die bei der konkreten Erkenntnissynthese im wissenschaftlichen Alltag am Werk sind, werden dabei nicht beachtet:

## 100. Die Konstitution als rationale Nachrekonstruktion

Das „Gegebene“ liegt im Bewußtsein niemals als bloßes, unverarbeitetes Material vor, sondern immer schon in mehr oder weniger verwickelten Bindungen und Gestaltungen. Die Erkenntnissynthese, die Verarbeitung des Gegebenen zu Gebilden, zu Vorstellungen der Dinge, der „Wirklichkeit“, geschieht meist unabsichtlich, nicht nach bewußtem Verfahren. [...]

Auch in der Wissenschaft geschieht die Verarbeitung, Gegenstandsbildung und Erkennung meist intuitiv und nicht in der rationalen Form logischer Schlüsse. [...]

Daß diese Erkenntnissynthese, nämlich die Gegenstandsbildung und die Erkennung oder Einordnung in Arten, intuitiv geschieht, hat den Vorzug der Leichtigkeit, Schnelligkeit und Evidenz. Aber die intuitive Erkennung [...] kann nur deshalb für weitere wissenschaftliche Verarbeitung verwertet werden, weil es möglich ist, die Kennzeichen [...] auch ausdrücklich anzugeben, mit der Wahrnehmung zu vergleichen und so die Intuition rational zu rechtfertigen. (Carnap [54], 138-139)

Für Mach ist „Leichtigkeit, Schnelligkeit und Evidenz“ gerade von Interesse, da es ihm um die Prinzipien der *Denkökonomie* geht. Deshalb ist Stevins Gedankenexperiment „geistreicher“, als wenn er „die ganze Tatsache nach allen Seiten klargelegt“<sup>50</sup> hätte. Die Denkökonomie muss erklären, warum dies so erscheint und warum es erfolgreich ist. Bei Carnap hingegen würde diese Episode gerade zugunsten

<sup>49</sup> Zur Rolle der einzelnen Motive bei Carnap vgl. Mormann [213] und Cohnitz und Rossberg [75], Kapitel 5.

<sup>50</sup> Mach [193], 52.

der Galileischen Ableitung des Prinzips der schiefen Ebene aus den Hebelgesetzen ersetzt werden. In der rationalen Rekonstruktion kommen die historischen „Abkürzungen“ nicht vor, sondern nur, was sich auf dem Hintergrund einer vorher festgelegten Erkenntnistheorie als gerechtfertigt erweist. Letzteres lässt sich sehr schön an folgendem Beispiel aus der Geschichte der Wissenschaftstheorie erläutern:

In der *Logik der Forschung* (Popper [260]) geht es Karl Raimund Popper (1902-1994) unter anderem darum, die Auffassung des „Neopositivismus“, dass die empirischen Wissenschaften durch die „induktive Methode“ charakterisiert werden können, durch die Auffassung zu ersetzen, dass die Methode der Wissenschaften ausschließlich hypothetisch-*deduktiv* ist. Nachdem Popper ein *a priori* Argument vorgetragen hat, warum das Induktionsprinzip, das seiner Meinung nach der neopositivistischen Auffassung zu Grunde liegt, nicht gerechtfertigt werden kann, erklärt er das „induktionslogische Vorurteil“ als Resultat einer Vermengung von psychologischen und erkenntnistheoretischen Fragestellungen:

Wir haben die Tätigkeit des wissenschaftlichen Forschers eingangs dahin charakterisiert, daß er Theorien aufstellt und überprüft.

Die erste Hälfte dieser Tätigkeit, das Aufstellen der Theorien, scheint uns einer logischen Analyse weder fähig noch bedürftig zu sein: An der Frage, wie es vor sich geht, daß jemandem etwas Neues einfällt [...], hat wohl die empirische Psychologie Interesse, nicht aber die Erkenntnislogik. [...]

Wir wollen also scharf zwischen dem Zustandekommen des Einfalls und den Methoden seiner logischen Diskussion unterscheiden und daran festhalten, daß wir die Aufgabe der Erkenntnistheorie oder Erkenntnislogik (im Gegensatz zur Erkenntnispsychologie) derart bestimmen, daß sie lediglich die Methoden der systematischen Überprüfung zu untersuchen hat, der jeder Einfall, soll er ernst genommen werden, zu unterwerfen ist. [...]

Sofern der Forscher seinen Einfall kritisch beurteilt, abändert oder verwirft, könnte man unsere methodologische Analyse auch als rationale Nachrekonstruktion der betreffenden denkpsychologischen Vorgänge auffassen. Nicht, daß sie diese Vorgänge so beschreibt, wie sie sich tatsächlich abspielen: sie gibt nur ein logisches Gerippe des Prüfungsverfahrens. Gerade das aber dürfte man wohl unter der rationalen Nachkonstruktion eines Erkenntnisvorgangs verstehen. (Popper [260], 6-7)

Sofern Wissenschaftler induktiv vorgehen, fällt dies bei Popper in die Psychologie der Forschung und ist kein Thema der Wissenschaftstheorie. *Was* jeweils zur Heuristik und was zur Rechtfertigung gehört, findet man nicht dadurch heraus, dass man wissenschaftliche Forschungsprozesse beobachtet, sondern dadurch, dass man *a priori* untersucht, durch welche Verfahren wissenschaftliche Aussagen gerechtfertigt werden können. Ist man im Gegensatz zu Popper der Auffassung, dass induktive Schlüsse sehr wohl rechtfertigen können, wird man die Grenze zwischen Entdeckungs- und Rechtfertigungszusammenhang entsprechend anders ziehen. Ein schönes Beispiel liefert Reichenbach in seiner Verteidigung der induktiven Logik, bei der er Poppers Vorwurf einfach umdreht:

The scientist who discovers a theory is usually guided to his discovery by guesses; he cannot name a method by means of which he found the theory and can only say that it appeared plausible to him, that he had the right hunch, or that he saw intuitively which assumption would fit the facts. Some philosophers have misunderstood this psychological description of discovery as proving that there exists no logical relation leading from the facts to the theory [...]. Inductive inference is for them guesswork inaccessible to logical analysis. These philosophers do not see that the same scientist who discovered his theory through guessing presents it to others only after he sees that his guess is justified by the facts. [...] The inductive inference is employed not for finding a theory, but for justifying it in terms of observational data.

The mystical interpretation of the hypothetico-deductive method as an irrational guessing springs from a confusion of *context of discovery* and *context of justification*. The act of discovery escapes logical analysis; there are no logical rules in terms of which a "discovery machine" could be constructed that would take over the creative function of the genius. But it is not the logician's task to account for scientific discoveries; all he can do is to analyze the relation between given facts and a theory presented to him with the claim that it explains these facts. In other words, logic is concerned only with the context of justification. And the justification of a theory in terms of observational data is the subject of the theory of induction. (Reichenbach [276], 230-231)

Der Hinweis auf diese Vorgehensweise soll keine Kritik darstellen, aber verständlich machen, weshalb das Gedankenexperiment, das bei Mach noch so eine große Rolle gespielt hat, in der Wissenschaftstheorie zunächst ignoriert wurde. Für Mach war das Gedankenexperiment automatisch von Interesse, weil es an prominenter Stelle in der Wissenschaftsgeschichte vorkam. In der frühen Wissenschaftstheorie – sei es im Neopositivismus oder kritischen Rationalismus – ist es nur dann von Interesse, wenn es sich irgendwie in die bestehende Auffassung von wissenschaftlichen Rechtfertigungsprozeduren einpassen lässt. Wie wir sehen werden bleibt da nicht viel. Bestenfalls gesteht Popper ihm die Rolle als Instrument zur Aufdeckung theorieinterner Widersprüche zu.

### 2.2.3 DIE WIEDERGEURT DES THEMAS AUS DER ENTWICKLUNG DER PHYSIK

In der frühen analytischen Philosophie in der Nachfolge des Wiener Kreises findet daher fast keine Auseinandersetzung mit dem Thema ‚Gedankenexperiment‘ statt. Ausnahmen bilden Carl Gustav Hempel und (der bereits erwähnte) Karl Raimund Popper.

Dass eine Beschäftigung mit dem Gedankenexperiment wieder aufkam – trotz der Grabrede Duhems – ist dabei wohl insbesondere der Tatsache zu verdanken, dass in der Zwischenzeit die Physik eine dramatische Entwicklung vollzogen hatte, in der an exponiertesten Stellen (Relativitätstheorie, Quantenmechanik) der Eindruck erweckt wurde, man forsche hier hauptsächlich mit Gedankenexperimen-

ten.<sup>51</sup> So findet man sowohl bei Hempel wie bei Popper Warnungen vor der suggestiven Kraft von Gedankenexperimenten in eben diesen aktuell diskutierten Bereichen.

Hempel thematisiert Gedankenexperimente in der Auseinandersetzung mit dem Max Weberschen Begriff des „Idealtypus“. Ihm geht es um die Frage, ob und (falls ja) wann radikal vereinfachende Idealisierungen in den Geistes- und Sozialwissenschaften noch explanativen Wert behalten. In diesem Zusammenhang thematisiert Hempel auch Max Webers Analyse der Rolle kontrafaktischer Gedankenexperimente in den Geschichtswissenschaften, wie dieser sie etwa in ‚Objektive Möglichkeit und Adäquate Verursachung in der historischen Kausalbetrachtung‘<sup>52</sup> diskutiert hat.

Hempel trifft hierbei eine Unterscheidung zwischen zwei „extremen“ Typen von Gedankenexperimenten („experiments-in-imagination“), die sich auch in den Naturwissenschaften finden ließen: *intuitive* Gedankenexperimente und *theoretische*:

An intuitive experiment-in-imagination is aimed at anticipating the outcome of an experimental procedure which is just imagined, but which may well be capable of being actually performed. Prediction is guided here by past experience concerning particular phenomena and their regularities, and occasionally by belief in general principles which are accepted as if they were *a priori* truths. [...]

The theoretical kind of imaginary experiment on the other hand, presupposes a set of explicitly stated general principles – such as laws of nature – and it anticipates the outcome of the experiment by deductive or probabilistic inference from those principles in combination with suitable boundary conditions representing the relevant aspects of the imagined experimental situation. (Hempel 1965, 164-165)

Theoretische Gedankenexperimente sind bloße Deduktionen aus Theorien und stipulierten Anfangsbedingungen, haben also eigentlich keinen „imaginären“ Teil. Hempel ordnet Webers kontrafaktische Betrachtungen deshalb unter die intuitiven Gedankenexperimente.<sup>53</sup> Zu diesen gehören laut Hempel auch die Gedanken-

<sup>51</sup> Vgl. hierzu die Darstellung in Kühne [175], 175-251.

<sup>52</sup> Weber [337], 102-131. Da wir uns an dieser Stelle auf die Rolle von Gedankenexperimenten in den Naturwissenschaften beschränken, muss diese Diskussion leider außen vor bleiben. Vgl. aber De Mey und Weber [84], Roberts [281].

<sup>53</sup> Ob diese Einordnung berechtigt ist, erscheint zweifelhaft. In ‚Objektive Möglichkeit und Adäquate Verursachung in der historischen Kausalbetrachtung‘ macht Weber deutlich, dass auch ein historisches Gedankenexperiment seine Gültigkeit durch Ableitung aus unserem „nomologischen“ (gesetzesartige Aussagen bzw. „Erfahrungsregeln“) und „ontologischem“ (Anfangs- und Randbedingungen bzw. „Tatsachen“) Wissen erhält, also eher einem „theoretischen Gedankenexperiment“ ähnelt: „Ein „Möglichkeits“urteil in dem Sinne, in welchem der Ausdruck hier gebraucht ist, bedeutet also stets die Bezugnahme auf Erfahrungsregeln. Die Kategorie der „Möglichkeit“ kommt also nicht in ihrer negativen Gestalt zur Verwendung, in dem Sinne also, daß sie ein Ausdruck unseres Nicht- resp. Nichtvollständig-Wissens im Gegensatz zum assertorischen oder apodiktischen Urteil ist, sondern gerade umgekehrt bedeutet sie hier die Bezugnahme auf ein positives Wissen von „Regeln des Geschehens“, auf unser „nomologisches“ Wissen, wie man zu sagen pflegt.“ (Weber [337], 113-114).

experimente Galileis, die er als fruchtbare Heuristiken betrachtet, naturwissenschaftliche Hypothesen zu entwerfen, denen aber keine Rolle bei der Bestätigung dieser Hypothesen zukommt.

Mit dem Hinweis, dass Gedankenexperimente also in den Entdeckungszusammenhang wissenschaftlicher Hypothesen gehören, ist zugleich auch klar, dass Hempel zu einer etwaigen „Logik des Gedankenexperiments“ nichts weiter zu sagen hat:

Galileo's dialogues contain excellent examples of this procedure, which show how fruitful the method can be in suggesting general theoretical insights. But, of course, intuitive experiments-in-imagination are no substitute for the collection of empirical data by actual experimental or observational procedures. This is well illustrated by the numerous, intuitively quite plausible, imaginary experiments which have been adduced in an effort to refute the special theory of relativity; and as for imaginary experimentation in the social science, its outcome is liable to be affected by preconceived ideas, stereotypes, and other disturbing factors. (Hempel [143], 165)

Während Hempel hier die jüngsten Gedankenexperimente gegen die Relativitätstheorie zum Anlass nimmt, sich kritisch zum Gedankenexperiment zu äußern, ist Poppers primäre Motivation, sich der Methode des Gedankenexperiments zuzuwenden, zwar ebenfalls, eine Auffassung Einsteins vor angeblichen *a priori* Einsichten in Schutz zu nehmen, allerdings in Bezug auf Einsteins Kritik an der (Bohrschen Interpretation der) Quantenmechanik. Entsprechend ist Poppers Beitrag, der 1959 als Anhang XI. der *Logik der Forschung* hinzugefügt wurde, auch betitelt: „Über den Gebrauch und Mißbrauch von Gedankenexperimenten besonders in der Quantentheorie“.<sup>54</sup>

Hauptsächlich geht es Popper – trotz seiner Beteuerung des Gegenteils – „darum, die eigene Interpretation der Quantentheorie zu verteidigen gegen andere Interpretationen, die von ihren Urhebern kein Stück weit aufgegeben wurden.“<sup>55</sup> Sowohl Poppers eigene Interpretation der Quantenmechanik, wie auch seine inhaltliche Kritik an alternativen Interpretationen braucht uns hier nicht zu interessieren. Wir werden diese nur insofern heranziehen, wie es der Erläuterung seiner methodologischen Vorstellungen dient.

Zunächst ist bemerkenswert, dass Popper Gedankenexperimente nicht nur als fruchtbare Heuristik sieht, sondern ihnen auch eine sinnvolle methodologische Rolle zuerkennt, sowohl was die Kritik an Theorien, wie auch was deren argumentative Stützung betrifft. Popper unterscheidet daher den *heuristischen*, den *kritischen* und den *apologetischen Gebrauch* von Gedankenexperimenten.

Der heuristische Gebrauch scheint weitgehend unproblematisch. Poppers Beispiele sind die Vorstellung, dass wir eine Substanz immer weiter in kleinere Stücke teilen, bis wir bei unteilbaren kleinsten Stücken angekommen sind (heuristische Basis des Atomismus), sowie Einsteins Experiment mit dem beschleunigten Auf-

---

<sup>54</sup> Vgl. zum Folgenden auch Krimsky [170].

<sup>55</sup> Kühne [175], 255.

zug (Heuristik zur lokalen Äquivalenz von Beschleunigung und Schwerkraft). Diesen Gebrauch nennt Popper „wichtig und berechtigt“.

Als Beispiel zum kritischen Gebrauch führt Popper Galileis Gedankenexperiment zum freien Fall verschieden schwerer Körper an. Popper stellt das Gedankenexperiment so dar, als würde Galilei die Annahme des Aristotelikers, dass die natürliche Geschwindigkeit eines schweren Körpers größer ist als die eines leichten, einfach *ad absurdum* führen und somit einen Widerspruch innerhalb der aristotelischen Theorie nachweisen (inwieweit dies eine zutreffende Darstellung ist, soll uns im nächsten Teil beschäftigen).

Was Popper genau als Beispiel für ein apologetisches Gedankenexperiment betrachten will, bleibt zunächst im Dunkeln. Er weist nur darauf hin, dass dieser Gebrauch historisch auf „die Diskussion des Verhaltens von Maßstäben und Uhren im Rahmen der speziellen Relativitätstheorie“ zurückgeht, der wohl zunächst heuristisch war, dann kritisch und schließlich apologetisch wurde. Auch im Dunkeln bleibt, warum dieser Gebrauch so besonders kritisch zu beurteilen ist, schließlich scheint eine „Apologie“ doch nichts als eine Verteidigung zu sein, und was sollte eine berechnete Verteidigung verbieten können? Das räumt Popper auch selbst ein<sup>56</sup>:

Nun ist der Gebrauch des Gedankenexperiments als kritisches Argument zweifellos berechtigt: man versucht damit zu zeigen, daß der Urheber einer Theorie bestimmte Möglichkeiten übersehen hat. Offenbar hat dann auch der Gegner das Recht, solchen kritischen Einwänden entgegenzutreten, etwa indem er nachweist, daß das vorgeschlagene Gedankenexperiment prinzipiell unmöglich ist und daß zumindest in dieser Hinsicht keine Möglichkeit übersehen wurde. (Popper [260], 398-399)

Obwohl Popper den Fehlschlag bei apologetischen Gedankenexperimenten besonders betont, gelingt es ihm nicht, einen prinzipiellen methodologischen Unterschied zwischen kritischen und apologetischen Gedankenexperimenten zu ziehen. Dass er einen solchen überhaupt ziehen möchte, ist durch seine Auffassung motiviert, dass es kein wissenschaftliches Verfahren zur positiven Bestätigung wissenschaftlicher Hypothesen gibt und dass die Verteidigung wissenschaftlicher Hypothesen gegen Kritik immer in Verdacht steht, gegen die Spielregeln der Wissenschaft zu verstoßen, da man Gefahr läuft seine Theorie zu „immunisieren“ und damit ihres empirischen Gehalts zu entleeren. Die Asymmetrie zwischen Bestätigung und Widerlegung, die Popper ansonsten zur Begründung seiner Bevorzugung kritischer Verfahren heranzieht, ist im Zusammenhang mit dem Gedankenexperiment – wie Popper es darstellt – allerdings nicht offensichtlich.

Laut Popper geht es beim Gedankenexperiment in kritischer Absicht ja darum, einen Widerspruch in einer Theorie aufzuzeigen. Dass *dabei* Konzessionen an die gegnerische Theorie gemacht werden müssen, *ist* offensichtlich. Konsistenz ist kei-

---

<sup>56</sup> Popper behauptet also keinesfalls, dass der apologetische Gebrauch von Gedankenexperimenten „generell unzulässig sei“, vgl. Kühne [175], 257. Derselbe Irrtum findet sich in Pulte [262].

ne monotone Eigenschaft einer Prämissenmenge. Fügt man neue Prämissen hinzu, kann eine Prämissenmenge diese Eigenschaft verlieren. Wenn der Nachweis von Widersprüchen in einer Theorie nicht trivial sein soll, muss es irgendwelche Beschränkungen geben, welche Annahmen im Gedankenexperiment gemacht werden dürfen und welche nicht. Wenn in einer „apologetischen“ Replik auf ein kritisches Gedankenexperiment dann beispielsweise argumentiert würde, dass das Gedankenexperiment Annahmen macht, die mit der angegriffenen Theorie ohnehin unvereinbar sind, stellt dies sicher *keine* Konzession an den „Gegner“ dar, es ist aber auch nicht klar, warum dies nicht gestattet sein sollte. Hier scheint eher die umgekehrte Asymmetrie vorzuliegen.

Gehen wir angesichts dieser offenen Fragen einfach davon aus, dass Popper apologetische Gedankenexperimente schlicht mengenmäßig häufiger unangenehm aufgefallen sind. Seine methodologischen Vorschläge zum Gedankenexperiment sind ohnehin für den apologetischen wie den kritischen Gebrauch formuliert:

Allgemein ist meines Erachtens der argumentative Gebrauch von Gedankenexperimenten nur dann berechtigt, wenn die Ansichten des Gegners klar ausgesprochen sind und wenn die Regel befolgt wird, daß *die eingeführten Idealisierungen Konzessionen an den Gegner oder wenigstens für ihn annehmbar sein müssen*. [...] Im Rahmen kritischer Argumentation ist jede Idealisierung unzulässig, die diese Regel verletzt. (Popper [260], 399)

Bei Popper findet sich kaum ein Argument, warum diese methodologische Forderung eigentlich sinnvoll sein soll. Sein einziger Hinweis, der für die Befolgung dieser Regeln spricht, folgt einer Darstellung der Reaktion Bohrs auf das Gedankenexperiment von Einstein, Podolsky und Rosen:

Wie man sieht, ist bei solchen Gedankenexperimenten die Gefahr sehr groß, daß jeder die Analyse nur so weit vortreibt, als es seiner These hilft, aber nicht weiter – eine Gefahr, die sich nur dann vermeiden läßt, wenn die oben angegebenen Regeln strikt befolgt werden. (Popper [260], 402)

Wenn man Popper zugestehen wollte, dass Bohr sich tatsächlich einer „Regelverletzung“ schuldig gemacht hat (was Popper ihm nicht eindeutig nachweist), kann man sich natürlich immer noch fragen, worin eigentlich das so überaus „gefährliche“ bestehen soll, wenn man eine Analyse nur so weit treibt, wie es der eigenen These hilft. Was aber an Poppers methodologischen Bemerkungen zum Gedankenexperiment besonders problematisch erscheint, ist die Tatsache, dass überhaupt nicht klar ist, *wann* man sich der Regelverletzung schuldig gemacht hat. Als Illustration hierzu betrachten wir den Austausch zwischen Bohr und Einstein über ein anderes Gedankenexperiment Einsteins auf der Sechsten Solvay-Konferenz (in der Interpretation Poppers).

Im Jahr 1930 trafen Bohr und Einstein anlässlich dieser Konferenz in Brüssel zusammen. Schon bei vorherigen Gelegenheiten hatte Einstein versucht, mit Gedankenexperimenten gegen die Heisenberg-Relationen zu argumentieren. Er versuchte insbesondere nachzuweisen, dass bestimmte im Experiment herbeiführbare

Umstände es gestatten, genauere Vorhersagen in Einzelfällen zu machen, als die Heisenberg-Relationen dies zu erlauben scheinen. Ein solcher Nachweis würde für die statistische Natur dieser Relationen sprechen und damit für die Unvollständigkeit der Quantenmechanik.<sup>57</sup> Bohrs Darstellung des Einsteinschen Gedankenexperiments ist in Abbildung 2.3-1 wiedergegeben.

Einstein bedient sich bei diesem Versuch, die Heisenberg-Relationen durch einen Trick zu umgehen, eines Gedankens aus der Relativitätstheorie, der Äquivalenz von Masse und Energie, die prominenterweise durch  $E = mc^2$  ausgedrückt wird. Kann man den im Gedankenexperiment beschriebenen Kasten vor und nach dem Austritt des Photons wiegen, kann man daher mit seiner Masse auch seine Energie bestimmen. Auf der anderen Seite erlaubt uns die im Kasten befindliche Uhr, die den Schieber steuert, genau den Austrittszeitpunkt des Photons zu bestimmen. Eine beliebig genaue Messung der Energie eines Photons zu einem genau bestimmten Zeitpunkt gehört aber zu den Dingen, die durch die Heisenberg-Relationen ausgeschlossen werden.<sup>58</sup>

Bohr erwiderte (angeblich sofort am nächsten Morgen), dass dieses Gedankenexperiment keine physikalische Möglichkeit darstellt: es ist selbst nicht vereinbar

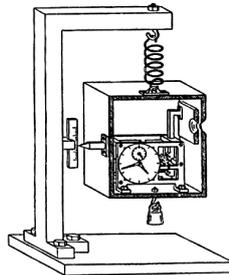


Abbildung 2.2-1

„[Die Apparatur] besteht aus einem Kasten mit einem Loch auf einer Seite, das durch einen Schieber geöffnet oder geschlossen werden kann, der mit Hilfe eines Uhrwerks im Inneren des Kastens bewegt wird. Wenn der Kasten am Anfang Strahlung enthält, und die Uhr so eingestellt ist, daß sich der Schieber zu einer gegebenen Zeit während eines sehr kurzen Intervalls öffnet, könnte man es erreichen, daß ein einzelnes Photon durch das Loch in einem Augenblick gelassen wird, der mit jeder gewünschten Genauigkeit bekannt ist. Weiterhin wäre es offenbar auch möglich, durch Wägen des Kastens vor und nach diesem Vorgang die Energie des Photons mit jeder gewünschten Genauigkeit zu messen – in striktem Widerspruch zur reziproken Unbestimmtheit von Zeit- und Energiegrößen in der Quantenmechanik.“ (aus Bohr [31], zitiert nach Held [142], 86; die Abbildung stammt aus Bohr [30], 227)

<sup>57</sup> Zu den Details vgl. Held [142], 84.

mit der allgemeinen Relativitätstheorie. Zunächst gibt Bohr eine „pseudo-realistische“ Darstellung des ursprünglich nur schematisch beschriebenen Experimentaufbaus nach Einstein (vgl. Abbildung 2.3-1). Der Kasten mit der Uhr ist selbst als Messobjekt aufzufassen, der uns indirekt über die Energie des Photons aufklären soll. Angenommen die Masse des Kastens wird durch eine Federwaage gemessen, so verändert sich durch Gewichtszu- oder -abnahme auch die Position des Kastens im Gravitationsfeld der Erde. Nach der allgemeinen Relativitätstheorie verursacht dies aber eine Enteichung der Uhr, die sich im Kasten ebenfalls durch das Gravitationsfeld der Erde bewegt. Bohr berechnet dann, dass diese Enteichung der Uhr dazu führt, dass die Heisenberg-Relationen nicht umgangen werden können.

Ob dies in letzter Konsequenz das Einsteinsche Argument mit Bezug auf die Allgemeine Relativitätstheorie aushebelt, sei dahingestellt.<sup>59</sup> Uns interessiert nur Poppers Zurückweisung des Bohrschen Arguments, bei dem er offensichtlich weder darauf abzielt, dass das Verhältnis von Quantenmechanik und allgemeiner Relativitätstheorie ungeklärt ist, noch darauf, dass die allgemeine Relativitätstheorie eine genaue Zeitmessung eventuell gar nicht ausschließt. Popper moniert nur, dass Bohr zur Rettung seiner Theorie eine davon scheinbar unabhängige einfach hinzuzieht:

Um ein kritisches Gedankenexperiment Einsteins zu entkräften, das auf dessen berühmter Formel  $E = mc^2$  beruhte, verwendet Bohr Argumente aus Einsteins Gravitationstheorie (d.h. aus der allgemeinen Relativitätstheorie). Aber  $E = mc^2$  ist aus der speziellen Relativitätstheorie und sogar aus nichtrelativistischen Gedankengängen ableitbar. Jedenfalls nehmen wir dadurch, daß wir  $E = mc^2$  annehmen, keinesfalls auch die Gültigkeit von Einsteins Gravitationstheorie an. Wenn wir daher, wie Bohr behauptet, bestimmte charakteristische Formeln der Einsteinschen Gravitationstheorie akzeptieren müßten, um die Widerspruchsfreiheit der Quantentheorie (in Verbindung mit  $E = mc^2$ ) zu retten, dann käme dies der merkwürdigen Behauptung gleich, daß die Quantentheorie der Gravitationstheorie Newtons widerspricht und ferner der noch merkwürdigeren Behauptung, daß die Gültigkeit von Einsteins Gravitationstheorie (oder zumindest der verwendeten charakteristischen Formeln, die zur Theorie des Gravitationsfeldes gehören) aus der Quantentheorie ableitbar ist. (Popper [260], 402)

Hier liegt angeblich ein Regelverstoß vor. Aber worin besteht er? Einstein hatte ein kritisches Gedankenexperiment vorgebracht und dabei eine Annahme gemacht, die für den Gegner offenbar annehmbar erschien:  $E = mc^2$ , was aus der speziellen

---

<sup>58</sup> Vgl. Held [142], 86.

<sup>59</sup> Kühne führt Gründe an, die dagegen sprechen: „Dieses Gegenargument ist problematisch. Nicht bloß deshalb, weil hier Bohr mit der allgemeinen Relativitätstheorie eine Prämisse heranzieht, deren allgemeine Verträglichkeit mit der Quantentheorie im Dunkeln liegt. Das offensichtliche Problem liegt darin, daß die (vorliegende) allgemeine Relativitätstheorie eine deterministische Theorie ist; die durch sie vorhergesagte Verlangsamung oder Beschleunigung der Uhr ist also prinzipiell exakt berechenbar und kann keinesfalls als Enteichung oder Unsicherheit der Uhrgenauigkeit interpretiert werden.“ (Kühne [175], 235).

Relativitätstheorie bereits folgt. Wir wissen ebenfalls, dass Einstein die spezielle Relativitätstheorie für „erkenntnistheoretisch mangelbehaftet“ hielt und glaubte, dass dieser Mangel durch die allgemeine Relativitätstheorie behoben ist.<sup>60</sup> Aber unabhängig davon, ob Einstein  $E = mc^2$  in diesem Beispiel aus der speziellen oder allgemeinen Relativitätstheorie abgeleitet haben wollte, die allgemeine Relativitätstheorie war Einsteins *eigene* Theorie und damit *unter Garantie* „für ihn annehmbar“, wie es Poppers methodologische Regel fordert. Wenn Bohr also einen Regelverstoß begeht, indem er eine Theorie des „Gegners“ heranzieht, um dessen „kritisches“ Gedankenexperiment auszuschalten, dann reduziert sich Poppers methodologischer Vorschlag offenbar darauf, dass nur solche Gedankenexperimente gestattet sind, die Poppers persönlichen Vorlieben in Bezug auf die Interpretation der Quantenmechanik entsprechen.

Sehen wir ab von Poppers eigenem unglücklichen Umgang mit seiner methodologischen Regel, scheint sie aber einen durchaus interessanten Kern zu haben: wenn Gedankenexperimente zu Überzeugungsänderungen beim „Gegner“ führen sollen, sollte man schon daher vermuten, dass sie nur dann erfolgreich sein können, wenn ihre Hintergrundannahmen für den Gegner annehmbar erscheinen. Was aber, wenn die erfolgreichsten Gedankenexperimente sich gerade dadurch auszeichnen, dass in ihnen Umstände vorkommen, die nicht nur keine eindeutigen Konzessionen darstellen, sondern die der bisherigen Vorstellungswelt des „Gegners“ gar nicht angehört haben? Eine solche Auffassung werden wir im nächsten Kapitel kennen lernen.

---

<sup>60</sup> Vgl. Einstein [99].

### 2.3 ‚GEDANKENEXPERIMENT‘ BEI KUHN

Betrachtet man die Entwicklung der Stellung der Wissenschaftstheorie im letzten Jahrhundert innerhalb der Philosophie, so kann man einen deutlichen Wandel beobachten.<sup>61</sup> Im Neopositivismus gehörte Wissenschaftstheorie, Philosophie der Mathematik und Logik zu den Kernbereichen der Philosophie. Wissenschaftstheorie war in gewisser Weise nur eine neue Form, Erkenntnistheorie zu betreiben. Wissenschaftliches Wissen ist das Paradebeispiel *a posteriori* gewonnenen, gut gerechtfertigten Wissens, weshalb eine Analyse der Rechtfertigungsprozesse in den Wissenschaften uns Einsicht in Erkenntnisprozesse im Allgemeinen verspricht:

Denn man könnte sagen, daß die wissenschaftliche Erkenntnis es uns ermöglicht, die Alltagserkenntnis gewissermaßen unter dem Vergrößerungsglas zu studieren, so daß, wenn wir die wissenschaftliche Erkenntnis betrachten, wir so etwas wie ein vergrößertes Bild der Alltagserkenntnis vor uns sehen. (Popper [260], XXII)

Ebenso behandelte die Sprachphilosophie die *Wissenschaftssprache* als Paradigma, zumindest im „idealsprachlichen“ Teil der analytischen Philosophie. Besonders deutlich ist dies von Carnap ausgedrückt worden:

Die Methode der Syntax [der Wissenschaftssprache], die im folgenden entwickelt werden soll, wird nicht nur der logischen Analyse wissenschaftlicher Theorien dienen, sondern auch der logischen Analyse der Wortsprachen. Wir werden zwar hier [...] symbolische Sprachen behandeln. Aber die syntaktischen Begriffe und Regeln können dann – nicht im einzelnen, aber ihrem allgemeinen Charakter nach – auch auf die Analyse der ungeheuer komplexen Wortsprachen übertragen werden. Das bisher übliche Vorgehen der direkten Analyse der Wortsprachen mußte ebenso scheitern, wie ein Physiker scheitern würde, wenn er von vornherein seine Gesetze auf die vorgefundenen Dinge, Steine, Bäume usw. beziehen wollte. (Carnap [51], 8)

Betrachtet man aber die Stellung der Wissenschaftstheorie im letzten Drittel des zwanzigsten Jahrhunderts, bietet sich ein anderes Bild.

For example, while the main preoccupation of orthodox epistemologists was the search for the proper definition of (personal) knowledge, philosophers of science ignored this altogether, being more concerned with the nature of change in theory preference among groups of scientists. More generally, philosophers of science now took it for granted that their problems and insights would come not from other parts of philosophy but rather from the history of science. (Bird [23], vii)

Diese Neuorientierung ist neben Philosophen wie Paul Feyerabend und Imre Lakatos insbesondere dem Wissenschaftshistoriker Thomas S. Kuhn zu verdanken.

---

<sup>61</sup> Vgl. hierzu Bird [23].

In seinem Hauptwerk *Die Struktur wissenschaftlicher Revolutionen*, das eher einen Beitrag zur Theorie der Wissenschaftsgeschichte darstellt als eine wissenschaftstheoretische Untersuchung<sup>62</sup>, argumentiert Kuhn, dass im Verlauf der Wissenschaftsgeschichte ein bestimmtes Muster feststellbar ist, das sich am Besten durch Bezug auf bestimmte institutionelle Strukturen der Wissenschaft erklären lässt, insbesondere durch Bezug auf die Art und Weise wie Wissenschaftler in ihrer Forschung von einem bestimmten Konsens ausgehen, den Kuhn als ‚Paradigma‘ bezeichnet.

Das Muster, das Kuhn in der Wissenschaftsgeschichte beobachtet, besteht darin, dass es (in bestimmten Wissenschaftsbereichen, die einen gewissen Grad der Reife erreicht haben) lange Phasen so genannter „normaler Wissenschaft“ gibt, in denen die Forschungstätigkeit einer Wissenschaft darauf beschränkt ist, eine akzeptierte Kerntheorie auszuarbeiten (indem ihre empirischen Konsequenzen abgeleitet und überprüft werden, sie auf neue Gebiete erweitert wird, gegebenenfalls durch sie postulierte Naturkonstanten gemessen werden, etc.), Phasen der „Krise“, in denen Schwierigkeiten, die sich bei der Durchführung normalwissenschaftlicher Aktivitäten mit der Zeit aufkumuliert haben, Zweifel an der Kerntheorie aufkommen lassen und zur Entwicklung von Alternativtheorien führen, und schließlich Phasen „wissenschaftlicher Revolution“, in denen die alte Kerntheorie durch eine neue ersetzt wird.

Was genau Forscher in Phasen der Revolution dazu führt, die bisher nahezu als Dogma betrachtete Kerntheorie durch eine neue zu ersetzen, also das „Paradigma“, das die bisherige Kerntheorie in ihrer Position gehalten hat, fallen zu lassen, hat Kuhn nicht immer besonders klar artikuliert. Es gibt daher verschieden starke Interpretationen der Kuhnschen Auffassung wissenschaftlicher Revolutionen:

Mit einem Wechsel des Paradigmas [in Phasen wissenschaftlicher Revolution] sind jedoch, gemäß der „radikalen“ Lesart Kuhns, alle gemeinsamen Rationalitätsstandards weggefallen, alle bisherigen Erfahrungsdaten werden neu interpretiert – die beiden Paradigmen sind, gemäß Kuhns bekannter Inkommensurabilitätsthese, rational unvergleichbar, inkommensurabel. Der Kampf um die Vorherrschaft findet im wesentlichen in der Form eines wissenschaftspolitischen Machtkampfes statt, in dem die Anhänger des alten Paradigmas schließlich aussterben, wodurch sich das neue Paradigma durchsetzt und eine neue normalwissenschaftliche Phase einläutet. (Schurz [289], 3)

Diese radikale, „irrationalistische“ Lesart hat zwar außerhalb der Wissenschaftstheorie die größere Popularität erlangt, Kuhns Einfluss auf die Entwicklung der Wissenschaftstheorie selbst liegt aber eher darin, die bis dahin vorherrschenden Konzeptionen rationalen wissenschaftlichen Forschens einer Revision zu unterwerfen und bisher übersehenen wissenschaftlichen Verfahren eine Rolle in der rationalen Rekonstruktion des Forschungsprozesses zukommen zu lassen. Deutlich

---

<sup>62</sup> Vgl. Bird [23].

ist dies von Wolfgang Stegmüller formuliert worden, der Kuhns Thesen entsprechend weniger radikal interpretiert:

Es scheint mir, daß Kuhns Analysen und historische Illustrationen die bisher stärkste Herausforderung der heutigen Wissenschaftstheorie darstellen. Ich empfinde diese Herausforderung als besonders groß, da ich davon überzeugt bin, daß Kuhn in wesentlichen Punkten im Recht ist. Ich glaube daher auch nicht, daß man mit seiner Herausforderung durch ‚rationalistische Polemik‘ fertig wird, wie diese von den Popperianern vorexerziert wird. Vielmehr gilt es, eine rationale Rekonstruktion von Kuhns Begriffen *der normalen Wissenschaft* sowie des Phänomens *der unmittelbaren Theorienverdrängung durch eine Ersatztheorie* zu liefern. Dies ist tatsächlich möglich. Allerdings muß man dazu bereit sein, *gewisse Schablonen rationalen wissenschaftlichen Verhaltens preiszugeben*. Es scheint mir nämlich, daß Kuhn für die Wissenschaftstheorie genau das Gegenteil von dem getan hat, was ihm die Kritiker vorwerfen: Er hat nicht die Irrationalität im Verhalten von Wissenschaftlern aufgezeigt, sondern *neue Dimensionen wissenschaftlicher Rationalität erblickt*. (Stegmüller [315], 74)

So gesehen schließt Kuhn wieder da an, „wo Mach aufgehört hatte.“<sup>63</sup> Wie bei Mach um die Prinzipien der Denkökonomie, geht es bei Kuhn um die Rationalität wissenschaftlichen Forschens, nur fangen beide nicht bei den Rationalitätsstandards, sondern bei der Wissenschaftsgeschichte an. Ihr Verlauf und die aufgrund dieses Verlaufs erreichten Erfolge sind erklärungsbedürftig, und eine Theorie wissenschaftlicher Rationalität ist so gut, wie sie diese Erklärungsleistung erfüllen kann.

In diesem Zusammenhang ist auch Kuhns Auseinandersetzung mit dem Thema ‚Gedankenexperiment‘ zu sehen. Wie Ulrich Kühne (Kühne [174], Kühne [175]) bereits aufgezeigt hat, kommt dem Gedankenexperiment bei Kuhn gerade in Phasen der Krise eine Vermittlungsfunktion zwischen den (nach der radikalen Interpretation) unvermittelbaren konkurrierenden Paradigmen zu. Damit gehört das Gedankenexperiment gerade zu den Elementen, die die nach bisherigen Rationalitätsstandards irrational erscheinende „Konversion“ zum neuen Paradigma „rationalisieren“ können:

Eine Krise, ausgelöst durch den Fehlschlag von Erwartungen und gefolgt von einer Revolution, steht im Mittelpunkt der Gedankenexperiment-Situationen [...]. Umgekehrt ist das Gedankenexperiment eines der wesentlichen analytischen Mittel, die während der Krise eingesetzt werden und dann zur grundlegenden theoretischen Neuorientierung beitragen. (Kuhn [173], 350)

Wie können Gedankenexperimente dies leisten?

---

<sup>63</sup> Kühne [175], 269.

### 2.3.1 DIE STANDARDAUFFASSUNG UND DAS PARADOX DES GEDANKEN-EXPERIMENTS

In seinem Aufsatz ‚Eine Funktion für das Gedankenexperiment‘, formuliert Kuhn seine Fragestellung genau entsprechend dem oben gesagten: falls es so ist, dass Gedankenexperimente in der Geschichte der Naturwissenschaften an entscheidender Stelle vorkommen, muss zumindest der Historiker in ihnen ein „manchmal sehr wirksames Werkzeug zur Verbesserung der Naturerkenntnis sehen.“ Von einem wissenschaftstheoretischen Standpunkt aus ist diese Rolle von Gedankenexperimenten jedoch rätselhaft: Gedankenexperimente beziehen sich auf Verhältnisse, die nicht untersucht sind, manchmal nicht einmal untersucht werden können (etwa weil sie physikalisch unmöglich sind). Nach Kuhn führt das zu drei Hauptproblemen: (a) Welche Bedingungen der Wirklichkeitsnähe gelten für Gedankenexperimente?, (b) Wie kann durch ein Gedankenexperiment eine neue Naturerkenntnis entstehen, wenn die empirischen Daten, die in ein Gedankenexperiment eingehen, doch bereits bekannt sein mussten? und (c) Was für neue Erkenntnisse lassen sich überhaupt durch ein Gedankenexperiment gewinnen?

Kuhn scheint mit der Reihenfolge dieser Fragen keine Systematik zu ihrer Beantwortung im Auge zu haben, was schon dadurch klar wird, dass die erste Frage sinnlos ist, wenn die Antwort auf die dritte lautet, dass sich aus Gedankenexperimenten keine neue Naturerkenntnis gewinnen lässt. Genau dies ist nämlich laut Kuhn die Standardauffassung vom Gedankenexperiment:

[Laut der Standardauffassung] bezieht sich die durch Gedankenexperimente erzeugte neue Erkenntnis nicht auf die Natur, sondern auf den theoretischen Apparat des Wissenschaftlers. Nach dieser Auffassung hat das Gedankenexperiment die Funktion, Verwirrungen aufzuklären, indem es den Wissenschaftler zur Erkenntnis von Widersprüchen in seinem bisherigen Denken zwingt. Im Unterschied zur Entdeckung neuer Tatsachen scheint die Beseitigung einer bestehenden Verwirrung keine zusätzlichen empirischen Daten zu erfordern. Auch braucht die Situation in der Natur nicht wirklich vorzukommen. Vielmehr gilt für das Gedankenexperiment, das lediglich Verwirrung auflösen soll, einzig die Bedingung der Wirklichkeitsnähe. Die vorgestellten Verhältnisse müssen so beschaffen sein, daß der Wissenschaftler seine Begriffe in der bislang normalerweise von ihm geübten Art und Weise anwenden kann. (Kuhn [173], 328)

Diese „Standardauffassung“<sup>64</sup> ist laut Kuhn zwar „ungeheuer einleuchtend“, aber eben nicht die ganze Wahrheit. Insbesondere sei unbefriedigend, dass sie die Situation des Forschers vor der Durchführung des Gedankenexperiments als „widersprüchlich“ oder „verworren“ bezeichne. Kuhns Ziel ist es, der Standardauffassung eine Alternative gegenüber zu stellen.

Zur Erläuterung der Standardauffassung, wie auch seiner alternativen Auffassung analysiert Kuhn ein Beispiel aus der Kinderpsychologie und eines aus der Wissenschaftsgeschichte.

---

<sup>64</sup> Kuhn hat dabei vermutlich Popper im Auge. Vgl. auch Kühne [175], 270.

### 2.3.2 BEGRIFFSWANDEL BEI KINDERN ALS MODELL FÜR WISSENSCHAFTLICHEN BEGRIFFSWANDEL<sup>65</sup>

Bei seinem Beispiel aus der Kinderpsychologie handelt es sich um einen Versuch Piagets aus dem Jahr 1946, bei dem er kleinen Kindern Zeichentrickfilme von Autorennen zeigte und sie danach befragte, welches der Autos aus dem Zeichentrickfilm das „schnellere“ gewesen sei.<sup>66</sup>

In diesen Trickfilmen werden ein blaues und ein rotes Auto gleichförmig auf einer geraden Strecke bewegt. Manchmal starten beide gleichzeitig vom selben Ausgangspunkt, manchmal startet eins später, holt das andere am Ziel aber ein, manchmal startet eins aus größerem Abstand zum Ziel, etc.

Eine bestimmte Altersgruppe von Kindern verwendet bei der Beurteilung, welches Auto „schneller“ gewesen ist, zwei verschiedene Kriterien. Einmal ein „Zielerreichungskriterium“, bei dem dasjenige Auto „schneller“ war, welches als erstes durchs Ziel fuhr, und dann noch ein Kriterium der „Wahrnehmungsunschärfe“, bei dem ein Auto als „schneller“ beurteilt wurde, das wesentlich später startete, das andere dennoch am Ziel erreichte (die Kinder meinten in diesen Fällen man habe es „sehen“ können, dass das eine Auto „schneller“ war).

Diese beiden Kriterien können zu widersprüchlichen Resultaten führen (beispielsweise, wenn ein Auto, das die Strecke mit höherer Geschwindigkeit durchfahren hat, später durchs Ziel kam, weil es später gestartet ist). Kinder, denen dies im Experiment auffällt, können daraus lernen und entwickeln einen neuen Begriff, indem sie ihren alten Begriff von „schneller“ möglicherweise aufspalten. So würde zumindest die Standardauffassung diese Situation beschreiben. Die Kinder haben zunächst einen in sich widersprüchlichen Begriff von „schneller“, die Experimentssituation bringt sie dazu, diesen Widerspruch festzustellen, daraufhin wird dieser Widerspruch durch Begriffswandel ausgeräumt. Damit Kinder aus dieser Situation lernen konnten, mußte zumindest gewährleistet sein, dass sie ihre Begriffe auf herkömmliche Weise anwenden konnten (deshalb geht es im Experiment um bewegte Autos und nicht um Kohlköpfe, die aus dem Nichts im Ziel auftauchen).

### 2.3.3 „IN MEINEN OHREN KLINGT DAS WIE EIN SCHROFFER WIDERSPRUCH. WAS MEINT IHR, SIGNORE SIMPLICIO?“

Als nächstes betrachtet Kuhn ein Gedankenexperiment Galileo Galileis, das im *Dialog über die beiden hauptsächlichsten Weltsysteme: Das Ptolemäische und das Kopernikanische* ziemlich am Anfang des ersten Tages vorkommt.<sup>67</sup> Wir wollen dieses

<sup>65</sup> Dieses Beispiel wird uns im Schlusskapitel noch einmal in einem anderen Zusammenhang begegnen.

<sup>66</sup> Vgl. Piaget [257], Kapitel 7; Kühne [175], 272; Kuhn [173], 329-332.

<sup>67</sup> Das Kuhn hier „ein für die bisherige Literatur neues Beispiel“ (Kühne [175], 269) betrachtet, ist übrigens nicht richtig. Mach erwähnt bereits Galileis Gedankengang und verweist in

Gedankenexperiment hier etwas ausführlicher besprechen, da es in der Regel relativ unverständlich und ohne den argumentativen Kontext präsentiert wird.<sup>68</sup> Wie auch in den *Discorsi* diskutieren hier Galileis *alter ego* Salviati, der aufgeschlossene neutrale Sagredo und der Aristoteliker Simplicio miteinander die Probleme der Aristotelischen Physik und die Vorteile der modernen Auffassungen.

Nachdem Salviati überzeugend dargelegt hat, warum man drei Dimensionen der Ausdehnung betrachtet, kommen die Dialogpartner auf die Ursachen von Bewegung und Ruhe zu sprechen. In diesem Zusammenhang trägt Salviati (ohne dabei irgendeine aristotelische Auffassung explizit angreifen zu wollen) die These vor, dass jeder Körper, der aus der Ruhe in Bewegung gebracht wird, dabei eine kontinuierliche Beschleunigung erfährt, also bei der Beschleunigung durch alle „Stufen geringerer Geschwindigkeit“ oder „größerer Langsamkeit“ hindurchgehen muss.<sup>69</sup> Dieses Thema weckt plötzlich das Interesse Sagredos:

**Sagr.** Ich höre von dieser Ansicht mit grossem Vergnügen, glaube aber, das wird in noch höherem Masse der Fall sein, wenn ihr mir erst ein Bedenken beseitigt habt; ich begreife nämlich nicht recht, wieso notwendig ein beweglicher Körper, der aus dem Zustande der Ruhe in eine Bewegung eintritt, zu der ein natürlicher Hang ihm innewohnt, alle vorhergehenden Grade der Schnelligkeit durchmachen muss, deren es zwischen einem beliebig vorgeschriebenen Grade und dem Zustande der Ruhe unendlich viele giebt: als wenn die Natur der Masse des Jupiter nicht gleich nach ihrer Schöpfung die kreisförmige Bewegung nebst der betreffenden Geschwindigkeit hätte zuerteilen können. (Galilei [118], 22)

Hierauf beginnt Salviati mit einem längeren Exkurs in dessen Zentrum das von Kuhn beschriebene Gedankenexperiment steht. Salviati vergewissert sich für seine Argumentation zunächst der Zustimmung seiner Dialogpartner zu der Ansicht, dass eine Kugel beim Fall immer größere Geschwindigkeit erlangt, je weiter sie sich vom Ausgangspunkt der Bewegung entfernt, bzw. je näher sie dem „angestrebten Mittelpunkt“ kommt (gemeint ist der Erdmittelpunkt als „Ziel“ der Bewegung)<sup>70</sup>, sowie dazu, dass zu jedem Zeitpunkt der Bewegung die Kugel genau auf eine solche Geschwindigkeit beschleunigt ist, wie erforderlich wäre, um die Kugel wieder zum Ausgangspunkt der Bewegung zurückzubringen. Als nächstes bittet Salviati seine beiden Dialogpartner, sich folgenden Fall vorzustellen. Gegeben sei eine schiefe Ebene  $CA$  und eine senkrechte  $CB$ :

---

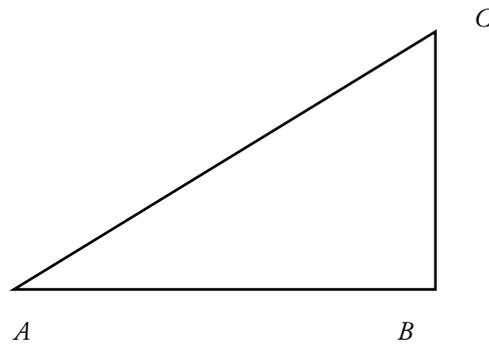
diesem Zusammenhang auch auf sein Kapitel ‚Über Gedankenexperimente‘ in *Erkenntnis und Irrtum*. Vgl. Mach [193], 154.

<sup>68</sup> Z.B. bei Kuhn selbst, der gar nicht erwähnt, weshalb sich Galilei überhaupt die Mühe macht, diese Argumentation durchzuführen.

<sup>69</sup> Galilei [118], 21.

<sup>70</sup> Galilei redet auch an dieser Stelle nicht von der *Dauer* des Falls, sondern nur von der zurückgelegten Strecke in Bezug auf den Ausgangs- oder Endpunkt der Bewegung. Diese Betrachtungsweise wird ihn zeitweilig gehindert haben, das Fallgesetz als abhängig von der Fallzeit zu formulieren. Vgl. hierzu aber Drake [90], Drake [91] und Mach [193], 154.

Die schiefe Ebene  $CA$  sei absolut glatt. Wenn nun eine „vollkommen runde Kugel von härtestem Stoff“ sich entlang  $CA$  bewegt und eine zweite derartige Kugel ent-



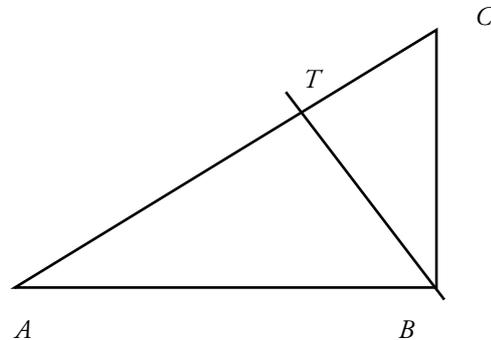
lang  $CB$  fällt, so sollte nach dem, was zuvor über die Geschwindigkeit in Relation zur Annäherung an den Mittelpunkt vereinbart wurde, der „Antrieb“ der beiden Kugeln nach Durchlaufen der jeweiligen Strecke genau gleich sein, da sich beide dem Mittelpunkt gleichviel genähert haben.

Allerdings sind sich die Dialogpartner ebenfalls einig, dass dieselbe Kugel auf der Ebene  $AB$  ruhen würde, sich auf der Ebene  $CA$  hingegen abwärts bewegte, diese Bewegung aber *langsamer* sein müsste als die Bewegung längs der Senkrechten  $CB$ . Sagredo wittert hier einen Widerspruch, den Salviati auf den Punkt bringt:

Salv. Umsomehr wird es Euch unrichtig vorkommen, wenn ich behaupte, dass die Geschwindigkeit des senkrecht und des schief fallenden Körpers genau gleich sind. Und doch ist dies vollkommen richtig, ebenso richtig wie die Behauptung, dass der Fall längs der Senkrechten schneller erfolgt als längs der schiefen Ebene. (Galilei [118], 25)

Salviati bittet seine Dialogpartner daraufhin, zunächst eine Begriffsklärung vorzunehmen. Was heißt es überhaupt, dass ein Körper schneller ist als ein anderer, bzw. wann sind zwei Körper gleich schnell? Als Antwort entwickeln sie, dass zwei Körper gleich schnell sind, wenn sich die Zeiten in denen sie sich durch eine Strecke bewegen in derselben Proportion zueinander verhalten, wie die zurückgelegten Strecken.

Laut Kuhn macht dies jedoch „alles nur noch schlimmer“. Vergleicht man mit dieser Definition einzelne Streckenteile und die Zeiten, in denen die Kugeln sich durch diese bewegen, so erhält man, beginnt man mit der senkrecht fallenden Kugel und der Strecke  $CB$ , das Ergebnis, dass sie in einer gegebenen Zeit eine größere Strecke durchläuft als die Kugel, die  $CA$  herabrollt. Ist erstere nämlich bei  $B$  angekommen, ist letztere gerade bei  $T$ .



Demzufolge wäre die senkrecht fallende Kugel schneller. Betrachtet man aber ein  $CB$ -langes Vergleichsstück am Ende der schiefen Ebene, dreht sich das Verhältnis genau um. Jetzt braucht die Kugel auf der schiefen Ebene weniger Zeit um dieselbe Strecke zurückzulegen, die die senkrecht fallende Kugel zurücklegt. Wählt man eine Strecke mit der Länge  $CB$  geschickt irgendwo auf der schiefen Ebene, wird die Zeit, die beide Kugeln brauchen um dieselbe Strecke zurückzulegen, sogar gleich sein. Das Ergebnis dieser Betrachtung präsentierte Salviati schon vor diesem Gedankengang:

**Salv.** Die Behauptung also, die Bewegung längs der Senkrechten sei schneller als längs der schiefen Linie, erweist sich nicht als allgemein richtig, wie ihr seht. Sie gilt nur bei Bewegungen, die vom Anfangspunkte, also von der Ruhelage, ihren Ausgang nehmen. (Galilei [118], 27)

Man muss also verschiedene Geschwindigkeitsbegriffe auseinander halten. Einmal den Begriff der Momentangeschwindigkeit in einem Punkt und dann den Begriff der Durchschnittsgeschwindigkeit für festgelegte Streckenabschnitte. Nachdem diese Klärung erreicht ist, kommt Salviati auf das eigentliche Thema zurück, die Frage, ob „ein schwerer Körper, der von der Ruhelage ausgeht, bei seinem Falle durch alle Stufen der Langsamkeit hindurchgehen muss, welche einer später von ihm erreichten Stufe der Geschwindigkeit vorangehen.“<sup>71</sup> Der dann folgende Gedankengang sei ebenfalls kurz skizziert:

Wir wissen nun, dass eine Kugel, die die schiefe Ebene nach  $A$  heruntergerollt ist, dort denselben „Antrieb“ hat wie eine Kugel, die von  $C$  senkrecht herunterfällt, in diesem Sinne beide Kugeln an diesem Punkt also „gleich schell“ sind (sie haben dieselbe Momentangeschwindigkeit). Wir können nun  $C$  beliebig nahe an  $B$  annähern.

Galilei führt dann an, dass wenn eine Kugel die schiefe Ebene von  $C$  nach  $A$  passiert hat, sie dort gerade soviel Antrieb (Momentangeschwindigkeit) besitzt, dass sie in derselben Zeit bei fortgeführter Bewegung auf der Horizontalen nun

---

<sup>71</sup> Galilei [118], 29.

die doppelte Strecke zurücklegen würde (dies zeigt Galilei erst in den *Discorsi*<sup>72</sup>). Dann kann man aber einen Punkt *C* so nahe an *B* finden, dass nach einem senkrechten Fall von *C* nach *B* die Kugel dort gerade so viel Antrieb hätte, wie sie benötigt um in hundert Jahren die doppelte Strecke von *C* nach *A* in der Horizontalen zurückzulegen. Das Gedankenexperiment macht also Galilei nur den Weg frei, um eine Gesetzmäßigkeit für die schiefe Ebene auf den freien Fall anwenden zu können, ohne dass Galilei sich mit dem ungeklärten Geschwindigkeitsbegriff der aristotelischen Physik in scheinbare Widersprüche verheddern muss.

Wieder kann man das Gedankenexperiment also nach der Standardauffassung interpretieren:

Die von Aristoteles zur Analyse der Bewegung verwandten Begriffe waren teilweise widersprüchlich, und das war während des Mittelalters nicht völlig ausgeräumt worden. Galileis Gedankenexperiment förderte diese Schwierigkeit zutage, indem es dem Leser die Paradoxie vor Augen stellte, die in ihrem Denken steckte. Damit half es ihnen bei der Veränderung ihres Begriffsapparates. (Kuhn [173], 338)

Und wieder ist das einzige Kriterium, dem das Gedankenexperiment genügen muss, um wirksam zu sein, dass die Standardbedingungen für die Anwendung der Begriffe ‚Schneller‘ und ‚Geschwindigkeit‘ gegeben sind.

#### 2.3.4 DIE INADÄQUATHEIT DER STANDARDAUFFASSUNG

Kuhns Haupteinwand gegen die Standardauffassung besteht darin, dass diese die Wissenschaftler vor der Durchführung des Gedankenexperiments zu schlecht wegkommen lässt. In gewisser Weise zieht er das *principle of charity* heran: Wenn die Standardauffassung die richtige Interpretation der historischen Situation gibt, dann müssen wir allen Wissenschaftlern, die während der 2000 Jahre seit Aristoteles mit dem Geschwindigkeitsbegriff gearbeitet haben, ein inkonsistentes Überzeugungssystem zuschreiben.

Zunächst stellt Kuhn fest, dass die „widersprüchlichen“ bzw. „verworrenen“ Begriffe, die durch Gedankenexperimente als fehlerhaft entlarvt werden, keine begrifflichen Unmöglichkeiten darstellen, wie z.B. der Begriff ‚rundes Quadrat‘. Letzterer ist in keiner möglichen Welt exemplifiziert, runde Quadrate sind logisch unmöglich. Ein aristotelischer Geschwindigkeitsbegriff hingegen ist in diesem Sinne nicht logisch unmöglich. In Welten, in denen ausschließlich gleichförmige Bewegung vorkommt, können Momentangeschwindigkeit und Durchschnittsgeschwindigkeit nie voneinander abweichen. Entsprechend muss in einer solchen Welt auch zwischen diesen beiden Geschwindigkeitsbegriffen nicht unterschieden

---

<sup>72</sup> Galilei [117], 60: „Theorem XVI. Propos. XXV. Wenn nach dem Fall längs einer geneigten Ebene die Bewegung in der Horizontalen fortgesetzt wird, so verhält sich die Fallzeit längs der Geneigten, zur Zeit der Bewegung längs irgendeiner Strecke in der Horizontalen, wie die doppelte Länge der geneigten zur horizontalen Strecke.“

werden. Es genügt, dass „ein Begriff, wie ein Gesetz oder eine Theorie in jeder Situation eindeutig anwendbar ist, die wir jemals erwarten.“<sup>73</sup>

Die Aristotelische Physik geht nun so vor, „als ob“ wir in einer solchen Welt leben würden. Indem sie einen Geschwindigkeitsbegriff „tief“ in ihr System „einbaut“, der zwischen Momentan- und Durchschnittsgeschwindigkeit nicht unterscheidet, beschreibt sie die Welt so, als ob in unserer Welt nur gleichförmige Veränderungen stattfänden. Der Aristotelische Geschwindigkeitsbegriff hat demnach empirische Konsequenzen. Es handelt sich nicht um eine bloße Konvention, die nach rein logischen Gesichtspunkten zu beurteilen ist, sondern er fungiert „nahezu als Naturgesetz“.

Dass dies 2000 Jahre lang funktionierte, liegt schlicht daran, dass Ereignisse, die zu Problemen bei der Begriffsanwendung hätten führen können, relativ selten vorkamen und daher erfolgreich „an den äußersten Rand“ des „wissenschaftlichen Bewusstseins“ verdrängt werden konnten. Der Begriff selbst war daher nicht „verworfen“ oder „widersprüchlich“, sondern er entsprach nicht der „gesamten Feinstruktur der Welt, auf die er passen sollte.“<sup>74</sup>

### 2.3.5 KUHN'S FUNKTION FÜR DAS GEDANKENEXPERIMENT

Nach Kuhn besteht die Funktion des Gedankenexperiments (bzw. einer wichtigen Teilklasse von Gedankenexperimenten) also nicht darin, Widersprüche in Theorien oder Begriffssystemen aufzudecken, und damit den Wissenschaftler über sein Theoriensystem zu unterrichten. Gedankenexperimente erinnern uns vielmehr daran, was wir bisher an den Rand unseres wissenschaftlichen Bewusstseins gedrängt haben, und zeigen auf, dass es sich um ernstzunehmende Anomalien handelt.

Nach dieser Auffassung erschließen Gedankenexperimente dem Wissenschaftler Informationen, die zwar einerseits bereits vorhanden sind, aber in einem anderen Sinne (noch) unzugänglich sind. Das Gedankenexperiment fügt sich dabei in Kuhns Theorie wissenschaftlicher Revolutionen ein.

Die Entwicklung eines ausgereiften wissenschaftlichen Spezialgebiets wird weitgehend durch ein eng verwobenes System von Begriffen, Gesetzen, Theorien und apparativen Methoden bestimmt. Dieses System legt u.a. fest, welche Probleme als zulässige wissenschaftliche Fragestellung im Rahmen des vorgegebenen Theoriengebäudes akzeptiert werden. Durch diese selektive Problemauswahl wird von den verfügbaren Daten nur eine begrenzte Teilmenge als relevant eingestuft und berücksichtigt. Auch diese Datenmenge wird die Theorie nicht vollständig bewältigen: Einige konfligierende Daten können zwar durch leichte theoretische Korrekturen in bestätigende Instanzen der Theorie transformiert werden. Andere werden sich jedoch zu dauerhaften Anomalien entwickeln, die die Verfechter des Theoriege-

---

<sup>73</sup> Kuhn [173], 341.

<sup>74</sup> Kuhn [173], 345.

bäudes nach Möglichkeit, wie Kuhn sich ausdrückt, „an den Rand ihres wissenschaftlichen Bewusstseins“ zu drängen versuchen. (Gähde [114], 193)

Gedankenexperimente *erinnern* nun an diese verdrängten Daten, die gerade dabei sind, wieder mehr ins Zentrum der wissenschaftlichen Aufmerksamkeit zu rücken. In diesen Gedankenexperimenten wird der Konflikt zwischen den Implikationen einer Theorie und den verdrängten Daten zugespitzt. Gleichzeitig bietet das Gedankenexperiment aber auch die Grundlage zur Neuorientierung. Nach Galileis Gedankenexperiment ist klar, dass der Aristotelische Geschwindigkeitsbegriff für das anvisierte Thema (Fallbeschleunigung) inadäquat ist.

Wenn wir zu den drei Kuhnschen Hauptfragen zurückkommen, die sich zum Thema Gedankenexperiment stellen, gelangen wir nun zu anderen Antworten als die „Standardauffassung“:

*(a) Welche Bedingungen der Wirklichkeitsnähe gelten für Gedankenexperimente?*

Gedankenexperimente müssen zumindest von Verhältnissen handeln, die in der Natur vorkommen könnten. Es reicht nicht, dass bloß die Bedingungen für die Anwendung eines bestimmten Begriffs gegeben sind, die Natur muss den „Widerspruch“, den das Gedankenexperiment aufweisen möchte, prinzipiell selbst präsentieren können. Es muss sich um eine wirkliche Anomalie handeln, von der – wenn sie nicht in der Erfahrung der Wissenschaftler bereits so vorgekommen ist – zumindest allen klar ist, dass sie so vorkommen könnte.

*(b) Wie kann durch ein Gedankenexperiment eine neue Naturerkenntnis entstehen, wenn die empirischen Daten, die in ein Gedankenexperiment eingehen, doch bereits bekannt sein mussten?*

Die neue Naturerkenntnis kann entstehen, weil die Daten zwar bekannt, aber verdrängt waren. Bisher war man der Meinung, dass es sich um eine Anomalie handelt, die durch neue apparative Methoden oder Ähnliches zum Verschwinden gebracht werden könnte. Das Gedankenexperiment zeigt auf, dass die bisher als problematisch aber marginal betrachteten Daten eindeutig von Relevanz für das Theoriegebäude sind, die Daten werden durch das Gedankenexperiment in diesem Sinne „aufgewertet“.<sup>75</sup>

*(c) Was für neue Erkenntnisse lassen sich überhaupt durch ein Gedankenexperiment gewinnen?*

Die Erkenntnisse, die man aus einem Gedankenexperiment gewinnt, betreffen nach Kuhn nicht ausschließlich das eigene Theoriegebäude. Man lernt nicht bloß, dass man bisher einen peinlichen logischen Schnitzer gemacht hat, sondern erfährt auch etwas über die Natur. Man erfährt, dass in das bisherige Begriffssystem des Theoriegebäudes bestimmte Erwartungen über das Verhalten der Natur eingebaut waren, Erwartungen, von denen man aus der Erfahrung bereits weiß, dass sie von der Natur nicht erfüllt werden, was man aber bisher nicht realisiert hat. Indem das

---

<sup>75</sup> Vgl. Gähde [114], 193.

Gedankenexperiment einen Widerspruch deutlich macht, zwingt es dazu, das Begriffssystem zu revidieren und damit die darin implizit enthaltenen Erwartungen der Feinstruktur der Natur weiter anzupassen.

### 2.3.6 KONZESSIONEN AN DEN GEGNER?

Vergleicht man Kuhns Auffassung vom Gedankenexperiment mit der Popper'schen „Standardauffassung“, wird außerdem deutlich, dass Poppers methodologische Forderung bei kritischen und apologetischen Gedankenexperimenten nur solche Idealisierungen zuzulassen, die Konzessionen an den Gegner darstellen, hier keinen rechten Sinn macht.

Wenn wir die Idealisierungen im Galileischen Gedankenexperiment betrachten, bestehen diese in der Hauptsache darin, dass die rollende, bzw. fallende Kugel absolut rund ist und die schiefe Ebene reibungsfrei. Aus diesem Grund kann Galilei überhaupt davon ausgehen, dass die beiden qualitativ identischen Kugeln am Ende der Strecken *CB* bzw. *CA* dieselbe Momentangeschwindigkeit haben, nämlich genauso viel, wie nötig wäre, um sie wieder zum Ausgangspunkt ihrer Bewegung zurückzubringen. Ohne diese Idealisierung gäbe es keinen Widerspruch mit einer konsequenten Anwendung eines Durchschnittsgeschwindigkeitskriteriums. Die senkrecht fallende Kugel ist eben schneller.

Würde das Gedankenexperiment Konzessionen an den Gegner machen, müsste es die „Welt“ so idealisieren „als ob“ alle Bewegung gleichförmig wäre, oder „als ob“ keine Differenzen zwischen Momentan- und Durchschnittsgeschwindigkeit auftauchen könnten. Ein Gedankenexperiment, das solche Konzessionen macht, könnte dann die Funktion, die Kuhn ihm zuweist, gerade nicht erfüllen.

Ganz im Gegenteil, muss die Welt im Gedankenexperiment geradezu „bösaartig“ sein. Fälle, die selten auftreten, die bisher erfolgreich verdrängt werden konnten, werden herangezogen, um einen Widerspruch zu produzieren.

Kuhn hat seine Vorstellung von der Funktionsweise des Gedankenexperiments nicht präzise ausgeführt. Dies ist in der gegenwärtigen wissenschaftstheoretischen Beschäftigung mit Gedankenexperimenten, der wir uns als nächstes zuwenden wollen, nachgeholt worden. Dabei ist nicht ausgeblieben, dass Korrekturen an Kuhns Auffassung vom Gedankenexperiment vorgenommen werden mussten.

### 3. Eine allgemeine Theorie des Gedankenexperiments

Durch das letzte Kapitel haben wir einen halbwegs angemessenen Überblick darüber gewinnen können, welche Rolle die Methode des Gedankenexperiments in der Wissenschaftstheorie der Naturwissenschaften spielt. Wir haben dabei feststellen können, dass man diese Rolle auf zwei sehr verschiedene Arten einschätzen kann. Entweder gibt man der Wissenschaftsgeschichte<sup>76</sup> den Vorzug, dann hat man es mit einer zusätzlichen Methode wissenschaftlichen Forschens zu tun, von der die in der Wissenschaftsgeschichte beteiligten Forscher anzunehmen scheinen, dass sie in den Rechtfertigungszusammenhang gehört. Dies ist die Vorgehensweise bei Mach und Kuhn. Beide haben dann die Schwierigkeit, die Funktionsweise dieser neuen Methode zu erklären.

Oder man gibt der erkenntnistheoretischen Analyse den Vorzug, und zählt nur das zu den Methoden, was sich als eine der als gerechtfertigt bekannten, bereits identifizierten Methoden rekonstruieren lässt. Diese Vorgehensweise fanden wir bei Duhem, Hempel und Popper. Dann ist die Bezeichnung ‚Gedankenexperiment‘ für eine Methode der Wissenschaften aber in der Gefahr, redundant zu sein. (Man erinnere sich an Meinong: Warum nennt man „Gedankenexperimentieren“ nicht einfach weiterhin ‚Denken‘?)

Die gegenwärtige Wissenschaftstheorie versucht zwischen diesen alternativen Vorgehensweisen einen Mittelweg zu finden. Einerseits ist man sich weitestgehend einig, dass sich hinter dem Gedankenexperiment nicht das lang erträumte Verfahren verbirgt, a priori Naturerkennnisse zu gewinnen (wobei James Robert Brown und Alexandre Koyré hier als Exotenpositionen ausgenommen sind<sup>77</sup>). D.h., wenn man durch Gedankenexperimente neue Erkenntnisse gewinnt, dann muss sich unter Rückgriff auf als gerechtfertigt anerkannte Prozeduren angeben lassen, wie Gedankenexperimente dies leisten.

---

<sup>76</sup> Natürlich sieht sich die weiter unten charakterisierte Position auch als eine Form von Wissenschaftsgeschichte. Hier ist gemeint, dass man tatsächlich versucht aus einer empirischen Analyse historischen wissenschaftlichen Forschens zu wissenschaftstheoretischen Einsichten vorzustoßen.

<sup>77</sup> Brown [38], Brown [39], Brown [40], Brown [41], Brown [42] und Koyré [168], Koyré [169] behaupten beide, dass gute Wissenschaft a priori gemacht wird und eine Untersuchung des Gedankenexperiments uns darüber aufklärt, wie man das anstellt. Wir werden weiter unten noch kurz auf diese Position eingehen. Koyré ist zu seiner Überzeugung hauptsächlich deswegen gekommen, weil seine historischen Analysen der Galileischen Experimente suggerierten, dass Galilei fast nie tatsächlich experimentiert habe. Dieser Befund wird heutzutage aber zurückgewiesen. Vgl. Kühne [175] und MacLachlan [197].

Andererseits können die Prozeduren, die dabei im Gedankenexperiment integriert werden, vielleicht spezifisch genug umrissen werden, um gerechtfertigter Weise von einer weiteren Methode sprechen zu können.

Die Auffassungen, die uns in diesem und dem nächsten Kapitel beschäftigen werden, sind aus zwei Gründen interessant.

*Erstens* handelt es sich bei den nun vorzustellenden Positionen um den gegenwärtigen „Konsens“, also um den „state of the art“ der wissenschaftstheoretischen Analyse des Gedankenexperiments in den Naturwissenschaften, bei dem – wie gesagt – versucht wird, eine Synthese der bisher betrachteten Ansätze herzustellen. Die Synthese besteht dabei (hauptsächlich) in Folgendem:

(1) Gedankenexperimente müssen argumentationstheoretisch verstanden werden. Gedankenexperimente sind typischerweise an Adressaten gerichtet. Diese Adressaten sollen zu einer rationalen Überzeugungsänderung gebracht werden. Diese Einsicht geht auf Popper zurück. Daraus folgt, dass die Funktionsweise von Gedankenexperimenten zum Teil mit argumentationstheoretischen Mitteln erfolgen muss.

(2) Gedankenexperimente machen in den meisten Fällen zusätzliche empirische Annahmen, sind also – contra Popper – keine bloßen internen Konsistenztests. Wie Mach und Kuhn schon argumentiert haben, ist klärungsbedürftig, woher diese zusätzlichen Annahmen ihre Rechtfertigung beziehen.

(3) Wenn die Rechtfertigung dieser empirischen Annahmen gelingen soll, muss sie aus der Erfahrung stammen. Mach hatte vorgeschlagen, dafür den Erfahrungsbegriff auf dasjenige „instinktive Wissen“ auszudehnen, das wir aus der Evolutionsgeschichte mitbekommen haben.

(4) Mach hat zu Recht darauf hingewiesen, dass Gedankenexperimente ihren „Witz“ im weiteren Verlauf der Wissenschaftsgeschichte verlieren können. Gedankenexperimente werden in bestimmten wissenschaftshistorischen Situationen eingesetzt, weil man auf andere Weise die intendierte rationale Überzeugungsänderung nicht erwirken kann.

*Zweitens* ist aus den nun vorzustellenden Erkenntnissen über Gedankenexperimente in den Naturwissenschaften auf die Rolle des Gedankenexperiments in der Philosophie zurück geschlossen worden.

### 3.1 DREI TYPEN VON GEDANKENEXPERIMENTEN

Wir wollen nun die einzelnen Fäden aufgreifen, und Gedankenexperimente in den Naturwissenschaften systematisch charakterisieren. Wie wir zu Anfang des zweiten Kapitels bereits bemerkt haben, sind wissenschaftliche Methoden zwar durch ihre Funktion zu charakterisieren, aber das schließt nicht aus, dass wissenschaftstheoretische Bezeichnungen ambig sind und Verfahren mit verschiedenen Funktionen bezeichnen. Wir wollen daher zunächst verschiedene Funktionen des „Gedankenexperimentierens“ unterscheiden, um dadurch diejenige Funktion identifizieren zu können, die das wissenschaftstheoretische Interesse ausgelöst hat.

## 3.1.1 KLÄRENDE GEDANKENEXPERIMENTE

Wie wir bei Hempel gesehen haben, kann man zwischen „theoretischen“ und „intuitiven“ Gedankenexperimenten unterscheiden. Theoretische Gedankenexperimente, wie Hempel sie charakterisiert, werden nicht vorgetragen, um eine gesetzesartige Aussage zu stützen oder zu erschüttern, sondern häufig nur in der Absicht, unterschiedliche empirische Konsequenzen rivalisierender gesetzesartiger Aussagen deutlich zu machen. Betrachten wir hierzu einige Beispiele, zwei aus der ferneren und eins aus der jüngeren Wissenschaftsgeschichte.

Im vierten Buch von *De Caelo* diskutiert Aristoteles die Ursachen dafür, dass sich manche Dinge nach oben (wie beispielsweise Feuer), andere (wie beispielsweise Erde) zum Erdmittelpunkt hin bewegen. Dabei räumt er das mögliche Missverständnis aus, dass dabei dem Erdmittelpunkt als solchem eine kausale Rolle zukommt:

Wenn man nämlich die Erde dorthin versetzte, wo jetzt der Mond ist, so wird sich jeder ihrer Teile durchaus nicht zu ihr hinbewegen, sondern dorthin, wo die Erde jetzt ist. (Aristoteles [2], 171)

Diese Textstelle ist so gedeutet worden, dass Aristoteles hier nicht beabsichtigt, die kausale Rolle des Mittelpunkts des Universums mit einem Gedankenexperiment argumentativ zu *etablieren*, sondern nur *deutlich machen* will, dass der Mittelpunkt der Erde, der mit dem Mittelpunkt des Universums zusammenfällt, nur qua Letzterem Endpunkt der Bewegung schwerer Körper ist. Es geht ihm also darum ein mögliches Missverständnis seiner Theorie auszuräumen.<sup>78</sup>

Duhem berichtet, dass William Gilbert seine „magnetische Philosophie“ als Alternative zu dieser Theorie verstanden haben wollte, wonach Schwere als Anziehung von Ähnlichem zu Ähnlichem betrachtet wird:

Alles, was irdisch ist, vereinigt sich mit der Erdkugel; und ebenso strebt alles, was mit der Sonne gleichartig ist, der Sonne, alle dem Mond angehörigen Dinge dem Monde zu und dasselbe gilt für alle Körper, die das Universum bilden. (Gulielmi Gilberti Colcestrensis, medici Londinensis: *De magnete, magneticis corporibus, et de magno magnete Tellure, physiologia nova*; Londini 1600, 115, zitiert nach Duhem [94], 309-310)

Kepler hingegen vertrat eine dritte Theorie, wonach weder der Mittelpunkt des Universums, noch die qualitative Gleichheit der Massen eine Rolle spielt, sondern Gravitation als Anziehungskraft zwischen jeglichen zwei Körpern wirkt. Der Unterschied zwischen diesen drei Auffassungen von Gravitation ist – laut Sorensen – von Marin Mersenne durch ein Gedankenexperiment erläutert worden<sup>79</sup>:

<sup>78</sup> Sorensen [304], 20; Duhem [94], 301.

<sup>79</sup> Diese Interpretation ist höchst zweifelhaft. Im Briefwechsel zwischen Mersenne und Jean Rey, auf den Sorensen verweist, erläutert Mersenne seine Position ausschließlich mit der Keplerschen Variante, während Rey in seinem Antwortschreiben auf der Gilbertschen Variante beharrt. Die Aristotelische Theorie wird gar nicht erwähnt. Die Gedankenexperimente

Suppose a man on the moon throws a terrestrial rock and lunar rock toward earth. On Aristotle's principle, both rocks travel to earth. On Gilbert's, the earth rock travels to earth while the moon rock falls back to the moon. On Kepler's, both rocks fall to the moon. (Sorensen [304], 21)

Auch hier sei das Gedankenexperiment nicht als argumentative Stützung oder Widerlegung einer der fraglichen Positionen zu verstehen, sondern als Erläuterung ihres jeweiligen Gehalts. Betrachten wir noch ein jüngeres Beispiel.

Der Unterschied zwischen der Allgemeinen Relativitätstheorie und der Newtonschen Gravitationstheorie wird häufig mit einem Gedankenexperiment erläutert. Man stelle sich vor, die Sonne verschwinde plötzlich aus dem Universum. Was würde passieren? Nach Newtonscher Auffassung würde die Erde sofort ihre Bahn verlassen, da die Gravitationskraft der Sonne aufhören würde auf die Erde zu wirken. Das Sonnenlicht, das mit Lichtgeschwindigkeit zur Erde unterwegs ist, würde allerdings die Erde (bzw. die Stelle, an der sie sich vorher befand) auch dann noch eine Zeitlang erleuchten (etwa 8 Minuten), nachdem die Sonne bereits aufgehört hat zu existieren.

Nach der Allgemeinen Relativitätstheorie hingegen würde die Erde sich noch genauso lange auf ihrer Bahn halten, wie auch das Sonnenlicht auf sie trifft, da sich nach allgemeiner Relativitätstheorie auch der Einfluss der Sonne auf die Bewegung der Erde nur mit Lichtgeschwindigkeit ausbreitet. Es bliebe also etwa 8 Minuten hell und genauso lange bliebe die Erde auch auf ihrer Bahn.<sup>80</sup>

In diesen Beispielen wird das Gedankenexperiment nicht argumentativ verwendet, in dem Sinne, dass ein bestimmtes Naturgesetz als wahre oder falsche Beschreibung der Wirklichkeit erwiesen werden soll, sondern die Gedankenexperimente dienen lediglich der Erläuterung des Gehalts bestimmter Naturgesetze anhand eines Beispiels.

Prinzipiell finden sich solche „theoretischen“ Gedankenexperimente auch als Erläuterungen in Physikbüchern:

Wir können die Gleichheit von Schwere und Trägheit auch durch ein Gedankenexperiment veranschaulichen, bei dem diese beiden Grundeigenschaften der Körper ihre Rollen vertauschen. Angenommen, es befinde sich jemand eines Morgens beim Erwachen in einem geschlossenen Raume, unserer Erde entrückt, fern von allen Himmelskörpern und frei von jeder Massenanziehung im Weltall schwebend. (Brunner [43], 440)

Dies ist beispielsweise der Anfang einer Schilderung des Einsteinschen Aufzug-Gedankenexperiments. Im Physiklehrbuch wird mit diesem Gedankenexperiment ein Sachverhalt „veranschaulicht“, nicht aber argumentativ gestützt.

---

scheinen außerdem durchaus argumentativ verwendet zu werden. Da diese Episode hier nur erläuternden Charakter hat, wollen wir dies hier nicht weiter verfolgen. Der geneigte Leser betrachte diese Stelle als klärendes Gedankenexperiment zur Erläuterung der Funktionsweise klärender Gedankenexperimente. Vgl. Duhem [94], 319-320.

<sup>80</sup> Z.B. Green [132], 56.

In diese Rubrik fallen aber auch Textaufgaben, bei denen ebenfalls unter kontrafaktischen Annahmen („[...] wenn wir vom Luftwiderstand absehen.“, „[...] wenn wir die Reibung außer Acht lassen.“) die empirischen Konsequenzen einer Theorie abgeleitet werden.

Sofern solchen Gedankenexperimenten eine besondere Rolle bei der Vermittlung naturwissenschaftlichen Wissen zu- oder abgesprochen wird, fällt dieses Thema in die Didaktik und damit in die empirische Psychologie.<sup>81</sup>

### 3.1.2 FUNKTIONALE GEDANKENEXPERIMENTE

Neben Gedankenexperimenten, die hauptsächlich der Explikation einer Theorie (gegebenenfalls im Vergleich mit anderen Theorien) dienen, gibt es noch eine weitere Gruppe von „Gedankenexperimenten“, die weder eine Theorie stützen noch schwächen sollen. Diese Gedankenexperimente nehmen vielmehr eine funktionale Rolle *innerhalb* einer Theorie ein. Ein Beispiel für ein funktionales Gedankenexperiment ist die Verwendung bestimmter kontrafaktischer Annahmen bei der Anwendung von Test-Theorien zur Fehlerkorrektur statistischer Daten.

In der „objektiven“ Wahrscheinlichkeitstheorie werden Wahrscheinlichkeiten als Häufigkeitswerte interpretiert. D.h. Aussagen, wie beispielsweise

Dass bei einem Würfelwurf die ‚3‘ oben liegt, hat die Wahrscheinlichkeit 1/6.

werden folgendermaßen interpretiert:

Die relative Häufigkeit von ‚3‘ in einer Wurfserie konvergiert gegen 1/6.

Wahrscheinlichkeitsaussagen beziehen sich also immer auf Häufigkeitsgrenzwerte. Nimmt man an, dass die Würfelwürfe in einer Sequenz voneinander unabhängig sind, kann man davon ausgehen, dass immer größere Stichproben den Häufigkeitsgrenzwert immer besser annähern.<sup>82</sup> Unter derselben Annahme werden auch zufällige Messfehler bei der statistischen Datenerhebung behandelt:

The basic idea of classical test theory is that there exists a ‚true score‘, which is to be conceived of as the observed score, stripped of its random error. The true score  $T$  is thus defined as the observed score  $O$  minus the random error  $E$ . This idea leads to the classic equation  $O = T + E$ . Of course, this definition is empty unless some procedure is specified to define what error actually is. [...] The idea is that, if we take measurements on many occasions, it is plausible to define the true score as the ex-

---

<sup>81</sup> Das heißt nicht, dass es keine logischen Kriterien für gute oder schlechte klärende Gedankenexperimente gibt. Für einen Theorienvergleich muss ein Gedankenexperiment beispielsweise einen Sachverhalt repräsentieren, der als Modell aller zum Vergleich herangezogenen Theorien betrachtet werden kann. Textaufgaben müssen sich berechnen lassen, etc.

<sup>82</sup> Vgl. Schurz [290], 183; Hacking [135].

peptation of the observed scores over repeated measurements, so that  $T = E(O)$ . (Borsbroom, Mellenbergh und Van Heerden [32], 382)

Bei der Anwendung statistischer Verfahren in psychologischen oder medizinischen Untersuchungen sind die Voraussetzungen für diese Überlegung aber nicht gegeben. Würde man tatsächlich Messungen an Subjekten wiederholt durchführen, könnte man nicht davon ausgehen, dass die einzelnen Messergebnisse voneinander unabhängig blieben. Die Versuchspersonen lernen, werden müde, gewöhnen sich an das Testverfahren, etc. Um dennoch eine Häufigkeitsinterpretation der Wahrscheinlichkeit anwenden zu können, bedient sich die Testtheorie eines Gedankenexperiments:

Suppose we ask an individual, Mr. Brown, repeatedly whether he is in favor of the United Nations; suppose further that after each question we “wash his brains” and ask him the same question again. Because Mr. Brown is not certain as how he feels about the United Nations, he will sometimes give a favorable answer and sometimes an unfavorable answer. Having gone through this procedure many times, we then compute the proportion of times Mr. Brown was in favor of the United Nations. (Lazarsfeld [180], 493–494, zitiert nach Lord und Novick [190], 29)

Hierbei handelt es sich um ein Gedankenexperiment, das ‚true score‘ für die psychologische Anwendung statistischer Methoden *semantisch* interpretiert. Es geht dabei nicht darum, ob  $T$  eine messbare Größe darstellt. Durch dieses „Gedankenexperiment“ wird die Syntax der Wahrscheinlichkeitstheorie (insbesondere der Testtheorie) so interpretiert, dass sie auf den psychologischen Fall anwendbar wird.

This type of thought experiment may best be characterized as functional: A functional thought experiment is not aimed at refuting or supporting a theory, but has a specific function within a theory. In the case of frequentist statistics, it functions as a semantic bridge, providing a real world interpretation for the abstract syntax of probability. (Borsbroom, Mellenbergh und Van Heerden [32], 384)

Durch ein solches Gedankenexperiment wird also die Menge der intendierten Modelle einer Theorie erweitert. Obschon Borsboom, Mellenbergh und Van Heerden solche funktionalen Gedankenexperimente nur in den *Life Sciences* zu finden glauben, hat Michael Stöltzner sie auch in der Physik identifizieren können:

Contrary to the authors’ [Borsboom, Mellenbergh, Van Heerden] assumption, there exist functional thought experiments within modern physics. General relativity and quantum mechanics have taught us that the theory of measurement can be part and parcel of the theory itself; already measuring a magnetic field by a charged test body can count as a functional thought experiments – though one of restricted interest. (Stöltzner [321], 246)

Ob man solche hypothetische Konstruktionen letztlich als ‚Gedankenexperiment‘ bezeichnen möchte, ist Geschmackssache. Die Frage, welche Adäquatheitskrite-

rien für funktionale Gedankenexperimente gelten, ist in der Wissenschaftstheorie bisher noch nicht hinreichend untersucht.

### 3.1.3 GEDANKENEXPERIMENTE ZUR ÜBERZEUGUNGSÄNDERUNG

Den beiden soeben betrachteten Arten von Gedankenexperimenten ist gemeinsam, dass mit ihnen *nicht* der Anspruch erhoben wird, neue Erkenntnisse über die physikalische Wirklichkeit zu verschaffen, und nicht der Zweck verfolgt wird, im Adressaten des Gedankenexperiments gerechtfertigte Überzeugungsänderungen auszulösen. Letzteres ist von manchen Autoren allerdings für eine Gemeinsamkeit aller paradigmatischen Gedankenexperimente erachtet worden.<sup>83</sup> In den letzten beiden Unterkapiteln haben wir gesehen, dass offenbar nicht alle (so genannten) Gedankenexperimente diese Funktion haben. Im historischen Überblick von Teil 2 haben wir jedoch auch gesehen, dass Gedankenexperimente, die eine rationale Überzeugungsänderung herbeiführen sollen, diejenigen Gedankenexperimente sind, die das besondere Interesse der Wissenschaftstheorie auf sich gezogen haben. Ihnen müssen wir im Folgenden also unsere besondere Aufmerksamkeit widmen, wenn es darum geht, eine allgemeine Theorie über die methodologische Rolle von Gedankenexperimenten zu entwickeln.

---

<sup>83</sup> Vgl. Gähde [113].

### 3.2 GEDANKENEXPERIMENTE ZUR ÜBERZEUGUNGSÄNDERUNG

An oben beschriebener Zielsetzung von Gedankenexperimenten, gerechtfertigte Überzeugungsänderungen beim Adressaten zu bewirken, sind zunächst mindestens zwei Aspekte erklärungsbedürftig:

- (i.) Wie können Gedankenexperimente zu Überzeugungsänderungen führen?
- (ii.) Wann sind diese Überzeugungsänderungen *gerechtfertigt*?

Die erste Frage bezieht ihre Brisanz aus der Tatsache, dass Gedankenexperimente im Lehnstuhl durchgeführt werden: „[T]he information you have leaving the armchair is the same as the information you had when you sat down. So how can you be better off?“<sup>84</sup> Ohne dass man neue Informationen erhält, scheint man keinen Anlass zu haben, sein Überzeugungssystem einer Revision zu unterziehen. Wie sollen Gedankenexperimente also funktionieren?

Die zweite Frage betrifft die Tatsache, dass die Überzeugungen, die im Gedankenexperiment etabliert werden, kontingente Sachverhalte betreffen, über die man nur aus der Erfahrung unterrichtet sein kann. Wenn es also zu Überzeugungsänderungen in Bezug auf solche Sachverhalte kommt, ohne dass neue Erfahrungen gemacht wurden, wieso sollten diese Überzeugungsänderungen gerechtfertigt sein?

#### 3.2.1 PSYCHOLOGISCHE MODELLE DES „LABORATORY OF THE MIND“

Die Brisanz der ersten Frage kann zum Teil dadurch abgeschwächt werden, indem man darauf hinweist, dass *manche* Gedankenexperimente tatsächlich neue Daten erheben, weil sie propositionale Einstellungen testen. In den Sozialwissenschaften und der Psychologie werden allgemeine Einstellungen von Testpersonen (beispielsweise in Bezug auf Normen) häufig dadurch abgefragt, dass man den Befragten so genannte „Vignetten“ vorlegt: kurze, konkrete, hypothetische Situationsbeschreibungen, die dann durch die Befragten in irgendeiner Hinsicht (beispielsweise normativ) beurteilt werden müssen. Solche Daten sind von unmittelbarer Relevanz für Theorien, die über die Einstellungen der Testpersonen Aussagen machen (also typischerweise außerhalb der Naturwissenschaften) und von mittelbarer Relevanz für Argumente *ad auctoritatem*.<sup>85</sup>

Letzteres ist in all jenen Fällen von Interesse, in denen es ausgewiesene Experten in einem Bereich gibt, deren Überzeugungen über diesen Bereich wertvolle Informationen darstellen.

---

<sup>84</sup> Sorensen [304], 76.

<sup>85</sup> Vgl. Sorensen [304], 76-86.

Es ist aber auch darüber spekuliert worden, was ein positives Modell der Lehnstuhlforschung sein könnte. So könnten Gedankenexperimente als *Erinnerungshilfe* fungieren:

Many thought experiments function as reminders. A standard physics thought experiment features a pilot who forgets to fasten his seat belt when performing a front-back loop. At the top of the loop, the pilot is upside down. Does he fall out of the plane? One is initially inclined to answer 'yes' on the grounds that gravity will pull him down. But since you figure that the question would not have been asked if the answer were so straightforward, you check for forces that might counteract gravity. Many people are led to think of centrifugal force because the hypothetical stunt reminds them of carnival rides. They increase the similarity by imagining the pilot doing a left-right loop. The pilot's feet would press against the floor, he would not be leaning just on his side. This comparison leads to a negative answer for the vertical loop case: centrifugal force would keep the pilot in the plane. Sometimes there is a follow-up question: does the speed of the plane matter? An affirmative answer is obtained by considering a sequence of slower and slower loops until the situation becomes equivalent to past ones in which you slowly turn a container upside down thereby dumping the contents. (Sorensen [304], 89-90)

Wir rufen in uns durch Gedankenexperimente systematisch Erinnerungen wach, die uns vielleicht bis dahin unzugänglich waren. Wie auch in anderen Mnemotechniken (beispielsweise der *loci*-Methode, bei der durch Assoziation unser räumliches Vorstellungsvermögen genutzt wird, um (typischerweise nicht-räumliche) Inhalte und ihre Ordnung besser erinnern zu können) wird im Gedankenexperiment häufig mit visuellen Vorstellungen operiert, was eine funktionale Verwandtschaft nahe legt.

Another similarity between thought experiments and the method of loci is the importance of making the main characters few but salient, the background familiar and unassuming, and the conditions of observation as normal as possible. (Sorensen [304], 92)

Ein anderes Modell betrachtet Gedankenexperimente als *Wissenstransformator*, bei dem *knowing how* in *knowing that* umgewandelt wird.<sup>86</sup> Dieses Modell wird uns später noch häufiger begegnen, da es in sinnvoller Weise auf viele philosophische Gedankenexperimente anwendbar ist, insbesondere wenn unsere Überzeugungen über Sprache zur Debatte stehen. Dieses Modell kann aber nur unzureichend erklären, was für eine Art *knowing how* in physikalischen Gedankenexperimenten umgewandelt wird.

Wie im Wissenstransformationsmodell kann neues Wissen dadurch entstehen, dass ein Teil unseres Verstandes etwas über einen anderen Teil unseres Verstandes lernt. Im *homunculi-Modell des Geistes* kann dies auch durch andere Weisen geschehen als bloß durch die Transformation von sprachlichem *knowing how* in

---

<sup>86</sup> Sorensen [304], 92-95. Diese Dichotomie lässt sich bekanntlich im Deutschen nicht durch äquivalente Ausdrücke wiedergeben.

*knowing that*. Hierbei wird davon ausgegangen, dass unser Geist aus Teilen besteht, die alle über einen gewissen Schatz an propositionalem Wissen verfügen, der für die spezifischen Aufgaben dieses Teils nötig ist.<sup>87</sup> Ein Gedankenexperiment führt dazu, dass Informationen von einem „homunculus“ zum anderen transportiert werden, der die Information anders verarbeiten kann, oder über Informationen verfügt, die mit jener fruchtbar in Verbindung gebracht werden können.

Ähnlich spekulativ ist die Auffassung, dass Gedankenexperimente die zur Verfügung stehenden Informationen so aufbereiten, dass sie besser durch den Verstand zu verarbeiten sind.<sup>88</sup> So fällt es uns manchmal schwer, auf einen konkreten Fall die relevanten Gesetzmäßigkeiten oder Prinzipien anzuwenden, während wir bei einem abstrakten, aber in allen relevanten Hinsichten analogen Fall diese Probleme nicht haben.

Die bisher am weitesten ausbuchstabierte psychologische Theorie zur Funktionsweise von Gedankenexperimenten betrachtet sie als dynamische mentale Modelle im Geist des Gedankenexperimentierers, die ausgehend von der narrativen Beschreibung des Gedankenexperiments einen Experimentablauf mental simulieren, wobei verschiedene Informationen (Propositionen, Modelle, Gleichungen) integriert werden. Diese Simulationen werden im Gehirn zum Teil durch den visuellen Kortex durchgeführt. Auf diese Weise lassen sich aus Informationen, die in narrativer Form präsentiert werden, schneller Schlussfolgerungen ziehen als durch Ableitungsverfahren aus einem Argument in propositionaler Form.<sup>89</sup>

The main idea centers on the fact that perceptual inferences are easy to make. By clustering connected information and making visual a chain of interconnected inferences the imagistic representations support a large number of immediate perceptual inferences. (Nersessian [223], 24)

Diese Auffassung kann gewisse Eigenschaften von Gedankenexperimenten erklären, wie beispielsweise, dass sie oft mit scheinbar überflüssigen Details ausgeschmückt sind, und warum bevorzugt Gedankenexperimente an Stelle ausführlicher Argumente benutzt zu werden scheinen.

Unter einem mentalen Modell versteht man dabei eine mentale Repräsentation besonderer Art:

[A] mental model is a structural analog of a real world or imaginary situation, event, or process that the mind constructs in reasoning. What it means for a mental model to be a structural analog is that it embodies a representation of the spatial and temporal relations among, and the causal structures connecting the events and entities depicted and whatever other information that is relevant to the problem solving task. (Nersessian [225], 11)

---

<sup>87</sup> Sorensen [304], 95-99; Dennett [85].

<sup>88</sup> Sorensen [304], 99-104.

<sup>89</sup> Vgl. Craik [79] für die Herkunft dieser Idee, Gooding [128], Nersessian [224], Mišević [210], Johnson-Laird [159] für die Ausarbeitung.

Die Fähigkeit, solche mentalen Modelle zu bilden, mag sich evolutionär durch den Vorteil geplanten Handelns gebildet haben. Es wird davon ausgegangen, dass die Informationen, die sich in solchen mentalen Modellen repräsentieren lassen, nicht auf eine bestimmte, evolutionär erworbene Menge beschränkt sind, sondern durch Lernen erweitert werden können. Dadurch ist zumindest die theoretische Möglichkeit adäquater mentaler Modelle auch für solche Situationen gegeben, die im Evolutionsprozess keine Relevanz besaßen. Gedankenexperimente nutzen diese Fähigkeit zur mentalen Modellierung und verbinden sie mit der Fähigkeit zur mentalen Simulation dynamischer Prozesse.

Diese psychologischen Spekulationen, die noch sehr in den Kinderschuhen stecken, sind durchaus interessant zum Verständnis der psychologischen Funktionsweise von Gedankenexperimenten, sind aber relativ uninformativ, wenn es um Fragen der methodologischen Funktionsweise von Gedankenexperimenten geht.<sup>90</sup> Interessant sind hier bestenfalls die Spekulationen darüber, *warum* mentale Simulationen physikalischer Prozesse zuverlässige Auskunft über das tatsächliche Verhalten physikalischer Prozesse geben sollten. Hier wäre aber zunächst zu klären, *ob* sie dies überhaupt tun. Auf diese Frage werden wir noch zurückkommen.

---

<sup>90</sup> Zu einer ähnlichen Schlussfolgerung kommt auch Häggqvist [137], 85: „The mental-model hypothesis is best viewed, it seems, as a theory about what goes on in the brain when someone performs a thought experiment. But such a theory has little normative weight. The normative question is: Does the argument whose premises are obtained from the thought experiment justify its conclusion? If they do, they do so regardless of how the thought experiment is implemented in the brain. And whether they justify the conclusion does not turn on whether the psychological mechanism involves mental models or not.“ Vgl. aber auch den nächsten Satz und den Abschnitt zur evolutionären Erkenntnistheorie in diesem Kapitel. Häggqvist scheint hier das Kind mit dem Bade auszuschütten. Wenn wir feststellen könnten, dass Gedankenexperimente auf einem *reliablen* kognitiven Prozess beruhen, könnte das sehr wohl die normative Frage beantworten. Diese Frage legen sich die Vertreter des Mentale-Modelle-Ansatzes allerdings tatsächlich nie vor (vgl. etwa Misselhorn [211] oder Mišćević [210]). Die Erfolgsgeschichte von *mental modelling* erschöpft sich meistens darin, dass man mit dieser Methode sehr gut ausknobeln kann, ob sich das neue Sofa besser an dieser oder jener Wohnzimmerwand machen wird.

### 3.2.2 GEDANKENEXPERIMENTE ALS PARADOXIEN

Neben diesen psychologischen Versuchen, die Funktionsweise von Gedankenexperimenten zu erklären, gibt es die konservativere Variante, den Sinn von (zumindest einigen) Gedankenexperimenten im Aufweis von Paradoxien zu sehen. Psychologisch würde ihre Rolle bei der Überzeugungsänderung dann etwa durch die Theorie der kognitiven Dissonanz, oder eine ähnliche Theorie erklärt.<sup>91</sup>

In der Theorie der kognitiven Dissonanz ist eine zentrale Hypothese, dass Personen durch kognitive Dissonanzen motiviert sind, diese Dissonanzen zu reduzieren. Unter kognitiver Dissonanz versteht man dabei eine Relation über Denkinhalten einer Person, wobei Dissonanz zwischen zwei kognitiven Elementen (bzw. Denkinhalten)  $p$  und  $q$  genau dann besteht, wenn die Person glaubt, dass  $p$ , glaubt, dass  $q$  und glaubt, dass nicht- $p$  aus  $q$  „objektiv“ folgt.<sup>92</sup>

Eine solche Theorie wäre das direkte psychologische Supplement zu Roy Sorensens „Cleansing Model“ der Funktionsweise von Gedankenexperimenten. Nach diesem Modell sollen Gedankenexperimente in erster Linie Inkonsistenzen im Überzeugungssystem aufweisen und dazu motivieren, das Überzeugungssystem auf konsistente Weise umzubauen. Sorensen erhebt dabei nicht den Anspruch, *alle* Funktionsweisen von Gedankenexperimenten zu erfassen, sondern nur eine wichtige Teilklass. Nach Sorensens Analyse lassen sich diese Gedankenexperimente auf eine von zwei Weisen als *reductio*-Argument rekonstruieren, entweder als „Necessity Refuter“ oder als „Possibility Refuter“.

#### 3.2.2.1 Necessity Refuter

Bei einem *necessity refuter*-Gedankenexperiment (NRG) soll nachgewiesen werden, dass eine bestimmte Überzeugung revisionsbedürftig ist, da sie etwas als unmöglich ausschließt, was aber tatsächlich möglich ist. In naturwissenschaftlichen Zusammenhängen denkt Sorensen dabei hauptsächlich an Naturgesetze oder Naturprinzipien, die bestimmte Sachverhalte als naturunmöglich ausschließen.

NRGs lassen sich am Besten in ihrem argumentativen Kontext charakterisieren, zumal NRGs ja dadurch definiert sind, was sie *zeigen* sollen. Ein solcher Argumentationskontext wird von Sorensen als ein Quintett zusammengenommen inkonsistenter Aussagen rekonstruiert:

- (N1) S „*Modal source statement*“: Hierzu gehören zunächst Gesetzesaussagen, Sorensen erweitert sein Modell aber auf semantische Thesen

<sup>91</sup> Festinger [104].

<sup>92</sup> Vgl. Bühler [44], 32. Ein Satz  $p$  folgt in diesem Sinne „objektiv“ aus einem Satz  $q$ , wenn er logisch aus  $q$  und dem Hintergrundwissen der Person folgt, oder – im Fall von Handlungsbeschreibungen und normativen Sätzen –  $p$  mit  $q$  deontisch konkordant ist. Zu den Voraussetzungen einer kognitiven Dissonanztheorie gehört, dass Subjekte zumindest scheinbar widersprüchliche Überzeugungen haben können.

(Realdefinitionen, Synonymiebehauptungen, Behauptungen über Implikationsbeziehungen), Thesen zur Überprüfbarkeit (nicht-Verifizierbarkeit, nicht-Falsifizierbarkeit, Unbeobachtbarkeit), Durchführbarkeitsbehauptungen, Zuschreibungen von Intentionen und Dispositionen, Gültigkeitsurteile und Theorien als Mengen solcher Aussagen.

- (N2)  $S \supset \Box I$  „*Modal Extractor*“: In diesem Schritt wird die relevante modale Implikation aus  $S$  abgeleitet.
- (N3)  $(I \wedge C) \Box \rightarrow W$  „*Counterfactual*“: Diese Aussage ist ein irrealer Konditionalsatz, welcher besagt, dass wenn  $I$  (die Implikation aus  $S$ ) und  $C$  (eine vorgestellte Situation) der Fall wären,  $W$  („weird consequence“) in dieser Situation ebenfalls der Fall wäre.<sup>93</sup>
- (N4)  $\neg \Diamond W$  „*Absurdity*“: die durch  $W$  beschriebene Situation wird zur Unmöglichkeit erklärt.
- (N5)  $\Diamond C$  „*Content Possibility*“: Durch diese Aussage wird schließlich behauptet, dass der Gehalt des Gedankenexperiments, also die durch  $C$  geschilderte Situation *möglich* ist.

Als Beispiel für eine solche Argumentation könnte die Standardinterpretation<sup>94</sup> des Galileischen Gedankenexperiments gegen das Aristotelische Fallgesetz dienen.

Der *Modal Extractor* entspräche etwa einem Nachweis, dass nach den aristotelischen Fallgesetzen schwere Körper eine Fallgeschwindigkeit erreichen müssen, die proportional zu ihrem Gewicht ist. Das *Counterfactual* bestünde in der Aussage, dass unter der Annahme dieser Gesetzesaussage zwei miteinander starr verbundene, verschieden schwere Objekte sowohl schneller als auch langsamer fallen als das schwerere von ihnen alleine fallen würde. Die *Absurdity* würde darauf hinweisen, dass ein solcher Umstand unmöglich ist, die *Content Possibility* in der Behauptung bestehen, dass es möglich ist, zwei verschieden schwere Körper starr zu verbinden und fallen zu lassen.

Wie wir sehen, wird der Argumentationskontext von Sorensen als modaler Kontext aufgefasst. Wir werden auf Modalitäten in diesem und den folgenden Kapiteln noch vieles sagen müssen, an dieser Stelle muss aber zunächst erklärt werden, weshalb in den Naturwissenschaften modale Überlegungen eine Rolle spielen sollen.

<sup>93</sup> Irreale Konditionalsätze werden hier wie in Lewis [185] als Sätze angesehen, die entgegen dem vermeintlichen umgangssprachliche Gebrauch nicht *bedeuten*, dass der „Sprecher eine Gegenposition zur Realitätserfahrung“ bezieht. Vgl. Helbig und Buscha [141], 202 u. 525. Es soll nicht automatisch *impliziert* sein, dass das Vorderglied in der aktualen Welt nicht der Fall ist, war oder sein wird.

<sup>94</sup> Unter ‚Standardinterpretation‘ verstehe ich hier die Interpretationen, die etwa von Popper [260] oder Brown [38] vorgeschlagen wurden.

In der Wissenschaftstheorie der Naturwissenschaften trifft man häufig auf die Auffassung, dass Naturgesetze *kontingent* seien, also gerade keine Notwendigkeiten ausdrücken. Dies soll in der Regel zum Ausdruck bringen, dass die Naturgesetze nicht schon aus logischen Gründen alleine gelten, es also mit der Logik durchaus vereinbar wäre, würden in unserer Welt andere Naturgesetze gelten.

Wenn Sorensen (und andere Autoren) von Naturgesetzen als „Notwendigkeiten“ reden, wollen sie diese Auffassung häufig gar nicht in Frage stellen, sondern darauf hinaus, dass Naturgesetze ebenfalls einen modalen Charakter haben, der von ihrer logischen Kontingenz unabhängig sein kann.<sup>95</sup> Man spricht dann häufig von „Naturnotwendigkeit“ oder „nomischer Notwendigkeit“<sup>96</sup>. Die „Gesetzesartigkeit“ wahrer Verallgemeinerungen erkennt man gerade daran, dass man sie auf ihren modalen Charakter untersucht.<sup>97</sup>

(N<sub>1</sub>) Alle Äpfel in diesem Korb sind rot.

(N<sub>2</sub>) Alle freien Körper in der Nähe der Erdoberfläche fallen mit Beschleunigung *g* nach unten.

Diese beiden wahren Verallgemeinerungen unterscheiden sich dadurch, dass (N<sub>1</sub>) die kontrafaktische Aussage ‚Wenn dieser (grüne) Apfel in diesem Korb wäre, wäre er rot.‘ nicht unterstützt, während (N<sub>2</sub>) die kontrafaktische Aussage ‚Wenn dieser Gegenstand (auf meinem Schreibtisch) frei wäre (und in der Nähe der Erdoberfläche), würde er mit Beschleunigung *g* nach unten fallen.‘ sehr wohl unterstützt.

Der modale Charakter von Naturgesetzen geht aber weiter. Naturgesetze machen auch über solche Sachverhalte Aussagen, die (z.T. aus prinzipiellen Gründen) in der aktuellen Welt nicht realisiert sind oder nicht realisiert sein können, also nur „in möglichen Welten“ der Fall sind. So erlauben Naturgesetze beispielsweise Prognosen über das Verhalten von Körpern unter vollständiger Reibungsfreiheit, obwohl Reibungsfreiheit nur angenähert realisiert werden kann. Nach Sorensens Analyse werden in Gedankenexperimenten die modalen Konsequenzen von Naturgesetzen (oder anderen Aussagen) nun mit anderen modalen Tatsachen in Konflikt gebracht. Schließlich präsentiert ein Gedankenexperiment keinen wirklichen Fall, der gegen eine Theorie oder Überzeugung sprechen würde (so wie ein reales Experiment dies tut) sondern einen möglichen Fall, der mit dem modalen Charakter der fraglichen Überzeugung inkonsistent ist.

---

<sup>95</sup> Vgl. aber Bigelow, Ellis und Lierse [22].

<sup>96</sup> Manchmal wird Naturnotwendigkeit von nomischer Notwendigkeit unterschieden. Der Unterschied bestünde darin, dass nomische Notwendigkeit eine Varianz in den Anfangsbedingungen zulässt. Das ist in der Regel auch mit ‚naturnotwendig‘ gemeint, da sonst Naturnotwendigkeit in deterministischer Auffassung nur eine mögliche Welt zulassen würde, die aktuelle.

<sup>97</sup> Vgl. Schurz [290], 388.

### 3.2.2.2 Possibility Refuter

Bei einem *possibility refuter*-Gedankenexperiment (PRG) soll ebenfalls nachgewiesen werden, dass eine bestimmte Überzeugung revisionsbedürftig ist. In diesem Fall soll allerdings nachgewiesen werden, dass die fragliche Überzeugung etwas als möglich erscheinen lässt, was tatsächlich unmöglich ist.

- (P1)  $S$
- (P2)  $S \supset \diamond I$  „*Possibility Extractor*“: In diesem Schritt wird die relevante modale Implikation aus  $S$  abgeleitet, im PRG eine Möglichkeitsaussage.
- (P3)  $(I \wedge C) \Box \rightarrow W$
- (P4)  $\neg \diamond W$
- (P5)  $\diamond I \supset \diamond (I \wedge C)$  „*Content Copossibility*“: Durch diese Aussage wird schließlich behauptet, dass die Aussage, die in (P2) aus dem *Modal Source Statement* hergeleitet wurde, nur dann wahr ist, wenn sie mit dem Gehalt des Gedankenexperiments vereinbar ist.

Als Beispiel für eine solche Argumentation könnte man Sorensens Interpretation des Machschen Arguments für die Transitivität der Gleichheit von Massen heranziehen. Als *Modal Source Statement* käme jede Theorie in Frage, die impliziert, dass, wenn zwei Massen einer dritten gleich sind, sie *nicht* untereinander gleich sein müssen.<sup>98</sup> Im *Possibility Extractor* würde ebendiese Implikationsbeziehung behauptet.  $C$  entspricht der Möglichkeit, dass drei elastische Körper,  $A$ ,  $B$  und  $C$ , sich frei um einen reibungsfreien starren Ring bewegen. Würde auf diese drei Körper zutreffen, dass die Masse von  $A$  gleich der Masse von  $B$  wäre und die Masse von  $B$  gleich der Masse von  $C$ , die von  $C$  aber größer als die Masse von  $A$ , dann müsste, wenn wir  $A$  eine bestimmte Geschwindigkeit geben, diese bei Kollision an  $B$  übergeben werden, diese wiederum bei Kollision an  $C$  übergeben werden. Wenn  $C$  jetzt aber wieder mit  $A$  kollidiert, sollte  $A$  aufgrund  $C$ s größerer Masse eine zusätzliche Beschleunigung erfahren, so dass sich die drei Körper bei fort-

---

<sup>98</sup> Man beachte, dass es eigentlich nötig wäre genauer anzugeben, was es heißen soll, dass eine Möglichkeit durch eine Theorie „impliziert“ wird. Dass eine Theorie eine Möglichkeit nur erlaubt, kann nicht gemeint sein: Man betrachte eine Axiomatisierung einer physikalischen Theorie mit  $n$  Axiomen. Die Axiome  $A_1$  und  $A_2$  seien logisch unabhängig, dann erlaubt  $A_1$  natürlich  $\diamond \neg A_2$ , in dem Sinne, dass  $\neg A_2$  nicht *nur* aufgrund von  $A_1$  bereits als Möglichkeit ausgeschlossen ist.  $\diamond \neg A_2$  wird aber durch  $A_2$  als physikalisch unmöglich ausgeschlossen, es lässt sich also eine mögliche Welt finden, die physikalisch unmöglich ist, aber in einem schwachen Sinne von  $A_1$  „impliziert“ wird, wenn eine Möglichkeit dann impliziert wird, wenn sie nicht logisch ausgeschlossen ist. Das kann nicht gemeint sein. (Ansonsten könnte jede Theorie durch irgendeine logisch unabhängige Theorie mit obigem Argument falsifiziert werden. Glücklicherweise werden wir uns in wenigen Abschnitten von PRGs verabschieden, weshalb wir diese Frage offen lassen können.)

gesetzten Umdrehungen von Runde zu Runde schneller um den Ring bewegen müssten (was  $W$  entspricht). Diese unendlich zunehmende Verstärkung widerspricht aber dem ersten Gesetz der Thermodynamik,  $W$  ist also physikalisch unmöglich.<sup>99</sup> *Content Copossibility* entspricht der Behauptung, dass die kontrafaktische Situation möglich sein muss, wenn  $I$  möglich ist.

### 3.2.2.3 Die Interpretation der Modaloperatoren in der Argumentrekonstruktion

Die Modaloperatoren erhalten hier und im Folgenden die übliche Interpretation, » $\Box\phi$ « ist wahr, gdw. » $\phi$ « in allen möglichen Welten wahr ist, die von der aktuellen Welt aus erreichbar<sup>100</sup> sind. » $\Diamond\phi$ « ist wahr, gdw. » $\phi$ « in mindestens einer möglichen Welt wahr ist, die von der aktuellen Welt aus erreichbar ist. » $\phi \Box \rightarrow \psi$ « ist wahr in Bezug auf ein Sphärensystem von möglichen Welten  $S$ , gdw. entweder (a) keine  $\phi$ -Welt zu irgendeiner Sphäre  $S$  in  $S@$  gehört, wobei  $S@$  auf die aktuelle Welt zentriert ist, oder (b) es mindestens eine Sphäre  $S$  aus  $S@$  gibt, die mindestens eine  $\phi$ -Welt enthält und » $\phi \supset \psi$ « in allen Welten in  $S$  wahr ist.<sup>101</sup> Negation, Konjunktion und Konditional haben ihre klassische Bedeutung.

Damit es sich um ein gültiges modales Argument handelt, muss der Möglichkeitsoperator in (N4) höchstens, der in (N5) jedoch von mindestens derselben

<sup>99</sup> An dieser Stelle gibt Sorensen eine Missrekonstruktion von Machs eigentlichem Argument. Mach zieht nicht die Thermodynamik als Begründung heran. In diesem Fall könnte er sich die Dramatisierung zur unendlichen Beschleunigung auch sparen (wie Sorensen selbst bemerkt). Nach dem ersten Gesetz der Thermodynamik wäre ja die Beschleunigung nach dem ersten Umlauf schon unmöglich. Mach will stattdessen ein Gedankenexperiment im Stil Stevins hervorbringen, das eine Konklusion hat, die *offensichtlich* unserer Erfahrung widerspricht. Machs Argument eignet sich dann aber nicht mehr so gut zur Explikation von PRGs, weshalb wir Sorensens Rekonstruktion betrachtet haben. Vgl. Sorensen [304], 154-155 und Mach [193], 243-244.

<sup>100</sup> Was hierbei als „erreichbar“ zählt, hängt davon ab, welche Art von Modalität jeweils intendiert ist. Falls » $\Box\phi$ « bedeuten soll, dass » $\phi$ « physikalisch notwendig ist, dann sind nur diejenigen möglichen Welten von der aktuellen Welt aus erreichbar, in denen dieselben Naturgesetze gelten, wie in der aktuellen Welt.

<sup>101</sup> Vgl. Lewis [185]. Timothy Williamson gibt eine interessante alternative Analyse der Modaloperatoren (und diskutiert ihren Zusammenhang zum Gedankenexperiment) in Williamson [348]. Williamson weist auch darauf hin, dass die subjunktiven Konditionale nicht sinnvoll durch strikte Implikationen ersetzt werden können, weil wir nicht alle möglichen Faktoren ausschließen können, die unsere Beurteilung des möglichen Falls vielleicht noch ändern könnten. Subjunktive Konditionale (wie » $(I \wedge C) \Box \rightarrow W$ «) sind zwar schwächer als strikte Implikationen, aber deswegen noch lange nicht so schwach wie die bloße Behauptung » $\Diamond(I \wedge C \rightarrow W)$ «. Unter einer solchen Interpretation käme es zwischen » $(I \wedge C) \Box \rightarrow W$ « und » $\neg \Diamond W$ « sonst zu gar keinem Widerspruch (es kann ja sein, dass » $\Diamond(I \wedge C \rightarrow W)$ « nur in den Welten wahr ist, in denen das Vorderglied des Konditionals *und* das Hinterglied falsch sind). Subjunktive Konditionale (wie » $(I \wedge C) \Box \rightarrow W$ «) sind andererseits auch schwächer als » $\Diamond(I \wedge C) \rightarrow W$ «. Diese Interpretation macht Kühne in Kühne [176], S. 377. Eine solche Formel würde aber besagen, dass das Hinterglied aktual der Fall ist, wenn das Vorderglied möglich ist. Wenn der Satz ‚Wenn es regnen würde, wäre die Strasse nass.‘ und der Satz ‚Es ist möglich, dass es regnet.‘ wahr sind, kann es natürlich trotzdem sein, dass die Strasse aktual staubtrocken ist, weil es nicht regnet und auch sonst kein Wasser auf die Strasse gekommen ist.

Modalität sein, wie der von I in (N2).<sup>102</sup> D.h., falls es so ist, dass die physikalischen Möglichkeiten eine echte Teilmenge der logischen Möglichkeiten darstellen und I eine physikalische Theorie ist, die nicht schon logisch notwendig ist, muss C *mindestens* physikalisch möglich sein, um ein relevantes Gegenbeispiel abzugeben, W darf dabei *höchstens* physikalisch unmöglich sein.<sup>103</sup> Ebenso muss bei einem PRG der zweite Möglichkeitsoperator in (P5) mindestens von derselben Modalität sein wie der in (P2), bzw. im Vorderglied von (P5).

Sorensen nimmt daher an, dass die fraglichen Modalitäten durch die intendierte Lesart von S festgelegt sind, nicht aber unabhängig durch eine Menge von (unterspezifizierten) Hintergrundannahmen (wie Häggqvist es vorschlägt).<sup>104</sup>

### 3.2.3 GEDANKENEXPERIMENTE ALS ARGUMENTE GEGEN „TARGETTHESEN“

An obiger Regimentierung ist kritisiert worden, dass die Rolle von (N2) bzw. (P2) unklar bleibt. Ein NRG soll das Source Statement kritisieren, folglich muss S bereits modaler Natur sein. (N2) bzw. (P2) ist also trivialerweise wahr, oder die Regimentierung ist inadäquat. Aus diesem Grund ist von Sören Häggqvist (vgl. Häggqvist [137]) folgende alternative Regimentierung vorgeschlagen worden:

- (1) T
- (2)  $T \supset (C \Box \rightarrow W)$
- (3)  $\neg \Diamond W$
- (4)  $\Diamond C$

,T' steht hier für ‚Target-Thesis‘. Diese These ist selbst bereits eine modale Aussage.

Häggqvist schlägt noch weitere Modifikationen der Sorensenschen Regimentierung vor. Hierzu gehört zunächst der Vorschlag, die Form einer *Argumentrekon-*

<sup>102</sup> Unter der Annahme eines „Zwiebelmodells“ der Notwendigkeiten, bei dem die logischen Notwendigkeiten eine echte Teilklasse der begrifflichen Notwendigkeiten bilden, diese eine echte Teilklasse der metaphysischen, diese eine echte Teilklasse der natürlichen, etc. Probleme mit diesem Modell werden wir in Kapitel 5 besprechen.

<sup>103</sup> Wie wir im nächsten Kapitel sehen werden, sind die Dinge sogar etwas komplizierter. Ob die physikalischen Möglichkeiten eine echte Teilmenge der logischen Möglichkeiten bilden, sei dahingestellt. Zu diesen Begriffen vgl. die „Zwiebel“ der Möglichkeiten und Notwendigkeiten in Kapitel 5.2.

<sup>104</sup> Mir ist schon nicht klar, wie „unterspezifizierte Hintergrundannahmen“ überhaupt etwas „determinieren“ können. Der Unterschied zu Häggqvist besteht aber hauptsächlich darin, dass bei Sorensen kein Holismus *vorausgesetzt* wird. Falls der Holismus wahr ist, kollabieren vielleicht alle Arten von Notwendigkeit in eine. Aber selbst in diesem Fall ist alles bisher Gesagte zutreffend (dann sind die Möglichkeitsoperatoren eben relativ zu *dieser* Modalität zu interpretieren).

struktion anzupassen (also Prämissen und Konklusion auszuweisen) und in der Reihenfolge der Prämissen typische Argumentationen nachzuahmen. Ein weiterer Vorschlag besteht darin, die Unmöglichkeit von  $W$  nicht von der fraglichen Modalität, sondern von  $C$  abhängig zu machen. Häggqvists so revidiertes Schema lautet folgendermaßen:

- (1)  $\diamond C$
- (2)  $T \supset (C \Box \rightarrow W)$
- (3)  $C \Box \rightarrow \neg W$

---

- (4)  $\neg T$

Häggqvist geht davon aus, dass dieses Schema auch alle PRGs erfassen kann (da die Modalität von  $T$  in der Regimentierung nicht festgelegt ist). Um dieses Schema mit etwas Inhalt zu füllen, sei kurz ein weiteres physikalisches Beispiel betrachtet.<sup>105</sup>

*Beispiel: Maxwells Dämon*

Maxwells Dämon-Gedankenexperiment ist von Sorensen und Häggqvist als mögliches Beispiel der oben angegebenen Regimentierung vorgeschlagen worden, obwohl beide sich eine detaillierte Ausführung sparen.

Dieses Gedankenexperiment von James Clark Maxwell ist wissenschaftshistorisch betrachtet insbesondere deswegen interessant, weil es eine ganze Industrie von Dämon-Exorzisten auf den Plan gerufen hat, die bis in die Gegenwart versuchen, verschiedene Spielarten des Maxwellschen Dämons als unmöglich zu entlarven, ohne dass in allen (sogar in den meisten) Fällen überhaupt klar ist, was das Ziel dieses Exorzismus ist und nach welchen Regeln der Dämon ausgeschaltet werden darf. Dies gilt in besonderem Maße für die „informationstheoretische“ Behandlung des Problems.<sup>106</sup> Für Maxwells ursprüngliches Gedankenexperiment liegen die Dinge glücklicherweise anders.

Maxwell erwähnt das Dämon<sup>107</sup>-Gedankenexperiment zum ersten Mal in einem Brief an Tait vom 11. Dezember 1867. Im Gedankenexperiment soll sich Tait einen Behälter mit Gas vorstellen, der durch eine Scheidewand in zwei Sektionen, A und B unterteilt ist:

Now conceive a finite being who knows the paths and velocities of all the molecules by simple inspection but who can do no work except open and close a hole in the

<sup>105</sup> Weder Sorensen, noch Häggqvist machen sich wirklich die Mühe ein solches Beispiel zu analysieren, obwohl Häggqvist bemerkt, dass es gar nicht so „transparent“ ist, dass sich überhaupt eine kritische Masse von Gedankenexperimenten auf diese Weise rekonstruieren lässt.

<sup>106</sup> Vgl. Earman und Norton [96], Earman und Norton [97], Bremer und Cohnitz [34], 15-19; Carnap [52], Leff und Rex [181], Norton [242].

<sup>107</sup> Die Bezeichnung ‚Dämon‘ stammt übrigens nicht ursprünglich von Maxwell selbst, sondern von William Thomson (Lord Kelvin).

diaphragm by means of a slid without mass. Let him first observe the molecules in A and when he sees one coming the square of whose velocity is less than the mean sq. vel. of the molecules in B let him open the hole and let it go into B. Next let him watch for a molecule of B, the square of whose velocity is greater than the mean sq. vel. in A, and when it comes to the hole let him draw the slide and let it go into A, keeping the slide shut for all other molecules. (Knott [165], 214; zitiert nach Earman und Norton [96], 438)<sup>108</sup>

Nach der Rekonstruktion von John D. Norton und John Earman, sowie nach der Auffassung Sören Häggqvists richtet sich Maxwell mit diesem Gedankenexperiment gegen die „universale“ oder „klassische“ Auffassung vom zweiten Gesetz der Thermodynamik. Falls ein solcher Dämon möglich wäre, würde A wärmer und B kälter „and yet no work has been done, only the intelligence of a very observant and neat-fingered being has been employed.“ Diese Möglichkeit wird aber durch den zweiten Hauptsatz der universal interpretierten Thermodynamik gerade ausgeschlossen:

- (\*) Wärme kann nicht ohne Arbeit von einem kälteren zu einem wärmeren Körper übergehen.<sup>109</sup>

Der Target-These entspricht hierbei (\*), die imaginäre Situation, C, ist in obigem Zitat geschildert. Laut T sollte in dieser Situation ausgeschlossen sein, dass der Dämon durch seine Tätigkeit die Temperatur in A erhöht. Die Aussage ‚Dem Dämon gelingt es nicht die Temperatur in A zu erhöhen.‘ entspräche W. Es scheint, dass Maxwell W aufgrund derselben Theorie ausschließen kann, die er auch heranzieht, um die Möglichkeit von C zu etablieren:

Only a little reflection is needed to realise that the Demon is made possible by the form of the velocity distribution law which Maxwell announced in 1860 for a gas in equilibrium. For no matter how sharply peaked this bell-shaped distribution is, it has tails which extend infinitely far. The Demon operates by sorting out molecules whose velocities lie sufficiently far out in the tails from those whose velocities lie in the main lump. (Earman und Norton [96], 440)

Da nach Maxwells Theorie die Geschwindigkeiten der einzelnen Moleküle statistisch in Form einer Glockenkurve verteilt sind, findet der Dämon sowohl in A, wie auch in B immer Moleküle, bei denen das Quadrat der Geschwindigkeit hoch bzw. niedrig genug ist, um A gegenüber B aufzuwärmen. Daher kann der Dämon wie beschrieben operieren, C ist also möglich.

Dass der Dämon damit den zweiten Hauptsatz verletzt, liegt daran, dass nach derselben Theorie die Temperatur des Gases in A abhängig davon ansteigt, wie es

<sup>108</sup> In Maxwells Theorie ist die Temperatur eines Gases proportional zum Durchschnitt der Quadrate der Geschwindigkeiten der Moleküle.

<sup>109</sup> Vgl. etwa Riecke [280], 650. Diese Version des zweiten Hauptsatzes wird dort als ‚Satz von Clausius‘ bezeichnet. Zu weiteren Formulierungen des zweiten Hauptsatzes, wie er hier von Maxwell attackiert wird, vgl. Earman und Norton [96], 464 (Appendix I).

dem Dämon gelingt, schnellere Moleküle von B nach A und langsamere von A nach B zu bringen (W, d.h. das Ausbleiben eines Temperaturanstiegs von A gegenüber B, trotz der Sortieraktion des Dämon also unmöglich ist).

Legt man eine solche Theorie zugrunde, muss der zweite Hauptsatz abgeschwächt werden. Maxwell intendierte dabei eine statistische Reformulierung des zweiten Hauptsatzes: „the 2nd law of thermodynamics has the same degree of truth as the statement that if you throw a tumblerful of water into the sea, you cannot get the same tumblerful of water out again.“<sup>110</sup>

Dieses Gedankenexperiment lässt sich also in der Tat in der von Häggqvist bzw. Sorensen vorgeschlagenen Form nachvollziehbar<sup>111</sup> regimentieren. Die Möglichkeiten, bzw. Notwendigkeiten, die hier zur Debatte stehen, scheinen sämtlich physikalischer Art zu sein.<sup>112</sup>

### 3.2.4 KUHN UND CONFLICT VAGUENESS

Gedankenexperimente auf diese Weise als Paradoxien aufzufassen, steht, wie wir gesehen haben, in der Tradition Poppers, der ja ebenfalls Gedankenexperimente als ein Instrument betrachtet hat, das zeigen soll, „daß gewisse Möglichkeiten bei der Formulierung [einer] Theorie außer acht gelassen wurden.“<sup>113</sup> Wie steht es aber mit Thomas S. Kuhns Beobachtung, dass Gedankenexperimente keine Paradoxien im Überzeugungssystem *aufweisen*, sondern diese zum Zwecke einer Revision des Begriffssystems erst *herstellen*?

Eine Möglichkeit, Kuhns Einsichten in diesem Rahmen wiederzugeben, besteht darin, die von Kuhn beschriebenen Fälle der Begriffsentwicklung als Auflösungen begrifflicher Vagheiten zu analysieren.<sup>114</sup> Ein Begriff ist (in diesem Sinne) *vage*, wenn er intrinsische Grenzfälle besitzt, d.h., dass es Fälle gibt, bei denen es nicht geklärt werden kann, ob sie unter einen bestimmten Begriff fallen oder nicht. Manche Grenzfälle lassen sich darauf zurückführen, dass der fragliche Begriff nach zwei Kriterien angewendet werden kann, die für den Grenzfall zu unterschiedlichen Ergebnissen kommen. Ein klassisches Beispiel hierfür ist die Frage, ob man einen zweiköpfigen Drachen „enthaupet“ hat, wenn man ihm nur einen

<sup>110</sup> Zitat von Maxwell aus einem Brief an John William Strutt (Lord Raleigh) vom 6. Dezember 1870. Zitiert nach Earman und Norton [96], 439.

<sup>111</sup> Da man nahezu jedes Argument in nahezu jede logische Form „pressen“ kann, ist dieses Resultat nicht besonders spektakulär. Solche Regimentierungen sind so gut, wie sie dabei helfen können, bestimmte Aspekte einer Klasse von Argumenten deutlich zu machen. Wie gut die Sorensen/Häggqvist-Analyse ist, wird sich im Folgenden zeigen. Vgl. hierzu auch die Debatte in Bunzl [45], Feldman [103], Sorensen [306], Sorensen [307].

<sup>112</sup> Die restliche Geschichte zu Maxwells Dämon, sowie die Frage, ob es sich um ein erfolgreiches Gedankenexperiment handelt, wollen wir hier übergehen. Zur Illustration reicht dieser Teil der Episode (vgl. aber auch den übernächsten Abschnitt). Für den Rest vgl. Earman und Norton [96], Earman und Norton [97], Norton [242].

<sup>113</sup> Popper [260], 399.

<sup>114</sup> Vgl. hierzu Sorensen [304], 167-185, Sorensen [309].

Kopf abgeschlagen hat. Nach einem Kriterium hat man etwas dann „enthauptet“, wenn man ihm einen Kopf abgetrennt hat, nach einem anderen Kriterium gilt etwas nur dann als „enthauptet“, wenn es keinen Kopf mehr hat.

Die von Kuhn betrachteten Gedankenexperimente scheinen Fälle zu sein, in denen wissenschaftliche Begriffe, die in diesem Sinne vage sind, mit Grenzfällen konfrontiert werden, die von wissenschaftlichem Interesse sind (weil es sich um intendierte Anwendungsfälle der Theorie handelt, die aber bisher nicht im Zentrum der Aufmerksamkeit standen). Das Gedankenexperiment macht auf die Vagheit aufmerksam, indem es das mögliche widersprüchliche Verdikt der verschiedenen Kriterien an einem Beispiel hervorhebt. Damit legt es gleichzeitig die Grundlage für eine Begriffsreform: der Begriff kann sich aufspalten in zwei neue Begriffe, die nach jeweils einem der Kriterien angewendet werden. Auch die Kuhnschen Fälle lassen sich damit als Paradoxien in Sorensens Sinn auffassen.

### 3.2.5 SIND GEDANKENEXPERIMENTE ARGUMENTE?

Diese Analyse scheint es uns zu erlauben, Gedankenexperimente zur Überzeugungsänderung als Argumente zu rekonstruieren. Da es um gerechtfertigte Überzeugungsänderung gehen soll, wäre eine solche Rekonstruktion erhellend: ein offensichtlicher Weg über den man zu *gerechtfertigten* Überzeugungsänderungen kommt, besteht darin, sich durch ein beweiskräftiges Argument überzeugen zu lassen.

Gegen die Auffassung, Gedankenexperimente seien ohne weiteres als Argumente rekonstruierbar, könnte man mehrere Einwände formulieren:

- (1) Die Analyse übersieht das eigentlich „experimentelle“ am Gedankenexperiment. Indem man ein Gedankenexperiment als Argument rekonstruiert, sieht man doch davon ab, dass Gedankenexperimente – wie Experimente – Ereignisse sind. Die Analyse übersieht also einen zentralen Aspekt.
- (2) Die Analyse identifiziert ein Gedankenexperiment mit einem bestimmten Argument. Häufig ist es jedoch so, dass dieselben Gedankenexperimente in ganz anderen Kontexten weiterverwendet werden, um möglicherweise eine ganz andere Theorie zu kritisieren. Dann handelt es sich immer noch um dasselbe Gedankenexperiment, aber nicht mehr um dasselbe Argument. Gedankenexperimente haben andere Identitätsbedingungen als Argumente.<sup>115</sup>

Zum ersten Einwand ist zu sagen, dass die Analyse von Gedankenexperimenten als Argumente bestimmter Art, oder als Teile von Argumenten, nicht bedeuten muss, dass man Gedankenexperimente auf Argumente *reduziert*. Gedankenexperimente

---

<sup>115</sup> Zur Debatte um diesen Punkt vgl. Bishop [25] und Hacking [136].

mente zur Überzeugungsänderung sind aber *auch* Argumente und lassen sich als solche analysieren.

Damit wird auch nicht behauptet, dass sich alle erkenntnistheoretischen Fragen erledigt haben, wenn man ein Gedankenexperiment als gültiges modales Argument rekonstruiert hat. Der folgende Vorwurf von Kühne scheint daher unbegründet:

Das „Gute“ der modallogischen Analyse von Gedankenexperimenten liegt darin, daß hiermit – falls man daran interessiert ist – eine Klassifikation von unterschiedlichen Argumentationstypen des Gedankenexperiments begründet werden kann. Das „Schlechte“ ist die von einigen Autoren gar nicht wahrgenommene Einsicht, daß durch die formallogische Analyse nicht der Funken einer Erklärung der Gedankenexperimente, ihres Erkenntnisgewinns oder ihrer Funktion innerhalb der Wissenschaften, erreicht werden kann. (Kühne [175], 291)

Zunächst einmal stimmt es gar nicht, dass kein „Funken“ einer Erklärung aus der Argumentrekonstruktion gewonnen werden kann. Wenn man zu jedem Gedankenexperiment eine *Targetthese* identifizieren kann, die das Gedankenexperiment zu widerlegen versucht, hat man offenbar schon einen guten Hinweis auf die Funktion von Gedankenexperimenten innerhalb der Wissenschaft.

Kühne wirft (modal)logischen Argumentrekonstruktionen zum Zwecke der Erläuterung des ursprünglichen Arguments außerdem eine Umkehrung der Erklärungsrichtung vor: Die Logik beschreibt die Regeln guten Argumentierens, daher erklären gute Argumente, warum die Logik bestimmte Regeln enthält, die Logik erklärt aber nicht, warum bestimmte Argumente gut sind:

Die Kalküle der Modallogik versuchen die akzeptierten Regeln für modales Argumentieren abzubilden und zu explizieren. Historisch betrachtet waren es die akzeptablen Argumentationsschemata (beispielsweise von „guten“ Gedankenexperimenten), die zur Auswahl geeigneter Axiome und Schlußregeln der Modallogik geführt haben (und auch heute noch zur Kritik von vorliegenden Kalkülen herangezogen werden), weshalb es zirkulär wäre, die Auswahl von akzeptablen Schlüssen von Gedankenexperimenten jetzt mit Eigenschaften der Modallogik begründen zu wollen. (Kühne [175], 294)

Auch dieser Einwand läuft wieder darauf hinaus, dass Argumentrekonstruktionen nicht erhellend sein können und daher am „eigentlichen“ Problem vorbeigehen. Wenn sich die Rekonstruktionstechnik (Logik) danach richtet, ob wir das Argument intuitiv für gültig halten, kann uns eine formale Argumentrekonstruktion nichts sagen, was wir nicht vorher schon wussten.

Hinter diesem Vorwurf steckt aber ein signifikantes Missverständnis der Logik oder ein zu enges Verständnis von Zirkularitäten. In der Tat versucht die Logik ein vortheoretisches Verständnis von logischen Folgerungsbeziehungen und gerechtfertigten Schlussfolgerungen in modelltheoretischen und beweistheoretischen Rekonstruktionen einzufangen. Dies gelingt der Logik nur zu einem bestimmten Grad. So gibt es Relevanzlogiken oder parakonsistente Logiken als Alternativvorschläge zur Standardlogik, weil die Stan-

dardlogik bestimmte Schlüsse als gerechtfertigt betrachtet, die intuitiv nicht gerechtfertigt erscheinen.<sup>116</sup> Trotz der Tatsache, dass auf diese Weise versucht wird (und historisch versucht worden ist), „die“ Logik unseren vortheoretischen Auffassungen anzugleichen, spielt sie dennoch *erfolgreich* eine normative Rolle als Lehre vom gültigen Schließen. Wie kann das sein? Die Antwort auf dieses Scheinproblem ist bekannt: es handelt sich um einen „guten“ Zirkel. Er beschreibt das Verfahren, mit dem wir zu einem *reflektierten Gleichgewicht* zwischen allgemein verbindlichen Normen und individuellen Einzelfallbeurteilungen kommen. Wir wollen diese Frage an dieser Stelle nicht im Detail erörtern. Hier sei nur soviel bemerkt: Modallogiken, wie auch die Standardlogik *funktionieren* als normative Theorien, auch wenn diese Normen durch die Methode des Gegenbeispiels weiterentwickelt worden sind und weiterentwickelt werden.<sup>117</sup>

Damit wird aber weder hier, noch sonst wo<sup>118</sup> behauptet, dass sich alle wissenschaftstheoretischen Probleme des Themas Gedankenexperiment durch eine Argumentrekonstruktion erübrigen würden. In vielen Hinsichten ist die Rekonstruktion erst die Grundlage der weiteren Analyse. Die erkenntnistheoretischen Fragen danach, wie sich die Prämissen des (so rekonstruierten) Arguments rechtfertigen lassen, können ja jetzt erst sinnvoll *gestellt* werden.

Im nächsten Unterkapitel werden wir auf die Frage, in welchem Sinne sich Gedankenexperimente auf Argumente reduzieren lassen, noch einmal eingehen. Es sei hier nur klargestellt, dass die Behauptung, dass sich Gedankenexperimente als Argumente rekonstruieren lassen, *nicht* impliziert, dass sie sich nicht auch in anderen Hinsichten analysieren lassen (psychologisch, historisch, rhetorisch, etc.), sie impliziert auch nicht, dass sämtliche erkenntnistheoretischen Fragen durch eine Argumentrekonstruktion erledigt sind.

Der zweite Einwand, dass Gedankenexperimente ein Eigenleben entwickeln können, unabhängig von dem argumentativen Kontext in dem sie einmal entstanden sind, scheint ein bloßer Streit um Worte zu sein. So hat Maxwells Dämon seit seiner Entstehung eine bemerkenswerte Entwicklung durchgemacht. Das Gedankenexperiment war zunächst gegen die strikte Notwendigkeit des zweiten Hauptsatzes der Thermodynamik gerichtet, der Dämon wurde aber von späteren Autoren dazu benutzt, die Möglichkeit eines Perpetuum mobiles zweiter Art zu etablieren, was der Thermodynamik direkt widerspricht. Der Dämon hat außerdem selbst eine Weiterentwicklung erlebt: wurde er ursprünglich als sehr kleines, flinkes intelligentes Wesen betrachtet, so wurde er später zu einem Klapptür-Mechanismus, einem Kraftfeld, oder einem Computer.<sup>119</sup> In manchen Versionen des Gedankenexperiments ist der Dämon selbst Teil des thermodynamisch beschriebenen Systems, in anderen operiert er außerhalb der Thermodynamik. Man kann

---

<sup>116</sup> Vgl. Bremer [33], Bremer und Cohnitz [34].

<sup>117</sup> Für eine detaillierte Auseinandersetzung mit diesem Problem vgl. Cohnitz und Rossberg [75], Kapitel 2.

<sup>118</sup> ...vielleicht mit Ausnahme von Rehder [275].

<sup>119</sup> Vgl. Earman und Norton [96], Leff und Rex [181].

in einem solchen Fall von der Entwicklung *eines* Gedankenexperiments sprechen, für die uns interessierenden erkenntnistheoretischen Fragen ist diese Redeweise aber nicht sinnvoll. Wenn wir untersuchen wollen, ob Gedankenexperimente zu gerechtfertigten Überzeugungsänderungen führen, interessieren wir uns dafür, warum *dieses* zu einem bestimmten Zeitpunkt vorgetragene Gedankenexperiment eine Revision *dieser* Überzeugung(en) rechtfertigen soll. Wandelt sich also das Argument, wandelt sich auch das Gedankenexperiment.

Man kann dem auch aus der anderen Richtung begegnen: Argumente werden gewöhnlich durch ihre Prämissen und ihre Konklusion identifiziert, dennoch ist es nicht umgangssprachlich selbstwidersprüchlich, von jemandem zu sagen, er habe im Grunde „dasselbe“ Argument noch einmal benutzt, nur mit veränderten Prämissen. Manchmal kann es sinnvoll sein, sich auf diese Weise auszudrücken. Manchmal ist es nützlicher, die Identitätsbedingungen für Argumente strikter zu fassen (wenn man beispielsweise vermeiden will, dass „dasselbe“ Argument gültig und ungültig sein kann).

Es stimmt einfach nicht, dass nur manche Redeweisen legitim sind, insbesondere, wenn es sich um einen quasi-theoretischen Term handelt. Wir haben bisher ‚Gedankenexperiment‘ zur Bezeichnung folgender verschiedener Dinge kennen gelernt:

- ( $I_1$ ) Das Gedankenexperiment als Argument, dessen Konklusion eine bestimmte Aussage über die Wirklichkeit ist und in dessen Prämissen auf bestimmte imaginäre Sachverhalte Bezug genommen wird. (Entspricht dem gesamten Haggqvist-Schema.)
- ( $I_2$ ) Das Gedankenexperimente als derjenige Teil einer solchen Argumentation, der nur aus der Beschreibung des imaginären Sachverhalts besteht und eine Aussage darüber trifft, was in diesem imaginären Fall in einem weiteren Sinne der Fall wäre, aber noch keine Schlussfolgerungen in Bezug auf eine wissenschaftliche Theorie oder die Wirklichkeit enthält. (Entspricht dem Schema ohne Konklusion und ohne Prämisse (2).)
- ( $I_3$ ) Das Gedankenexperiment als derjenige Teil einer solchen Argumentation, der nur aus der Beschreibung eines imaginären Sachverhalts besteht, ohne dass ein Urteil darüber gefällt würde, was in diesem Fall in einem weiteren Sinne der Fall wäre. (Entspricht dem C aus Prämisse (1) des Schemas.)
- ( $I_4$ ) Der psychische Akt des Vorstellens von ( $I_3$ ). (Etabliert Prämisse (1) des Schemas.)
- ( $I_5$ ) Der psychische Akt des Urteilens in Bezug auf ( $I_4$ ). (Etabliert die Schritte (2) und (3) des Schemas.)
- ( $I_6$ ) Der psychische Akt des Schlussfolgerns in Bezug auf ( $I_1$ ). (Etabliert die Konklusion des Schemas.)

Es ist ziemlich offensichtlich, dass es sich hier um ganz verschiedene Dinge handelt. Wenn man will, könnte man diese Liste noch um *konkrete* psychische Akte erweitern (wenn man beispielsweise zählen will, wie oft man ein Gedankenexperiment in diesem Sinne wiederholt hat). Welche Verwendungsweise man dabei auszeichnen möchte, ist Geschmackssache. Wir werden ab jetzt  $T_1$  auszeichnen und an allen abweichenden Stellen Bezug auf die hier getroffenen Unterscheidungen nehmen.

### 3.2.6 DIE RECHTFERTIGUNG VON (PHYSIKALISCHEN) GEDANKENEXPERIMENTEN DURCH DIE EVOLUTIONÄRE ERKENNTNISTHEORIE

Wie wir gesehen haben, sind manche Autoren der Meinung, dass in Gedankenexperimenten bestimmte „instinktive Erkenntnisse“ (Mach) oder „mentale Simulationen“ (Nersessian, Sorensen) physikalischer Abläufe Auskunft über tatsächlich mögliche physikalische Abläufe geben. Diese physikalischen „Intuitionen“ spielen nach unserer Rekonstruktion insbesondere immer dann im Gedankenexperiment eine Rolle, wenn die physikalische Möglichkeit von C behauptet wird, oder die physikalische Unmöglichkeit von W, gegeben C. Für diese Urteile stellt sich ein erkenntnistheoretisches Problem: während reale Experimente, das, was physikalisch (un)möglich ist, durch Beobachtung an *tatsächlichen* physikalischen Abläufen etablieren, scheint der Inhalt unserer instinktiven Erkenntnisse und der Ausgang mentaler Simulationen zunächst von *uns* abzuhängen, nicht aber von den physikalischen Vorgängen, die wir damit beurteilen wollen.

Mach, Sorensen und Nersessian geben darauf folgende Replik: unsere instinktiven Erkenntnisse, wie auch unsere Fähigkeit, mentale Modelle physikalischer Abläufe zu bilden, sind uns durch die Evolution mitgegeben. Für unseren reproduktiven Erfolg wird es besser gewesen sein, dass diese instinktiven Urteile wahr sind und unsere mentalen Simulationen mit den Abläufen in der Welt harmonieren. Es ist also davon auszugehen, dass die Ergebnisse dieser Fähigkeiten im Großen und Ganzen zuverlässig sind.<sup>120</sup>

Diese Erklärung legt dabei bereits bestimmte Grenzen nahe: Sehr kleine und sehr große Dinge und ihr physikalisches Verhalten sind für unseren Überlebenskampf vermutlich irrelevant gewesen, warum also sollten wir gute Gedankenexperimentierer in der Quantenmechanik oder der Relativitätstheorie sein? D.h. in den Bereichen, in denen wir (z.B. laut Hempel und Popper) besonders viele Gedankenexperimente vorfinden, sollten sie eigentlich am Wenigsten verloren haben.

---

<sup>120</sup> Obwohl Sorensen an mehreren Stellen verspricht, eine moderne evolutionsbiologische Erklärung für unsere instinktive Einsicht in physikalische Zusammenhänge zu geben, gehen auch seine Ausführungen nicht über die soeben aufgeführten Behauptungen hinaus. Vgl. insbesondere Sorensen [305]. Diese dürftigen Erklärungsskizzen sind natürlich nicht hinreichend für eine halbwegs plausible evolutionäre Erklärung. Vgl. hierzu insbesondere Maffie [198] und Sober [303].

Ein zweites, gravierenderes Problem besteht darin, dass unsere instinktiven oder intuitiven physikalischen Urteile nicht auf erklärungsbedürftige Weise zuverlässig sind, sondern in alarmierendem Maße *unzuverlässig*.

In mehreren psychologischen Studien ist nachgewiesen worden, dass unsere „intuitive Physik“ eher eine mittelalterliche Theorie stützen würde als die Newtonsche Physik. Diese Intuitionen werden auch nicht signifikant dadurch „verbessert“, dass man Physik studiert. Die intuitiven Beurteilungen physikalischer Vorgänge bleiben unzuverlässig.<sup>121</sup> Nimmt man an, dass es nicht spezifische Überzeugungen sind, die uns durch die Evolutionsgeschichte mitgegeben werden, sondern überzeugungsbildende Mechanismen, dann deutet dieser Befund außerdem stark daraufhin, dass etwas beim *mental modelling* schief geht. Dieser Mechanismus mag gut geeignet sein, will man schnell erinnern, wie viele Fenster die eigene Wohnung hat, dass er bei theoretischer Physik helfen soll, ist aber mehr als fraglich.<sup>122</sup> Wenn Physikabsolventen durch *mental modelling* zu falschen Konsequenzen kommen, und diese Resultate nicht dadurch wegerklärt werden können, dass die falschen Konsequenzen Atavismen der Evolutionsgeschichte sind, ist offenbar der Prozess, mit dem diese Konsequenzen generiert wurden, unzuverlässig.

Diese Problematik, die zunächst für physikalische Gedankenexperimente spezifisch ist, kann hier nicht erschöpfend behandelt werden. Aber es gibt eine Möglichkeiten, die Problematik zu entspannen, die mit Themen dieser Arbeit in Verbindung steht:

Sofern physikalische Intuitionen tatsächlich den Kern eines Gedankenexperiments ausmachen, scheinen sie sich in der Regel auf sehr konkrete und primitive physikalische Zusammenhänge zu beziehen, für die tatsächlich angenommen werden darf, dass die Intuitionen hier zuverlässig sind. Physikalische Intuitionen sind in diesem Sinne der „Kern eines Gedankenexperiments“, wenn der Ausgang des Gedankenexperiments, also das, was durch das „mind's eye“ Beobachtbares passiert, bereits der Targetthese widerspricht. In diesen „trivialen“ Fällen sollte man annehmen dürfen, dass die instinktive Reaktion zuverlässig ist. Galileis und Stevins Gedankenexperimente sind offenbar von diesem Typ. Wir werden diese These im nächsten Kapitel etwas genauer untersuchen.

Die Kehrseite dieser These ist, dass Gedankenexperimente, in denen offensichtlich keine physikalisch trivialen Umstände vorkommen, physikalische Intuitionen auch nicht den Kern des Gedankenexperiments ausmachen dürften. Ein solcher Fall liegt beispielsweise im Gedankenexperiment von Einstein, Podolsky und Rosen vor, bei dem ein physikalisch nicht-trivialer Versuchsablauf die Grundlage des Gedankenexperiments bildet.<sup>123</sup>

---

<sup>121</sup> Für eine Übersicht vgl. Nersessian und Resnick [226]. Für die empirischen Untersuchungen vgl. Caramazza, McCloskey und Green [48], Clement [67], Halloun und Hestenes [140], McDermott [204] und Viennot [335].

<sup>122</sup> Insbesondere erscheint die Empfehlung, mentale Simulationen anstelle von propositionalen Argumenten zu verwenden, angesichts dieser Ergebnisse als höchst fragwürdig.

<sup>123</sup> Vgl. zum Folgenden Cohnitz [73], Wheeler und Zurek [340], Held [142], Düster [95].

Einstein, Podolsky und Rosen (EPR) versuchten (nach herkömmlicher Auffassung) mit ihrem Gedankenexperiment nicht die Falschheit der Quantenmechanik nachzuweisen, schon gar nicht, indem sie versucht hätten, empirische Konsequenzen der Quantenmechanik als unintuitiv erscheinen zu lassen.<sup>124</sup> Was gezeigt werden sollte, war nur, dass eine bestimmte *Interpretation* der Quantenmechanik falsch ist, insbesondere die Interpretation Niels Bohrs, nach der die Quantenmechanik als *vollständige* Theorie der physikalischen Wirklichkeit betrachtet werden kann.

Dies sollte dadurch gezeigt werden, dass, gegeben ein bestimmtes Kriterium für „physikalische Wirklichkeit“, die Quantenmechanik die Ableitung von mehr Elementen der physikalischen Wirklichkeit erlaubt, als die Quantenmechanik eigentlich zulässt. Das herangezogene Kriterium war Folgendes:

If, without any way disturbing a system, we can predict with certainty (i.e., with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity. (Einstein, Podolsky und Rosen [100], 138)

Das Argument von EPR bestand dann darin, nachzuweisen, dass, gegeben ein bestimmter Experimentaufbau, die Quantenmechanik mit „Gewissheit“ die Prognose des Werts von zwei physikalischen Größen (P und Q) erlaubt, obwohl die Quantenmechanik zugleich behauptet, dass nicht beiden Größen simultane physikalische Wirklichkeit zukommt. Die Unvollständigkeit der Quantenmechanik sollte also in Abhängigkeit von orthodoxer Quantenmechanik und dem auf bestimmte Weise interpretierten Realitätskriterium aufgewiesen werden.

One could object to this conclusion on the grounds that our criterion of reality is not sufficiently restrictive. Indeed, one would not arrive at our conclusion if one insisted that two or more physical quantities can be regarded as simultaneous elements of reality *only when they can be simultaneously measured or predicted*. On this point of view, since either or the other, but not both simultaneously, of the quantities P and Q depend upon the process of measurement carried out on the first system, which does not disturb the second system in any way. No reasonable definition of reality of P and Q could be expected to permit this. (Einstein, Podolsky und Rosen [100], 141)

Obwohl der Ablauf des beschriebenen Experiments die *Grundlage* dieses Arguments bildet, bestand der kontroverse Teil des EPR-Gedankenexperiments *nicht* im empirischen Ausgang des geschilderten Experiments.<sup>125</sup> Der Ausgang des Experiments ist ganz das, was die Quantenmechanik prognostiziert. Die Kontroverse besteht nur hinsichtlich der Frage, was als zufrieden stellende Erklärung eines solchen Phänomens zählt. EPR argumentieren mit einem epistemologischen Prinzip

---

<sup>124</sup> Das ist – wohlgemerkt – auch versucht worden. In der Regel geschah dies aber mit Gedankenexperimenten, die auf der Makroebene spielten (wie etwa „Schrödingers Katze“). Die Ausnahmen von der Regel müssen – da erfolglos – hier nicht wegeklärt werden.

<sup>125</sup> Das wird häufig übersehen (zum Beispiel von Peijnenburg und Atkinson [252]).

und der Annahme, dass empirisch äquivalente Theorien entwickelt werden können, die diesem Prinzip besser entsprechen. Dass die zweite dieser intuitiven Annahmen sich als empirisch falsch herausstellen *könnte*, war damals vermutlich<sup>126</sup> unklar. Zunächst musste das EPR-Gedankenexperiment von David Bohm in den 50er Jahren umgearbeitet werden, was dann den dafür Weg ebnete, dass John Bell 1964 seine berühmten Ungleichungen formulieren konnte.<sup>127</sup> In den frühen 80er Jahren hat Alain Aspect (vgl. Aspect [4]) dann experimentell nachgewiesen, dass die Bellschen Ungleichungen falsch sind, was die Hintergrundannahme falsifizierte, dass es bestimmte zur Quantenmechanik empirisch äquivalente Theorien gibt, die das EPR-Kriterium für physikalische Wirklichkeit besser erfüllen als die Quantenmechanik.

Die *Intuitionen*, die im Gedankenexperiment eine Rolle spielten, bezogen sich also einerseits auf Gütekriterien für physikalische Theorien, andererseits auf die Frage nach der Möglichkeit von Alternativtheorien bestimmter Art. Letzteres war zwischen den diskutierenden Parteien nicht Gegenstand der Kontroverse (Bohrs Replik besteht eben nicht in der Zurückweisung dieser Annahme, oder in einer Korrektur am vorgestellten Experimentablauf), Ersteres schon (Bohr versucht das EPR-Prinzip für physikalische Wirklichkeit als „essentiell ambig“ zu kritisieren<sup>128</sup>). Die *Intuitionen* bezogen sich also nicht auf den nicht-trivialen Experimentausgang, sondern auf ein Problem, das eher im Bereich der Philosophie der Naturwissenschaften anzusiedeln ist. Insofern handelt es sich bei solchen Gedankenexperimenten vielleicht eher um wissenschaftstheoretische bzw. naturphilosophische Gedankenexperimente. Weshalb philosophische Gedankenexperimente funktionieren, wird uns in den nächsten Kapiteln beschäftigen. Ob sich von da aus das epistemologische Rätsel der Zuverlässigkeit nicht-trivialer Gedankenexperimente in den Naturwissenschaften lösen lässt, können wir erst am Ende beurteilen. Wir werden in Kapitel 5 auf diese Fragen zurückkommen.

### 3.2.7 SUBSTANTIELLE UND AKZIDENTELLE EIGENSCHAFTEN VON GEDANKEN-EXPERIMENTEN

Wie wir aus der Geschichte der Wissenschaftstheorie erfahren konnten, liegt von den verschiedenen Funktionen, die ein Gedankenexperiment haben kann, die Funktion der gerechtfertigten Überzeugungsänderung im Mittelpunkt des Interesses. Es wird allgemein davon ausgegangen, dass Gedankenexperimente in der Wissenschaftsgeschichte hauptsächlich dazu intendiert waren, beim Adressaten eine gerechtfertigte Überzeugungsänderung herbeizuführen. Ulrich Gähde hat, sich auf

---

<sup>126</sup> Vgl. Peijnenburg und Atkinson [252].

<sup>127</sup> Vgl. Peijnenburg und Atkinson [252], Bell [19], Bell [20].

<sup>128</sup> Bohr [29].

diese Klasse von Gedankenexperimenten konzentrierend, folgende substantielle Eigenschaften von Gedankenexperimenten formuliert:<sup>129</sup>

- (G1) Das Ziel jedes Gedankenexperiments besteht darin, begründete Überzeugungsänderungen beim Adressaten zu bewirken.
- (G2) Um dieses Ziel zu erreichen, werden entweder bestimmte denk- oder vorstellbare Sachverhalte beschrieben oder Experimente geschildert.
- (G3) Das Ziel der Überzeugungsänderung wird erreicht, ohne dass die entsprechenden Sachverhalte als real angenommen oder diese Experimente *de facto* ausgeführt werden müssen.

Die Eigenschaft (G3) ist – so formuliert – auch keine Eigenschaft aller, sondern nur eine Eigenschaft aller *erfolgreichen* Gedankenexperimente, sofern es solche gibt. In jedem Fall ist es aber so, dass (G3) vom Gedankenexperimentierer intendiert wird. Ähnliches gilt für (G2), wenn man zulassen möchte, dass es auch misslungene „Gedankenexperimente“ geben kann. Adäquater wäre daher folgende Formulierung:

- (G1) Das Ziel jedes Gedankenexperiments besteht darin, begründete Überzeugungsänderungen beim Adressaten zu bewirken.
- (G2') Um dieses Ziel zu erreichen, werden Sachverhalte beschrieben oder Experimente geschildert, die denkbar oder vorstellbar sein sollen.
- (G3') Das Ziel der Überzeugungsänderung soll erreicht werden, ohne dass die entsprechenden Sachverhalte als real angenommen oder diese Experimente *de facto* ausgeführt werden müssen.

Gedankenexperimente zur Überzeugungsänderung lassen sich, wie wir bei Sorensen und Häggqvist gesehen haben, häufig als ein Argument rekonstruieren, bei der die Überzeugung, deren Revision intendiert wird, als Targetthese fungiert: ihre Negation ist die Konklusion des Arguments. Dies ist auch sehr plausibel: Häufig kommen wir zu gerechtfertigten Überzeugungsänderungen dadurch, dass wir uns durch ein gutes Argument überzeugen lassen.

Bei naturwissenschaftlichen Gedankenexperimenten scheint der als vorstellbar behauptete Sachverhalt häufig neben seiner Vorstellbarkeit noch die Bedingung erfüllen zu müssen, dass er physikalisch möglich ist. Wie in (G3') aber richtig festgestellt wird, ist es in der Regel nicht erforderlich, dass der Sachverhalt realisierbar ist (wie wir aus der Wissenschaftsgeschichte gelernt haben, ist es aber auch nicht erforderlich, dass der Sachverhalt unrealisierbar ist, viele Gedankenexperimente konnten (zumindest näherungsweise) realisiert werden).

---

<sup>129</sup> Diese Charakterisierung stammt von Ulrich Gähde, vgl. Gähde [113].

Urteile darüber, was physikalisch möglich ist, werden dabei manchmal intuitiv gefällt, wobei die Rechtfertigung dieser Urteile fragwürdig ist, zumindest sobald physikalische Sachverhalte beurteilt werden, die in Bereichen liegen, mit denen wir nicht vertraut sind. Wie wir gesehen haben, ist aber nicht jedes Gedankenexperiment in der Physik auf physikalische Intuitionen aufgebaut. Manchmal – wie möglicherweise im EPR-Gedankenexperiment – wird die Rechtfertigung für die physikalische Möglichkeit eines Sachverhaltes einer Theorie entnommen. In diesen Fällen scheint aber auch keine gerechtfertigte Überzeugungsänderung in Bezug auf eine empirische Frage *intendiert*, sondern eher in Bezug auf eine wissenschaftstheoretische oder naturphilosophische, für die vielleicht andere Intuitionen eine Rolle spielen.

Die Zuverlässigkeit unserer physikalischen Intuitionen ist demnach nicht unbedingt von Relevanz für die Beurteilung *aller* physikalischen Gedankenexperimente. Wir erinnern uns an Hempels Unterscheidung zwischen theoretischen und intuitiven Gedankenexperimenten: wenn man ‚theoretisch‘ so versteht, dass in diese Gedankenexperimente nur keine *physikalischen* Intuitionen eingehen, könnte Hempels Charakterisierung zweier *Extremtypen* hier erneut angewendet werden. Hempels These, dass es sich um Extremtypen handelt, zwischen denen ein Kontinuum von Zwischentypen gibt, wäre dann die These, dass es zwischen physikalischen und philosophischen Gedankenexperimenten nur graduelle Unterschiede gibt. Ob dem so ist, kann erst beurteilt werden, wenn wir genauer wissen, was eigentlich philosophische Gedankenexperimente auszeichnet. Ob dann damit viel gewonnen ist, wird wiederum davon abhängen, ob das (vermutlich ebenfalls auftretende) Rechtfertigungsproblem in der Philosophie leichter gelöst werden kann.

### 3.3 DIE ELIMINIERBARKEIT VON GEDANKENEXPERIMENTEN

In diesem Kapitel wollen wir uns zum Abschluss unseres Ausflugs in die Philosophie der Naturwissenschaften der Frage zuwenden, ob Gedankenexperimente eliminierbar sind. In einem näher zu spezifizierenden Sinn scheinen manche Gedankenexperimente tatsächlich nicht eliminierbar zu sein, was dafür sprechen würde, dass sie Bestandteil jeder rationalen Rekonstruktion der Wissenschaftsgeschichte sein müssen.

#### 3.3.1 DIE NORTON-BROWN DEBATTE UND DIE FRAGE DER ELIMINIERBARKEIT

Wie wir oben bereits bemerkt haben, beruht die Debatte, ob Gedankenexperimente Argumente sind, zum Teil auf Scheinproblemen (wie der Frage, ob Gedankenexperimente und Argumente dieselben diachronen Identitätsbedingungen besitzen).

Wie wir sahen, konnte man einen solchen Grund finden, wenn man Gedankenexperimente in großer Nähe zu realen Experimenten sieht. Natürlich lassen sich auch Experimente als Argumente rekonstruieren, indem man durch die im Experiment gewonnenen Beobachtungssätze und Zusatzannahmen für oder gegen eine physikalische Theorie „argumentiert“. Würde man aber behaupten, Experimente seien nichts anderes als Argumente? Vermutlich bekäme man in diesem Fall zu Recht vorgeworfen, dass man dabei viele Dinge ausblendet, die ein Experiment gerade zum Experiment machen, wie das systematische Beobachten, gegebenenfalls das Eingreifen in Naturvorgänge oder die künstliche Herstellung bestimmter naturanaloger Abläufe im Labor, etc.<sup>130</sup>

Diejenigen, die hartnäckig Gedankenexperimente mit Argumenten identifizieren wollen, sehen sich hiervon in der Regel unbeeindruckt, was daran liegen mag, dass sie die Tatsache, dass sich Gedankenexperimente im Geist abspielen, zu sehr als Indikator dafür nehmen, dass es sich dabei im wesentlichen um Schlussfolgerungsprozesse handelt (die sich in Argumentform gut rekonstruieren lassen).

James Robert Brown<sup>131</sup> ist dagegen der Meinung, dass dabei ein wesentlicher Aspekt von Gedankenexperimenten übersehen wird, nämlich der, dass neben den Schlussfolgerungsprozessen im Geist auch Beobachtungen gemacht werden. Beobachtungen, die im Grunde analog zu den Beobachtungen im richtigen Experiment zu verstehen sind, nur mit dem Unterschied, dass wir keine konkreten Objekte beobachten, sondern abstrakte, und dass wir dafür nicht unsere (technisch erweiterten) fünf Sinne benutzen, sondern einen sechsten.

Wir wollen uns mit dieser Position nicht länger auseinandersetzen; sie ist aus offensichtlichen Gründen<sup>132</sup> schon von nahezu jedem kritisiert worden, der sich mit dem Thema Gedankenexperiment beschäftigt hat. Einer der hartnäckigsten Kritiker dieser Position ist John Norton.<sup>133</sup> Er vertritt die Brown genau entgegen gesetzte Meinung, dass sich Gedankenexperimente sehr wohl als Argumente rekonstruieren lassen, ja dass sie zugunsten nicht-gedankenexperimenteller Argumente ersetzt werden können. Tamar Szabó Gendler hat Nortons These folgendermaßen zusammengefasst:

**The Elimination Thesis:** Any conclusion reached by a (successful) thought experiment will also be demonstrable by a non-thought-experimental argument.  
(Gendler [121], 34)

---

<sup>130</sup> Auch bei ‚Experiment‘ handelt es sich hier allerdings nur um einen Streit um Worte.

<sup>131</sup> Vgl. Brown [38], Brown [39], Brown [40], Brown [41], Brown [42].

<sup>132</sup> Das Hauptproblem besteht darin, dass als Schluss auf die „beste“ Erklärung des Erfolgs von Gedankenexperimenten ein platonisches Reich von Naturgesetzen postuliert wird, sowie eine mysteriöse Fähigkeit diese wahrzunehmen, ohne dass dafür irgendein Sinnesorgan existieren würde (mal ganz zu schweigen von der Tatsache, dass auch ein solches Organ noch damit zu kämpfen hätte, wie man mit abstrakten Objekten in kausalen Kontakt kommt). Rundheraus zu leugnen, dass Gedankenexperimente jemals funktioniert haben, scheint eine einfachere und wesentlich leichter zu verteidigende These zu sein.

<sup>133</sup> Vgl. Norton [238], Norton [239], Norton [240], Norton [241].

Obwohl wir dafür argumentiert haben, dass Gedankenexperimente als Argumente rekonstruiert werden können, und kein Anlass dafür besteht, anzunehmen, dass wir dabei eine mysteriöse Fähigkeit platonische Objekte wahrzunehmen, übergangen haben, scheint Nortons These zu stark zu sein. Wie Tamar Szabó Gendler überzeugend argumentiert hat, sind zumindest manche Gedankenexperimente nicht eliminierbar.

Die Eliminationsthese besagt, dass jedes gute Gedankenexperiment in ein nicht-gedankenexperimentelles Argument überführt werden kann, ohne dass dies die Überzeugungskraft des Arguments reduzieren würde. Eine *Elimination* bestünde zunächst darin, die narrative Präsentation des Gedankenexperiments in eine Reihe expliziter Prämissen umzuformen, die dazu hinreichen, die gewünschte Konklusion zu erschließen. In einem zweiten Schritt müssten in dieser Prämissenmenge all jene Prämissen, die sich auf hypothetische oder kontrafaktische Umstände, sowie auf Einzeldinge beziehen, durch Prämissen ersetzt werden, die eine solche Bezugnahme nicht enthalten. Falls die Eliminationsthese wahr ist, bewahrt eine solche Umformung die Überzeugungskraft des ursprünglichen Gedankenexperiments.

„Überzeugungskraft“ bezieht sich auf die rechtfertigende Rolle, die Gedankenexperimente in der Wissenschaftsgeschichte gespielt haben. Es geht dabei nicht um persönliche Überzeugungsergebnisse von Einzelpersonen, die – aus psychologischen Gründen – ohne ein bestimmtes Gedankenexperiment vielleicht nicht eingetreten wären, noch um die Frage, ob man diejenigen Überzeugungen, die man zu einem Zeitpunkt in der Wissenschaftsgeschichte durch ein Gedankenexperiment gewonnen hat, nicht in einem späteren Stadium der Wissenschaftsgeschichte auf andere Weise gewinnen konnte. Wie Mach bereits bemerkte, kann es sein, dass in einer reifen Wissenschaft Gedankenexperimente überflüssig werden, weil man längst andere, bessere Evidenz besitzt. Das bedeutet aber noch nicht, dass die Gedankenexperimente auch bereits zu den früheren Zeitpunkten überflüssig waren.

The proper reading of “demonstrative force” makes the Elimination Thesis epistemologically interesting. On this reading, demonstrative force concerns the role that thought experiments play in living bodies of knowledge: after the moment of discovery and before the end of inquiry. It concerns whether a particular conclusion based on a particular process of reasoning (thought experiment) is thereby *justified* – whether if such process leads to true beliefs, those beliefs should count as knowledge. (Gendler [121], 36)

Wir haben bereits dafür argumentiert, dass Gedankenexperimente als Argumente rekonstruiert werden können, haben sie aber als Argumente rekonstruiert, in denen explizit modale Prämissen vorkommen und in denen Bezug auf Einzeldinge genommen wird. Wenn Norton Recht hat, greift diese Rekonstruktion noch zu kurz, weil die eigentliche Einsicht in die Funktionsweise von Gedankenexperimenten nur dadurch erreicht werden kann, dass man die nicht-modalen und universalen Prämissen offen legt, die für den Erkenntniszuwachs verantwortlich sind.

### 3.3.2 NICHTELIMINIERBARKEIT

Gendlers These ist, dass wir unsere Argumentrekonstruktion weit genug getrieben haben. Gedankenexperimente lassen sich nicht als nicht-gedankenexperimentelle Argumente rekonstruieren, ohne dass diese Argumente ihre Überzeugungskraft einbüßen.<sup>134</sup> Gendler argumentiert dabei im Grunde gegen zwei Thesen, die „Dispensability Thesis“ und die „Derivability Thesis“, wobei beide als Präzisierung dessen gemeint sind, was Norton ursprünglich mit seiner *Elimination Thesis* ausdrücken wolte:

**The Dispensability Thesis:** Any good scientific thought experiment can be replaced, without loss of demonstrative force, by a non-thought-experimental argument.

**The Derivability Thesis:** The justificatory force of any good scientific thought experiment can only be explained by the fact that it can be replaced, without loss of demonstrative force, by a non-thought-experimental argument. (Gendler [121], 37-38)

Es ist klar, dass, wenn die erste der beiden Thesen falsch ist, auch die zweite falsch sein muss. Wir werden uns daher auf die erste These konzentrieren. Gendler versucht auch noch nachzuweisen, dass, selbst wenn sich eine Argumentrekonstruktion finden lässt, die dieselbe Überzeugungskraft hat wie das Gedankenexperiment, dies keinen Grund darstellt, anzunehmen, dass dieses Argument erklärt, warum das Gedankenexperiment überzeugen kann. Letzterer Punkt geht über die bloße Negation der *Derivability Thesis* hinaus, indem darin implizit die Behauptung zu stecken scheint, dass auch ein gedankenexperimentelles Argument, das gültig ist und dessen Prämissen überzeugungskräftig sind, keine Erklärung für die Überzeugungskraft des Gedankenexperiments ist. Letzterem haben wir im letzten Kapitel widersprochen. Da Gendler tatsächlich nur für die Negation der *Derivability Thesis* argumentiert und nur in ihren „Conclusions“ weiter reichende Schlüsse zieht, besteht kein Grund, diese weitere These einer nochmaligen Untersuchung zu unterziehen.

Was wir uns stattdessen etwas genauer ansehen werden, ist die Argumentation gegen die *Dispensability Thesis*. Wir werden die Nichteliminierbarkeit von Gedankenexperimenten dabei zunächst an Gendlers Beispiel der Galileischen Widerlegung des Aristotelischen Fallgesetzes erläutern und dann am Fall von Stevins Gedankenexperiment zum Prinzip der schiefen Ebene.

#### 3.3.2.1 Beispiel Galilei

Obwohl Galileis Argument gegen das Aristotelische Fallgesetz für Wissenschaftstheoretiker *das* Paradebeispiel für ein gutes Gedankenexperiment ist, ist seine Interpretation unter Wissenschaftshistorikern alles andere als unkontrovers. Es ist zunächst nicht klar, was das Gedankenexperiment eigentlich zeigen soll. Es ist

---

<sup>134</sup> Ähnliche Überlegungen finden sich bei Arthur [3].

auch nicht klar, welche Voraussetzungen es macht. Wir werden diese Probleme hier nicht in aller erforderlichen Gründlichkeit behandeln können. In erster Linie soll es sich hier ja auch um ein Beispiel handeln, das illustriert, wie die These von Gendler zu verstehen ist, weshalb wir nicht behaupten müssen, dass die folgende Darstellung das letzte Wort zur wissenschaftshistorischen Interpretation dieser Episode ist. Dennoch sind einige Bemerkungen vonnöten.

Das Gedankenexperiment von Galilei, das am „ersten Tag“ in *Unterredungen und mathematische Demonstrationen über zwei neue Wissenszweige, die Mechanik und die Fallgesetze betreffend* von dem uns bereits bekannten Salviati vorgetragen wird, soll – nach Gendlers Interpretation – gegen „das Aristotelische Fallgesetz“ gerichtet sein. Die Targetthese ist demnach diese:

- (T) Die natürliche Geschwindigkeit eines fallenden Körpers ist linear proportional zu seinem Gewicht.

Hierzu sind drei Dinge zu bemerken:

Erstens ist nicht klar, in welchem Sinne dieses Gesetz von irgendjemandem so vertreten worden ist, wie es hier vorkommt, nämlich als ein quantitatives Gesetz, das sich dazu eignet, die Differenz in der Fallgeschwindigkeit zweier Körper auf Grund der Differenz ihres Gewichts zu prognostizieren.

Bei Aristoteles selbst wird diese Thematik unter anderem in *De Caelo* und in *Physica* diskutiert. Es scheint offensichtlich, dass sich Galilei insbesondere auf die Diskussion in *Physica* bezieht. Dort diskutiert Aristoteles Argumente, die gegen die Annahme eines Vakuums sprechen. In der Tat sagt er dabei auch Folgendes:

We observe that bodies which have a preponderance in weight or lightness, if alike in other ways, travel faster over equal space intervals, and in the ratio which the magnitudes have to each other. So they should be traveling in the void in the same way. (Apostle [1], 75; Übersetzung von Aristoteles, *Physica*, Δ, 216a, 13-17)

Der Argumentationszusammenhang, in dem Aristoteles sein „Fallgesetz“ formuliert, lässt aber Zweifel zu, ob es sich hierbei um ein Gesetz handelt, das er selbst vertreten möchte<sup>135</sup>, oder nur um ein Gesetz, das seine Gegner vertreten<sup>136</sup>, oder ob es sich hier nicht um eine Nominaldefinition<sup>137</sup>, also gar kein Gesetz handelt.

Es ist auch nicht klar, ob zu Galileis Zeiten dieses Gesetz überhaupt noch größere Anerkennung genossen hat. Simon Stevin berichtet bereits 1586 von Experimenten, die dieses Fallgesetz empirisch widerlegen.<sup>138</sup> Galilei berichtet auch selbst von solchen Experimenten und zitiert sie durch Sagredo unmittelbar vor (!) dem Gedankenexperiment.

Zweitens kommt das Gedankenexperiment auch in Galileis Dialog im Zusammenhang mit Argumenten gegen die mögliche Existenz eines Vakuums vor. Aris-

<sup>135</sup> Vgl. z.B. Duhem [93], 361; Sambursky [286], 130f.

<sup>136</sup> Vgl. Wolff [353], 15ff.

<sup>137</sup> Vgl. Kühne [175], 386.

<sup>138</sup> Vgl. Kühne [175], 380.

toteles hatte geschlossen, dass auch verschieden schwere Körper im Vakuum gleich schnell fallen müssten, weil es für sie dort keinen Grund mehr gebe, mit unterschiedlichen Geschwindigkeiten zu fallen. Dies widerspricht aber der vorausgesetzten Annahme des Fallgesetzes, ist also absurd. Es wird demnach gerade das gleichschnelle Fallen verschieden schwerer Körper im Vakuum zum Nachweis einer Inkonsistenz in der Annahme des Vakuums verwendet. Wenn Galilei auf dieses Argument antworten will (und so scheint es), dann kann Galilei nicht nur den Nachweis erbringen wollen, dass oben zitiertes Fallgesetz für Körper im Vakuum falsch ist (das ist ja gerade der Punkt des Aristoteles).<sup>139</sup>

Drittens ist nicht klar, ob Galileis Gedankenexperiment ausschließlich die Widerlegung dieser Targetthese zur Konsequenz haben soll, oder auch positiv durch das Gedankenexperiment gezeigt werden soll, dass alle Körper im Vakuum gleich schnell fallen, also ein alternatives Fallgesetz *etabliert* werden soll.<sup>140</sup>

Wir wollen von diesen Schwierigkeiten an dieser Stelle aber absehen und Galileis Gedankenexperiment im Sinne Gendlers als Versuch der Widerlegung von (T) verstehen.

Die Textstelle, in der Galilei sein Gedankenexperiment von Salviati vortragen lässt, liest sich folgendermaßen:

*Salv.* Ohne viel Versuche können wir durch eine kurze, bindende Schlussfolgerung nachweisen, wie unmöglich es sei, dass ein grösseres Gewicht sich schneller bewege, als ein kleineres, wenn beide aus gleichem Stoff bestehen; und überhaupt alle jene Körper, von denen Aristoteles spricht. [...]

Wenn wir zwei Körper haben, deren natürliche Geschwindigkeit verschieden sei, so ist es klar, dass, wenn wir den langsameren mit dem geschwindigeren vereinigen, dieser letztere von jenem verzögert werden müsste, und jener, der langsamere, müsste vom schnelleren beschleunigt werden. [...]

Aber wenn dies richtig ist, und wenn es wahr wäre, dass ein [schwerer] Stein sich z.B. mit 8 Maass Geschwindigkeit bewegt, und ein [leichterer] Stein mit 4 Maass, so würden beide vereinigt eine Geschwindigkeit von weniger als 8 Maass haben müssen, aber beide Steine zusammen sind doch [schwerer], als jener [schwerere] Stein war, der 8 Maass Geschwindigkeit hatte; mithin würde sich nun der [schwerere] langsamer bewegen, als der [leichtere]; was gegen Eure Voraussetzung wäre. Ihr seht also, wie aus der Annahme, ein [schwerer] Körper habe eine grössere Geschwindigkeit als ein [leichterer] Körper, ich Euch weiter folgern lassen konnte, dass ein [schwererer] Körper langsamer sich bewege als ein [leichterer]. (Galilei [116], 57-58)

Es scheint, als sollte sich dieses Gedankenexperiment ohne weiteres als Argument wiedergeben lassen. Versucht man es im Sinne Nortons als nicht-gedankenexperimentelles Argument zu rekonstruieren, gelangt man zu folgender Rekonstruktion<sup>141</sup>:

<sup>139</sup> Vgl. Norton [239], 20.

<sup>140</sup> Vgl. Brown [39], Atkinson und Peijnenburg [6].

<sup>141</sup> Vgl. Gendler [120], Gendler [121], vgl. aber auch die Diskussion dieser Analyse in Atkinson und Peijnenburg [6] und Schrenk [288].

- (1) Natürliche Geschwindigkeit ist „mediativ“.
- (2) Schwere ist „additiv“.
- 
- (3) Natürliche Geschwindigkeit ist nicht direkt proportional zur Schwere.

Prämisse (1) besagt, dass wenn ein Körper A eine natürliche Geschwindigkeit  $g_1$  besitzt und ein Körper B eine natürliche Geschwindigkeit  $g_2$ , die natürliche Geschwindigkeit eines aus A und B zusammengesetzten Körpers zwischen  $g_1$  und  $g_2$  liegt.

Prämisse (2) hingegen besagt, dass die Schwere eines so zusammengesetzten Körpers sich als Summe der Schwere der Teile A und B ergibt.

Der Aristoteliker kann aber sein Fallgesetz, dass Körper proportional zu ihrem Gewicht fallen, aufrecht halten, wenn er (1) oder (2) zurückweist. Gendler gibt vier Möglichkeiten an, wie der Aristoteliker zu einer konsistenten Verteidigung seines Fallgesetzes kommen könnte:<sup>142</sup>

- (V1) Natürliche Geschwindigkeit ist für zusammengesetzte Körper nicht determiniert.
- (V2) Schwere ist für zusammengesetzte Körper nicht determiniert.
- (V3) Natürliche Geschwindigkeit und Schwere sind mediativ für Körper, die *verbunden* sind, und additiv für Körper, die *vereinigt* sind.
- (V4) Natürliche Geschwindigkeit und Schwere werden durch einen Grad der Verbundenheit (C) bestimmt (wobei C einen Wert zwischen 0 und 1 annehmen kann, 1 für völlig „vereinigte“ Körper, 0 für bloß „verbundene“ Körper), so dass die Geschwindigkeit bzw. Schwere von einem System aus den Körpern  $K_1$  und  $K_2$  mit den jeweiligen Schweren bzw. natürlichen Geschwindigkeiten  $s_1$  und  $s_2$  durch folgende Formel wiedergegeben wird:  

$$(C)(s_1+s_2) + (1-C)((s_1+s_2)/2)$$

In (V1) und (V2) wird die Applizierbarkeit des Gesetzes auf den Fall zusammengebundener Körper bestritten. In (V3) und (V4) wird das Gesetz erweitert, wobei in (V4) durch die Einführung einer neuen physikalischen Eigenschaft vermieden wird, starke Diskontinuitäten in der Natur anzunehmen.

Laut Gendler müsste eine erfolgreiche Elimination des Gedankenexperiments diese Auswege für den Aristoteliker versperren. Wenn dies aber in einer Weise geschehen soll, die ohne Bezugnahme auf (aktuelle oder mögliche) Einzeldinge und deren Eigenschaften auskommt, müssen zusätzliche universale Annahmen ge-

---

<sup>142</sup> Vgl. Gendler [121], 43-44. Vgl. auch Schrenk [288], der ebenfalls nachweist, dass dem Aristoteliker verschiedene Auswege offen stehen (ebenso Atkinson und Peijnenburg [6] und Norton [239]).

macht werden. Zu diesen Annahmen könnten beispielsweise die folgenden gehören (da sie (V1)-(V4) blockieren können):

- (A) Die natürliche Geschwindigkeit und das Gewicht von physikalischen Körpern sind immer physikalisch determiniert.
- (B) Die Entifizierung (entification) von physikalischen Körpern ist nicht physikalisch bestimmt, d.h. die physikalischen Gesetze sind invariant gegenüber der Frage ob etwas ein Teil eines Gegenstandes ist oder nicht.

(A) besagt, dass jeder physikalische Körper, wie auch immer seine Form sein mag, ein bestimmtes Gewicht und eine bestimmte, fixierte natürliche Geschwindigkeit hat. (B) besagt hingegen, dass physikalisch unbestimmt ist, ob wir einen zusammengebundenen Körper als *einen* Körper oder als *zwei* Körper, die durch eine Verbindung aneinander gebunden sind, oder als unbestimmt viele Körper, die durch interne Kräfte zusammen gehalten werden, betrachten. Die Frage danach, um wie viele Gegenstände es sich handelt, ist demnach keine Frage, deren Antwort durch irgendeine nachweisbare physikalische Eigenschaft festgelegt wäre.

Nimmt man (A) und (B) an, lässt sich Galileis Gedankenexperiment durch ein Argument ersetzen. Allerdings ist der Überzeugungsgrad, den man (A) und (B) insbesondere zu Galileis Zeiten entgegenbringen konnte, deutlich geringer als der Überzeugungsgrad gegenüber der Konklusion des Gedankenexperiments.

Dies liegt daran, dass der von Galilei attackierte Aristoteliker zwar *so etwas wie* (A) und (B) ebenfalls vertreten muss, aber nicht unbedingt genau (A) und (B). In einem expliziten Argument könnte der Aristoteliker die Annahme von (A) und (B) problematisieren und sich von Galileis Argument entsprechend unbeeindruckt zeigen. Das Besondere an Galileis Gedankenexperiment ist gerade, dass die Annahmen (A) und (B) im hypothetischen Fall spezialisiert werden. Die Spezialisierung erfolgt auf eine Situation, in der (A) und (B) für unproblematisch erachtet werden können. Es mag sein, dass sich an (A) und (B) in ihrer allgemeinen Form Zweifel formulieren lassen, aber für den von Galilei beschriebenen hypothetischen Fall lassen sich solche Zweifel eben nicht formulieren.

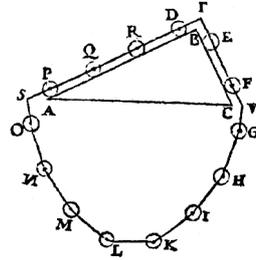


Abbildung 3.3-1

### 3.3.2.2 Beispiel Stevin

Ähnliche Überlegungen lassen sich auf den Fall des „Kugelkranzbeweises“ durch Simon Stevin (1586) anwenden. Mit diesem Gedankenexperiment etabliert Stevin das Gesetz der schiefen Ebene in seiner Grundlegung der Statik, *De Beghinse-len der Weeghconst.*<sup>143</sup> Man stelle sich ein Prisma vor, das durch das Dreieck ABC repräsentiert wird, dessen Grundseite AC parallel zum Horizont ist, und dessen Seite AB doppelt so lang sei wie die Seite BC (vgl. Abbildung 3.3-1<sup>144</sup>). Stevin möchte zeigen, dass die Kräfte, die benötigt werden, jeweils eine gegebene Kugel E auf der Ebene BC und eine gleich schwere und gleich große Kugel D auf der Ebene AB in Ruhe zu halten, im selben Verhältnis zueinander stehen, wie die Längen der jeweiligen Ebenen.

Dazu stelle man sich eine Kette von 14 Kugeln vor, die alle gleich groß und gleich schwer sind, im gleichen Abstand zueinander aufgereiht sind und frei in Richtung der Kette rollen können. Diese Kette wird nun über das Prisma gelegt, so dass vier der Kugeln auf AB zu liegen kommen, zwei auf BC und die restlichen 8 Kugeln unterhalb von AC „durchhängen“. Betrachten wir nur die Kugeln auf den Ebenen, so kann die Zugkraft, die die Gruppe der vier Kugeln auf AB auf die zwei Kugeln auf BC ausübt, gleich der Zugkraft in der umgekehrten Richtung oder kleiner oder größer als diese sein. Angenommen sie wäre größer, so würden die Kugeln von BC auf die Ebene AB „übergezogen“, angenommen sie wäre kleiner, wäre dies umgekehrt. In beiden Fällen würden die unten durchhängenden Kugeln nachrücken und man wäre wieder in der Ausgangssituation, dass sich auf der einen Ebene vier und auf der anderen Ebene zwei Kugeln befinden. Die Annahme, dass die Zugkräfte ungleich sind, würde damit zu einer ewig nicht endenden Bewegung führen, was als *falsch* angenommen wird („t'welck valsch is“). Da die Zugkräfte der durchhängenden Kugeln (G-O) auf die Kugeln auf den Ebenen

<sup>143</sup> Wir haben dieses Gedankenexperiment schon in der Darstellung von Mach kennen gelernt, vgl. Abbildung 2.1-2. Die Darstellung in diesem Kapitel orientiert sich am Text von Stevin. Vgl. Stevin [316], 175-176.

<sup>144</sup> Aus Stevin [316], 176.

je gleich sind, können diese aus dem Modell abgezogen werden. Folglich stehen die Kräfte, die benötigt werden, jeweils eine gegebene Kugel E auf der Ebene BC und eine gleich schwere und gleich große Kugel D auf der Ebene AB in Ruhe zu halten, im selben Verhältnis zueinander, wie die Längen der jeweiligen Ebenen.

Auch bei diesem Gedankenexperiment ist es fraglich, ob es sich ohne weiteres in ein nicht-gedankenexperimentelles Argument überführen lässt. Eine zentrale Prämisse einer solchen Rekonstruktion wäre sicherlich so etwas wie

(C) Ein Perpetuum mobile ist unmöglich.

Dieses Prinzip wurde ja auch bei Machs Rekonstruktion angeführt. Zu Stevins Zeiten (also um 1586) war dies als eine allgemeine Überzeugung in der Physik nicht etabliert.<sup>145</sup> Das Verbot eines Perpetuum mobile taucht außerdem nicht in den Postulaten Stevins auf.<sup>146</sup> Wieder scheint es so zu sein, als ob der Überzeugungsgrad, den man (C) zu Stevins Zeiten entgegenbringen konnte, deutlich geringer ist als der Überzeugungsgrad gegenüber der Konklusion des Gedankenexperiments.

Betrachtet man diesen Fall aber in Analogie zum vorher behandelten Gedankenexperiment Galileis, so scheint auch in diesem Fall das Gedankenexperiment seine Überzeugungskraft daraus zu ziehen, dass diese problematische Hintergrundannahme auf einen unproblematischen Einzelfall spezialisiert wird. Was auch immer im Allgemeinen zum Thema Perpetuum mobile zu sagen sein mag, so sieht jedenfalls keines aus. Stevin konnte nicht ausschließen, dass man irgendwie ein Perpetuum mobile konstruieren kann, aber er konnte ausschließen, dass man es mit einer Kette und einem Prisma konstruieren kann.

Dass eine rationale Rekonstruktion von Gedankenexperimenten, die Gedankenexperimente durch Argumente ersetzt, zumindest manchmal nicht möglich ist, liegt also daran, dass versteckte Hintergrundannahmen in die Darstellungen eingehen, die zum Zeitpunkt der Durchführung des Gedankenexperiments als allgemeine Aussagen nicht gerechtfertigt werden konnten, von denen die jeweiligen Autoren aber annehmen konnten, dass sie spezialisiert auf den kontrafaktischen Fall nicht problematisch sind.<sup>147</sup>

<sup>145</sup> Vgl. hierzu auch Kühne [175], 357-365.

<sup>146</sup> Vgl. Kühne [175], 327; Stevin [316].

<sup>147</sup> Die besondere Funktion von Gedankenexperimenten ist damit an bestimmte historische Situationen geknüpft, zu denen eine bestimmte Wissensbasis vorlag. Diese historische „Relativität“ von Gedankenexperimenten, die auch Mach schon betonte, ist aber nicht zu verwechseln mit der historischen Relativitätsthese von Gedankenexperimenten, derzufolge Gedankenexperimente erst ab Galilei und seiner Auffassung von einem physikalischen Phänomen möglich waren. Zu letzterer These vgl. McAllister [202], McAllister [203]. Man könnte sich fragen, ob es sich bei der Nicht-Eliminierbarkeit tatsächlich um ein bemerkenswertes Charakteristikum von Gedankenexperimenten handelt, oder ob es sich hierbei nicht eigentlich um ein bereits bekanntes Charakteristikum enthymematischer Argumente handelt (vgl. etwa die Diskussion in Crick [80]). *Enthymematischen* Argumenten fehlt mindestens eine Prämisse, um formal gültig zu sein. Ebenso scheint der nicht-gedankenexperimentellen Argumentrekonstruktion mindestens eine Prämisse zu fehlen.

## 3.3.3 HINTERGRUNDANNAHMEN

Wenn die Auffassung Gendlers insofern richtig ist, dass manche Gedankenexperimente nicht eliminierbar sind, und dies der Fall ist, weil es sich um Argumente handelt, deren (möglicherweise versteckten) Prämissen gerade deswegen als gerechtfertigt und überzeugungskräftig erscheinen, weil sie auf einen unproblematischen Fall spezialisiert werden, dann kann man unsere bisherigen methodologischen Überlegungen erweitern:

Um erfolgreich zu sein, müsste ein Gedankenexperiment zunächst nachweisen, dass der betrachtete Fall in den Anwendungsbereich der Theorie fällt, oder – wie im Fall Stevin – nachgewiesen werden, dass der beschriebene Fall eine Generalisierung zulässt.<sup>148</sup> Beides ist wichtig, weil Gedankenexperimente gerade deshalb ein nützliches Mittel zu einem bestimmten Zeitpunkt im Forschungsprozess darstellen, wenn es keine bekannten allgemeinen Prinzipien gibt, die die Bezugnahme auf den besonderen Fall im Gedankenexperiment ersetzen könnten. Ebenso sollten Kritiken, die ein Gedankenexperiment angreifen, nicht bloß in dem Nachweis bestehen, dass eine – zur eliminativen Rekonstruktion nötige – allgemeine Annahme falsch ist, sondern dass sie spezialisiert auf den im Gedankenexperiment geschilderten Fall falsch ist.

Schließlich sollte der Rat zur Abstinenz von Gedankenexperimenten überdacht werden. Wenn es Phasen im Forschungsprozess gab, in denen ein Gedankenexperiment zu gerechtfertigten Überzeugungsänderungen geführt hat, und es zu diesem Zeitpunkt keine Möglichkeit gab, diese Änderung auf eine alternative Weise zu erzielen, handelt es sich offenbar um eine fruchtbare Methode, auf die nicht

---

Nun ist es bei enthymematischen Argumenten in der Regel nicht leicht, die fehlende Prämisse zu ergänzen – es sind häufig mehrere, verschieden starke Prämissen möglich, um das enthymematische Argument zu einem gültigen Argument zu ergänzen, genauso ist es bei der nicht-gedankenexperimentellen Argumentrekonstruktion nicht leicht, die fehlende Prämisse zu ergänzen; wie wir gesehen haben gibt es auch dort verschieden starke Optionen. Im Gegensatz zu nicht-gedankenexperimentellen Enthymemen sind *Gedankenexperimente* keine Argumente, denen einfach eine Prämisse fehlt. Die nicht-gedankenexperimentellen Rekonstruktionen, die John Norton im Auge hat, sind Argumente, denen einfach eine Prämisse fehlt. Wir haben in diesem Unterkapitel gesehen, dass diese – dem allgemeinen Argument fehlende – Prämisse im Gedankenexperiment-Argument dadurch *geliefert* wird, dass ein Fall vorkommt, in dem diese Prämisse instantiiert ist. Das, was im enthymematischen Argument fehlt, wird im Gedankenexperiment also gerade geliefert. Das soll ja der besondere Witz von Gedankenexperimenten sein: Statt eines enthymematischen Arguments, bei dem Entscheidendes fehlt, wird ein Fall präsentiert, in dem der fehlende Baustein für ein allgemeines Argument auf unproblematische Weise vorkommt: eine Kette über einem Prisma ist kein Perpetuum mobile, zwei Kugeln kann ich noch so feste oder lose verbinden wie ich will, ohne dass dies Einfluss auf ihre Fallgeschwindigkeit haben könnte.

<sup>148</sup>In Stevins Fall scheint eine Verallgemeinerung deswegen möglich, weil (a) das Gesetz für die Extrema die richtigen Werte liefert (also für den Fall einer horizontalen Ebene und einer genau vertikalen), (b) der Fall zwischen diesen Extrema arbiträr gewählt wurde. Man beachte, dass dieses Ergebnis durch eine empirische Untersuchung eines beliebigen Falls zwischen den Extrema nicht so hätte erzielt werden können (jedenfalls nicht ohne die Zusatzannahme, dass in den Limit-Bereichen keine chaotischen Verhältnisse auftreten. Vgl. zu solchen Problemen auch Batterman [13].

verzichtet werden sollte. Näheres zu der Frage, unter welchen Umständen Abstinenz ratsam ist und wann nicht, werden wir aber erst am Ende dieser Untersuchung beantworten können.

## 4. Gedankenexperimente in der Philosophie

A boy is about to go on his first date, and is nervous what to talk about. He asks his father for advice. The father replies: "My son, there are three subjects that always work. The famous three 'F'. These are 'Food', 'Family', and ... 'Philosophy'."

The boy picks up his date and they go to a soda fountain. Ice cream sodas in front of them, they stare at each other for a long time, as the boy's nervousness builds. He remembers his father's advice, and chooses the first topic. He asks the girl: "Do you like potato pancakes?" She says "No", and the silence returns.

After a few more uncomfortable minutes, the boy thinks again of his father's suggestion and turns to the second item on the list. He asks, "Do you have a brother?" Once more, the girl says "No" and there is silence again.

The boy then plays his last card. He thinks of his father's advice and asks the girl: "If you had a brother, would he like potato pancakes?"

Imaginäre, kontrafaktische Fälle kommen in der Philosophie häufig vor und werden dort ebenfalls häufig als ‚Gedankenexperimente‘ bezeichnet. Auch diese Verwendungsweise des Ausdrucks mag mehr oder weniger berechtigt sein. Manche Autoren verwehren sich strikt dagegen, die „Gedankenexperimente“ in der Philosophie in Zusammenhang mit Gedankenexperimenten in den Naturwissenschaften zu betrachten.

George Bealer hält eine solche Bezeichnung für eine philosophische Tätigkeit für fehl am Platze, da es die methodologische Autonomie der Philosophie untergräbt und zu begrifflichen Verwirrungen führt:

In recent philosophy there has been an unfortunate blurring of traditional terminology. Rational intuitions about hypothetical cases are often being erroneously called thought experiments. [...] [T]o call [such cases] thought experiments is, not only to invite confusion about philosophical method, but to destroy a once useful term. (Bealer [14], 207-208)

Aber nicht nur aus Besorgnis um die Autonomie der philosophischen Methodologie, auch in (umgekehrter) Abgrenzung der Naturwissenschaften von der Philosophie wird diese Bezeichnungsweise kritisiert:

In Disziplinen wie der Moralphilosophie, die bisher nie Experimente benutzt hatten, mag es wohl als bloße begriffliche Verwirrung abgetan werden, daß man dort neuerdings behauptet, mit „Gedankenexperimenten“ zu forschen. [...] In den nicht empirischen Wissenschaften sind „Gedankenexperimente“ eine *Anmaßung*, Ergebnisse mit der Überzeugungsstärke eines physikalischen Experiments zu gewinnen [...]. (Kühne [175], 2)

Andere Autoren wiederum denken, dass diese Bezeichnungsweise berechtigt ist und es tatsächlich in Philosophie und Naturwissenschaften eine solche gemeinsame Methode gibt.<sup>149</sup> Es mag zwar sein, dass sich philosophische und naturwissenschaftliche Gedankenexperimente graduell unterscheiden, aber methodologisch betrachtet seien sie gleich.

Entsprechend könnten methodologische Einsichten, die man in Bezug auf erfahrungswissenschaftliche Gedankenexperimente gewonnen hat, auch auf philosophische Gedankenexperimente übertragen werden. In der Regel geschieht dies in der Hoffnung, die Methodologie der Philosophie zu *verbessern*. Man ist häufig der Auffassung, dass Gedankenexperimente in den Wissenschaften zwar zu klaren Ergebnissen führen, Gedankenexperimente in der Philosophie hingegen mit hoher Wahrscheinlichkeit nur eine weitere end- und fruchtlose Debatte lostreten.<sup>150</sup>

Nachdem wir in den letzten beiden Teilen einen Eindruck davon bekommen haben, was naturwissenschaftliche Gedankenexperimente sind, werden wir uns in diesem Kapitel den Namensvettern aus der Philosophie zuwenden. Dabei werden wir zunächst die Frage untersuchen, ob die Charakteristika für Gedankenexperimente, die wir im letzten Teil zusammengestellt haben, auch auf die Beispiele in der Philosophie zutreffen. Im nächsten thematischen Hauptteil (Teil 5-9) werden wir uns dann ansehen, was Gedankenexperimenten in der Philosophie ihren schlechten Ruf eingebracht hat.

---

<sup>149</sup> Vgl. Sorensen [304], Häggqvist [137], Wilkes [345], Gähde [113], Peijnenburg und Atkinson [252], Gendler [121].

<sup>150</sup> Vgl. Wilkes [345], Peijnenburg und Atkinson [252].

#### 4.1 GEDANKENEXPERIMENTE IN DER THEORETISCHEN PHILOSOPHIE

In diesem Kapitel wollen wir einige paradigmatische Gedankenexperimente aus der theoretischen Philosophie vorstellen und prüfen, inwiefern sie die Charakteristika erfüllen, die wir in Teil 3 zusammenstellen konnten. Wir werden uns ansehen, ob sie von ihrer Argumentstruktur zu unserem Schema aus 3.1 passen, ob sie dazu intendiert sind, Überzeugungsrevisionen herbeizuführen, wie sie dies ggf. bewerkstelligen, etc.<sup>151</sup>

##### 4.1.1 SEARLES CHINESISCHES ZIMMER

Beginnen wir gleich mit einem der bekanntesten „Gedankenexperimente“ der analytischen Philosophie, dem „Chinesischen Zimmer“ von Searle.

Searle richtet sich in ‚Geist, Gehirn, Programm‘ (Searle [294]) gegen die so genannte „starke AI“ (für ‚artifizielle Intelligenz‘, auch ‚KI‘ für ‚künstliche Intelligenz‘). Dabei geht es ihm um die Frage, welchen Status Computersimulationen menschlicher kognitiver Fähigkeiten zugesprochen werden soll. Charakteristisch für das Forschungsprogramm der starken AI sind dabei folgende funktionalistische Thesen:

- (AI<sub>1</sub>) Jede Maschine, deren Verhalten allein dadurch bestimmt ist, dass in ihr ein (syntaktisch spezifiziertes) Computerprogramm implementiert ist, und die den Turing-Test besteht, *besitzt* kognitive Fähigkeiten.
- (AI<sub>2</sub>) Rein syntaktisch spezifizierte Computerprogramme können als Erklärungsmodell für menschliche kognitive Fähigkeiten dienen.

AI<sub>2</sub> bedeutet, dass wir Fragen nach der Funktionsweise des menschlichen Geistes dadurch beantworten können, dass wir die Funktionsweise eines Computerpro-

---

<sup>151</sup> Um die Geduld des Lesers nicht über Gebühr zu strapazieren, wird dies mit abnehmender Ausführlichkeit geschehen. Was wir hier nicht vorhaben, ist diese Gedankenexperimente hier inhaltlich zu bewerten. Im Gegensatz zu den Namensvettern in den Naturwissenschaften (abgesehen von Maxwells Dämon in seiner modernen Spielart), sind einige der hier vorgestellten Gedankenexperimente noch in der Diskussion, und es würde zu weit führen, die jeweils relevanten Hintergrundinformationen bereitzustellen. Nach welchen Kriterien man philosophische Gedankenexperimente bewerten sollte, werden wir ohnehin erst in Kapitel 9 beantworten können. (Wer also nun gespannt zum Abschnitt zu „Gehirnen im Tank“ vorblättert, um dort zu erfahren, ob er eins ist, oder möglicherweise eins ist, oder notwendigerweise keins ist, wird hier enttäuscht werden.)

gramms zitieren, das den menschlichen Geist (im Sinne des Turing Tests) simuliert. Der „Turing Test“ besteht hierbei schlicht darin, ob ein menschlicher Beobachter, der nur Informationen über sprachlichen Input und Output seines Gegenübers hat, feststellen kann, ob es sich um einen Menschen oder einen Computer handelt. Nach der ursprünglichen Intention Alan Turings<sup>152</sup> soll durch diesen Test überprüfbar werden, ob eine Maschine kognitive Fähigkeiten besitzt. AI<sub>1</sub> kann also auch als die These verstanden werden, dass der Turing-Test tatsächlich ein Test für das Vorliegen kognitiver Fähigkeiten bei einer Maschine ist: „Täuschend“ echt menschliche kognitive Fähigkeiten zu simulieren, ist streng genommen keine „Täuschung“ mehr, die Maschine *besitzt* dann diese Fähigkeiten. Wenn AI<sub>1</sub> wahr ist, ist AI<sub>2</sub> offensichtlich ebenfalls wahr (nicht aber unbedingt umgekehrt).

Searles „Gedankenexperiment“ vom Chinesischen Zimmer soll gegen diese Thesen sprechen:

[S]tellen Sie sich vor, Sie wären in ein Zimmer eingesperrt, in dem mehrere Körbe mit Chinesischen Symbolen stehen. Und stellen Sie sich vor, daß Sie (wie ich) kein Wort Chinesisch verstehen, daß Ihnen allerdings ein auf Deutsch verfaßtes Regelwerk für die Handhabung dieser Chinesischen Symbole gegeben worden wäre. Die Regeln geben rein formal – nur mit Rückgriff auf die Syntax und nicht die Semantik der Symbole – an, was mit den Symbolen gemacht werden soll. Eine solche Regel mag lauten: ‚Nimm ein Kritzel-Kratzel-Zeichen aus Korb 1 und lege es neben ein Schnörkel-Schnarkel-Zeichen aus Korb 2‘. Nehmen wir nun an, daß irgendwelche anderen Chinesischen Symbole in das Zimmer gereicht werden, und daß Ihnen noch zusätzliche Regeln dafür gegeben werden, welche Chinesischen Symbole jeweils aus dem Zimmer herauszureichen sind. Die hereingereichten Symbole werden von den Leuten draußen ‚Fragen‘ genannt, und die Symbole, die Sie dann aus dem Zimmer herausreichen, ‚Antworten‘ – aber dies geschieht ohne Ihr Wissen. Nehmen wir außerdem an, daß die Programme so trefflich und Ihre Ausführungen so brav sind, daß Ihre Antworten sich schon bald nicht mehr von denen eines chinesischen Muttersprachlers unterscheiden lassen. (Searle [293], 31; Übersetzung Beckermann [18], 290-291)

Vergleicht man diese Situation damit, dass man Searle stattdessen Fragen in Englisch hereingereicht hätte, die er dann auf Englisch beantwortet und wieder herausgereicht hätte, stellt man einen Unterschied fest: Im ersten Fall, versteht Searle kein Chinesisch. Er versteht die „Fragen“ nicht und auch nicht die „Antworten“, die er gibt, obwohl dies für einen Chinesischen Muttersprachler außerhalb des Zimmers nicht zu erkennen ist. Im zweiten Fall ist dies klarerweise anders. Hier *versteht* er die Fragen und die Antworten in einem ganz offensichtlichen Sinn. Tatsächlich spricht dieses Argument gegen AI<sub>1</sub> (und unterminiert AI<sub>2</sub> damit nur indirekt (was Searle auch zugibt<sup>153</sup>)).

Dieses „Gedankenexperiment“ scheint unserer bisherigen Charakterisierung zu entsprechen. Searles Intention ist offenbar eine gerechtfertigte Überzeugungsänderung beim Adressaten auszulösen, wobei dieser ein Proponent der starken AI ist.

<sup>152</sup> Turing [330].

<sup>153</sup> Searle [294], 237.

Dass Searle der Auffassung ist, dass sein Gedankenexperiment zu einer *gerechtfertigten* Überzeugungsänderung führen müsste, lässt sich insbesondere daran ablesen, dass Searle sich die Mühe macht, Argumente, die gegen die Überzeugungskraft des Gedankenexperiments gerichtet sind, einzeln aufzuführen und im Detail zurückzuweisen.<sup>154</sup>

In der Tat wird auch ein Sachverhalt geschildert, der als vorstellbar dargestellt wird, und schließlich soll sein Argument als Widerlegung von AI<sub>1</sub> hinreichen, ohne dass der geschilderte Sachverhalt realisiert wird.

Es scheint ebenfalls klar, dass sich Searles Gedankenexperiment als Argument in die von Häggqvist vorgeschlagene Form bringen lässt:

Zunächst wird die Möglichkeit eines Chinesischen Zimmers etabliert, in dem eine Person bestimmte Zeichen erhält und wieder nach draußen reicht, nachdem sie sie entsprechend bestimmter Vorschriften manipuliert hat, und zwar so, dass ein Muttersprachler denkt, die Person im Zimmer verstünde Chinesisch. Für diese Situation folgt nun aus AI<sub>1</sub>, dass dieses symbolverarbeitende System die Fragen und Antworten tatsächlich versteht. In dieser Situation ist diese Konsequenz aber absurd. Searle würde englische Fragen und Antworten „verstehen“, in diesem Sinne „verstehen“ er die chinesischen jedoch nicht. Aus dieser Absurdität wird auf die Negation der Targetthese, AI<sub>1</sub>, geschlossen.

Außerdem scheint es auch in diesem Fall so zu sein, dass eine universal schwer zu etablierende These hier auf einen (zumindest vermeintlich) unproblematischen Fall spezialisiert wird, was erklärt, weshalb das Format des Gedankenexperiments gewählt wurde:

Meine Kritiker weisen darauf hin, daß es viele verschiedene Grade des Verstehens gibt; daß „Verstehen“ nicht einfach ein transitives Verb ist; daß es sogar verschiedene Arten und Ebenen von Verstehen gibt und häufig nicht einmal der Satz vom ausgeschlossenen Dritten eindeutig auf Feststellungen anwendbar ist, die die Form „x versteht y“ haben; daß es in vielen Fällen nicht eine einfache Tatsachen-, sondern eine Ermessensfrage ist, ob y von x verstanden wird; und so weiter. Zu all diesen Bemerkungen möchte ich sagen: Na klar, na klar; aber all das hat nichts mit dem zu tun, worum es hier geht. Es gibt klare Fälle, wo „Verstehen“ im Sinne des Wortes statthat, und ebenso klare Fälle, wo es nicht statthat; und nichts sonst als diese zwei Fälle brauche ich im Zusammenhang dieser Argumentation. (Searle [294], 239)

In einem nicht-gedankenexperimentellen Argument wäre die Vagheit des Ausdrucks ‚Verstehen‘ womöglich ein Problem. Spezialisiert auf den Fall im Gedankenexperiment, scheint aber klar, was mit ‚Verstehen‘ gemeint ist. So, wie wir von Searle sagen würden, dass er englische Fragen und Antworten „verstehen“ würde, würde er die chinesischen nicht „verstehen“, gleichgültig, wie dieser Verstehensbegriff korrekt expliziert aussieht.

---

<sup>154</sup> Vgl. Searle [294], 240-264.

## 4.1.2 MARY, DIE NEUROPHYSIOLOGIN

Wir bleiben zunächst in der Philosophie des Geistes und betrachten ein weiteres Zimmer, diesmal eines, das Frank Jackson entworfen hat, nicht für sich, sondern für die Neurowissenschaftlerin Mary.

Wie bereits an früherer Stelle erwähnt, ist der „Physikalismus“ eine Bezeichnung für eine Vielzahl metaphysischer Positionen. *Eine* solche Position besteht darin, anzunehmen, dass alle Tatsachen der Welt physikalische Tatsachen sind. In der Terminologie möglicher Welten bedeutet dies, dass ein minimales physikalisches Duplikat unserer Welt ein Duplikat unserer Welt *simpliciter* ist. Wenn zwei mögliche Welten also hinsichtlich ihrer Naturgesetze und physikalisch spezifizierten Anfangs- und Randbedingungen identisch sind, sind sie auch in allen anderen Hinsichten identisch.<sup>155</sup> In diesem Sinne sind alle Tatsachen physikalische Tatsachen.

In ‚Epiphenomenal Qualia‘ (Jackson [157]) richtet sich Frank Jackson gegen den Physikalismus, also gegen

(P) Alle Tatsachen sind physikalische Tatsachen.

Sein mittlerweile recht berühmt gewordenes Gedankenexperiment handelt von Mary, einer brillanten Neurowissenschaftlerin, die seit ihrer Geburt gezwungen ist, in einem schwarz-weiß-grauen Raum zu leben und nie Farben gesehen hat. Sie ist allerdings Expertin für Wahrnehmungsphysiologie und besitzt *alle* physikalischen Informationen darüber, was vorgeht, wenn Menschen mit normalem Wahrnehmungsapparat farbige Objekte (wie beispielsweise rote Gegenstände) wahrnehmen.

Eines Tages wird Mary nun aus ihrer seltsamen Situation entlassen und sie sieht zum ersten Mal eine reife Tomate. Bei diesem ersten Farbeindruck ihres Lebens bildet sie die Überzeugung: ‚So ist es also, wie es ist, Rot zu sehen‘. Mary hat dadurch neues Wissen erworben. Offenbar handelt es sich dabei nicht um Wissen über eine neu entstandene Tatsache, sondern um eine Tatsache, die auch schon vor diesem Ereignis bestand, dass es *so* ist, wie es ist, Rot zu sehen.

Es scheint ganz offensichtlich, dass sie etwas über die Welt und über unsere visuelle Erfahrung der Welt lernt. Aber dann ist der Schluß unausweichlich, dass ihr bisheriges Wissen unvollständig war. Auf der anderen Seite hatte sie aber *alle* physikalischen Informationen. *Also* kann man mehr als diese Informationen haben; und der Physikalismus ist falsch. (Jackson [157], 130; Übersetzung Beckermann [18], 390)

Auch hier ist die Intention Jacksons, eine gerechtfertigte Überzeugungsänderung beim Adressaten auszulösen, in diesem Fall beim Physikalisten. Mary wusste alle physikalischen Tatsachen. Sie hat etwas Neues gelernt, als sie eine rote Tomate gesehen hat. Diesem neuen Wissen entspricht eine Tatsache, die offenbar nicht zu

---

<sup>155</sup> Vgl. Jackson [158], 12. Diese Analyse wird in Kapitel 7.2 noch genauer erläutert.

den physikalischen Tatsachen zählt (sonst hätte sie sie ja gekannt), also ist die Auffassung, dass es nur physikalische Tatsachen gibt, falsch. Argumentiert wird wieder mit einem vorstellbaren Sachverhalt und Jackson scheint nicht anzunehmen, dass man tatsächlich eine Neurowissenschaftlerin so aufwachsen lassen muss, um das Argument zu erhärten.

Wieder kann das Gedankenexperiment auch als Argument in Häggqvists Schema wiedergegeben werden: Zunächst wird die Möglichkeit etabliert, dass eine Neurowissenschaftlerin ohne Farbwahrnehmung aufwächst, die nach ihrem Studium in die Freiheit entlassen wird und zum ersten Mal eine Farbwahrnehmung hat. Aus dem Physikalismus wird für diese Situation abgeleitet, dass Mary durch dieses Erlebnis kein neues Tatsachenwissen gewinnen kann. Diese Prognose ist aber absurd: es ist ganz offensichtlich, dass Mary etwas Neues lernt, was sich auch nicht durch die Entstehung einer neuen Tatsache weganalisieren lässt. Folglich ist die Targetthese, dass alle Tatsachen physikalische Tatsachen sind, falsch.

Wenn wir auch hier danach fragen, welche Hintergrundannahmen es nahe gelegt haben mögen, die Form eines Gedankenexperiments zu wählen, so sind hier insbesondere Annahmen über Wissen und Tatsachenwissen zu nennen. ‚Alles Wissen ist Tatsachenwissen.‘, wie es eine Rekonstruktion des Arguments ohne Bezug auf den hypothetischen Fall anführen könnte, ist alles andere als unproblematisch. Bekanntlich gibt es auch nicht-propositionales Wissen (z.B. *knowing how*), weshalb diese mögliche Prämisse einfach falsch ist. Im hypothetischen Fall scheint es aber zunächst tatsächlich so zu sein, dass Mary neues propositionales Wissen erwirbt, dass sich auf eine ihr bisher unbekannte Tatsache bezieht (und nicht etwa nur auf einen neuen Aspekt einer ihr bisher bekannten Tatsache), etc.

Dass der Übergang von ‚Mary lernt etwas Neues.‘ zu ‚Mary lernt eine neue Tatsache.‘ nicht unbestritten geblieben ist<sup>156</sup>, ist kein Einwand dagegen, dass Jackson diesen Schluss für diesen Fall als unproblematisch erachtet hat. Wie wir sehen werden, betrifft die Frage nach der Gültigkeit dieses Übergangs die Frage, wie überzeugend dieses Gedankenexperiment ist. Diese Frage soll uns hier aber nicht beschäftigen. (Wir werden aber in 5.3 hierauf zurückkommen.)

#### 4.1.3 GETTIER-FÄLLE

Zur Abwechslung wollen wir die Philosophie des Geistes erst einmal verlassen und uns in der Erkenntnistheorie umsehen. Lange Zeit war es in der Erkenntnistheorie relativ unbestritten, dass eine Person P weiß, dass q, gdw. P die gerechtfertigte, wahre Überzeugung hat, dass q.<sup>157</sup> In ‚Is Justified True Belief Knowledge?‘ (Gettier [123]) wurde diese Analyse des Wissensbegriffs von Edmund L. Gettier angegriffen. Dazu beschrieb er zwei hypothetische Fälle. Der erste ist der Folgende:

<sup>156</sup> Vgl. Beckermann [18], 391, Perry [255].

<sup>157</sup> Vgl. Bremer und Cohnitz [34], 124-125; Audi [7], 214-216. Bereits Russell diskutiert in Russell [285] ähnliche Fälle (insbes. S. 170-171).

Suppose that Smith and Jones have applied for a certain job. And suppose that Smith has strong evidence for the following conjunctive proposition:

(d) Jones is the man who will get the job, and Jones has ten coins in his pocket.

Smith's evidence for (d) might be that the president of the company assured him that Jones would in the end be selected, and that he, Smith, had counted the coins in Jones's pocket ten minutes ago. Proposition (d) entails:

(e) The man who will get the job has ten coins in his pocket.

Let us suppose that Smith sees the entailment from (d) to (e), and accepts (e) on the grounds of (d), for which he has strong evidence. In this case, Smith is clearly justified in believing that (e) is true.

But imagine, further, that unknown to Smith, he himself, not Jones, will get the job. And, also, unknown to Smith, he himself has ten coins in his pocket. Proposition (e) is then true, though proposition (d), from which Smith inferred (e), is false. In our example, then, all of the following are true: (i) (e) is true, (ii) Smith believes that (e) is true, and (iii) Smith is justified in believing that (e) is true. But it is equally clear that Smith does not *know* that (e) is true; for (e) is true in virtue of the number of coins in Smith's pocket, while Smith does not know how many coins are in Smith's pocket, and bases his belief in (e) on a count of the coins in Jones's pocket, whom he falsely believes to be the man who will get the job. (Gettier [123])

Gettier argumentiert weiterhin, dass dieser Fall zeige, dass die Analyse des Wissensbegriffs, wie wir sie oben zitiert haben, keine hinreichenden Bedingungen für das Vorliegen von Wissen angibt, da alle Bedingungen in dem hypothetischen Fall erfüllt sind, obwohl es sich *nicht* um Wissen handelt.

Klarerweise zielt Gettier auf eine Überzeugungsänderung beim Adressaten des Gedankenexperiments, nämlich auf die Überzeugung, dass die Standardanalyse von ‚Wissen‘ hinreichende Bedingungen für das Vorliegen von Wissen angibt. Diese Änderung soll erreicht werden, indem ein hypothetischer Sachverhalt geschildert wird, und es scheint nicht so zu sein, dass Gettier davon ausgeht, dass man diesen Sachverhalt in der Realität erst nachstellen muss.

Von der Argumentstruktur her ist es wiederum so, dass eine bestimmte Situation als möglich dargestellt wird. Für diese Situation macht die kritisierte Überzeugung eine Prognose: Smith müsste Wissen, dass der Mann, der den Job bekommen wird, zehn Münzen in seiner Tasche hat. Das ist aber absurd: Smith glaubt ja, dass Jones den Job bekommt und glaubt nur deswegen, dass der, der den Job bekommt, zehn Münzen in der Tasche hat. Folglich kann auf die Negation der Targetthese geschlossen werden: die Standardanalyse des Wissensbegriffs gibt keine hinreichenden Bedingungen für das Vorliegen von Wissen an.

Am Spezialfall, der in diesem Gedankenexperiment betrachtet wird, wird deutlich, dass es eine Überzeugung, obwohl sie gerechtfertigt ist, nicht auf die richtige Weise zustande gekommen sein kann, um als Wissen zu zählen. Da der Begriff der Rechtfertigung selbst aus der Erkenntnistheorie stammt, ist dies mit universalen Prämissen, ohne Bezug auf den hypothetischen Fall, nicht nachweisbar, solange

man nicht sagen kann, was es bedeutet, dass eine Überzeugung auf „richtige“ Weise zustande gekommen ist. Diese Debatte wurde aber von Gettier erst initiiert. Auch hier ist also verständlich, warum die Form des Gedankenexperiments gewählt wurde.

#### 4.1.4 ZWILLINGSERDE

Betrachten wir als nächstes ein Beispiel aus der Sprachphilosophie. In seinem Artikel ‚Die Bedeutung der Bedeutung‘ (Putnam [263]) möchte Hilary Putnam die so genannte „traditionelle Bedeutungstheorie“ kritisieren. Diese Theorie charakterisiert er durch zwei Annahmen:

- (TB<sub>1</sub>) Um einen Ausdruck zu verstehen, muss man sich in einem bestimmten psychischen Zustand befinden.
- (TB<sub>2</sub>) Die Bedeutung eines Ausdrucks (im Sinne von ‚Intension‘) bestimmt seine Extension (d.h. aus Intensionsgleichheit folgt Extensionsgleichheit).

Putnam argumentiert dafür, dass (TB<sub>1</sub>) in der traditionellen Bedeutungstheorie so zu verstehen ist, dass einen Ausdruck zu verstehen, bedeuten soll, seine Intension zu kennen, und dass die psychischen Zustände, von denen in (TB<sub>1</sub>) gesprochen wird, im Sinne eines methodischen Solipsismus nur den psychischen Zustand desjenigen Subjektes betreffen, von dem wir sagen wollen, dass es einen Ausdruck versteht (die Intension eines Ausdrucks kennt). Dass die beiden Thesen so verstanden unzutreffend sind, will Putnam „mit Hilfe von etwas Sciencefiction“<sup>158</sup> zeigen.

Man stelle sich vor, es gäbe eine Zwillings-Erde (Zwerde), die unserer Erde in allen Hinsichten gleicht, bis auf die Tatsache, dass es auf der Zwerde kein H<sub>2</sub>O in Flüssen und Seen gibt, sondern eine andere klare, geschmack- und geruchslose Flüssigkeit deren (recht komplizierte) chemische Molekülstruktur durch ‚XYZ‘ abgekürzt wird. XYZ wird von den Bewohnern der Zwerde im Zwerdendeutsch auch ‚Wasser‘ genannt.

Ein Raumschiff von der Erde besucht nun die Zwerde, und zunächst sind die Erdlinge der Überzeugung, dass es sich bei der klaren Flüssigkeit, die sie in den Seen und Meeren der Zwerde finden und die die Zwerdlinge ‚Wasser‘ nennen, um Wasser, wie sie es kennen, handelt. Chemische Untersuchungen ergeben aber, dass es eine andere Molekülstruktur hat, nämlich XYZ. Als Bericht senden sie folgende Nachricht zu Erde:

Auf der Zwerde bedeutet das Wort ‚Wasser‘ XYZ.

Nun drehe man die Uhr um etwa 200 Jahre zurück, vor die Entdeckung der Molekularstruktur von Wasser. Auf Zwerde und Erde sitzt jeweils ein Mann namens

---

<sup>158</sup> Putnam [263], 31.

Oskar vor einem Glas klarer Flüssigkeit und hat Durst. Oskar<sub>1</sub> auf der Erde und Oskar<sub>2</sub> auf der Zwerde wissen beide das Gleiche über die Flüssigkeit, die sie ‚Wasser‘ nennen. Sie sind in diesem Moment möglicherweise im selben psychischen Zustand (im Sinne von TB<sub>1</sub>). Aber beide beziehen sich mit ‚Wasser‘ nicht auf dasselbe. Oskar<sub>1</sub> bezeichnet mit ‚Wasser‘ H<sub>2</sub>O, Oskar<sub>2</sub> hingegen XYZ.

Das heißt, daß die Extension des Ausdrucks „Wasser“ (und auch seine Bedeutung im intuitiven, vorthoretischen Sinne) keine Funktion allein des psychischen Zustandes des Sprechers ist. (Putnam [263], 34)

Putnams „Gedankenexperiment“ soll die traditionelle Bedeutungstheorie unterminieren, also zu einer gerechtfertigten Überzeugungsänderung führen. Der dazu beschriebene, als vorstellbar präsentierte Sachverhalt braucht dafür nach Putnams Ansicht offenbar nicht realisiert zu werden, um eine gerechtfertigte Überzeugungsänderung herbeizuführen.

Der als möglich behauptete Sachverhalt besteht darin, dass zwei Subjekte sich in demselben psychischen Zustand befinden, der ihrer jeweiligen Kenntnis der Bedeutung von ‚Wasser‘ entspricht. Nach der traditionellen Bedeutungstheorie müssten sie sich also beide mit ‚Wasser‘ auf dasselbe beziehen. Das ist aber absurd. Oskar<sub>1</sub> bezieht sich auf H<sub>2</sub>O, Oskar<sub>2</sub> auf XYZ. Entweder ist also die Intension kein psychischer Zustand, oder Intensionsgleichheit impliziert keine Extensionsgleichheit.

Zu den Hintergrundannahmen gehört, dass sich die Bedeutung von ‚Wasser‘ in der betrachteten 200 Jahre umspannenden Geschichte nicht verändert hat (was man schwerlich im Allgemeinen von Bedeutungen sagen will, wenn man noch keine Theorie der Bedeutung natürlicher Artbegriffe voraussetzen möchte). Diese Theorie scheint für Wasser plausibel, genau wie die implizite Annahme, dass weder Oskar<sub>1</sub>, noch Oskar<sub>2</sub> *eigentlich* ein disjunktives Prädikat verwenden, wenn sie den Ausdruck ‚Wasser‘ verwenden.

#### 4.1.5 GEHIRNE IM TANK

Auch ein anderes „Gedankenexperiment“ Putnams (aus Putnam [264]) hat es zu einiger Berühmtheit gebracht, das „Gedankenexperiment von den Gehirnen im Tank“. Inhaltlich schließt es in gewisser Weise an die soeben vorgetragenen Überlegungen an.

Wie wir gesehen haben, argumentiert Putnam gegen die Auffassung, dass (zumindest manche) mentale Repräsentationen nur aufgrund der psychischen Zustände, in denen wir uns befinden, einen bestimmten Bezug haben. Repräsentationen müssen vielmehr auf bestimmte Weise zustande gekommen sein, um einen bestimmten Bezug zu haben.

Im Fall von ‚Wasser‘ in obigem Beispiel, muss die mentale Repräsentation von Oskar<sub>1</sub> eben durch Kontakt mit H<sub>2</sub>O zustande gekommen sein. Dies war Oskar<sub>2</sub> nicht möglich, weshalb sich ‚Wasser‘ im Zwerdendeutsch nicht auf H<sub>2</sub>O bezieht.

Die These der „Neuen Theorie der Referenz“ ist, dass es Ausdrücke unserer Sprache gibt (Eigennamen, Prädikate für natürliche Arten, Demonstrativpronomen in Äußerungssituationen), deren gegenwärtige Referenz davon abhängt, in welcher kausalen Kontinuitätsbeziehung die gegenwärtige Verwendung des Ausdrucks zu einem ursprünglichen Taufakt steht.

‚Wasser‘ auf der Zwerde bedeutet XYZ, weil ein Vorfahre von Oskar<sub>2</sub> den Ausdruck ‚Wasser‘ ostensiv für die klare Flüssigkeit in den Seen, Flüssen und Meeren der Zwerde einführte, und Oskar<sub>2</sub> diesen Ausdruck in dieser Tradition verwendet. Solche Kausalketten bestehen, oder bestehen nicht, unabhängig davon, ob wir irgendeine Kenntnis davon haben. Worauf sich unserer Ausdrücke beziehen, mag uns daher in gewisser Weise a priori unzugänglich sein, aber wie wir sehen werden, können wir a priori wissen, worauf sie sich *nicht* beziehen.

Die Hypothese, dass wir (einzeln oder kollektiv) nur eine Außenwelt vorgegaukelt bekommen, während die Wirklichkeit radikal anders ist als wir glauben, ist eine in der Philosophie gängige skeptische Hypothese. Eine *skeptische Hypothese* ist eine alternative Erklärung der mir zugänglichen Daten, die lokal oder global mit meinen Überzeugungen in Widerspruch steht. Wenn ich keine Gründe finden kann, die skeptische Hypothese als alternative Erklärung auszuschließen, soll sie (global oder lokal) meine Wissensansprüche unterminieren, da es in diesem Fall so scheint, als hätte ich keine Rechtfertigung dafür, ein Überzeugungssystem (mein aktuelles) einem alternativen Überzeugungssystem (dem *mit* der skeptischen Hypothese) vorzuziehen. Globale skeptische Hypothesen sind daher Gegenstand der allgemeinen Erkenntnistheorie, während lokale skeptische Hypothesen erkenntnistheoretische Probleme in Einzelbereichen zuspitzen.

Auch Putnams Argument ist so verstanden worden, als ginge es hierbei um erkenntnistheoretische Fragen.<sup>159</sup> Putnam selbst stellt die folgenden Überlegungen aber in einem metaphysischen Zusammenhang an, im engeren Sinne im Zusammenhang mit der Frage nach der Kohärenz des metaphysischen Realismus. Wir werden Putnams Argument auch so zu analysieren versuchen.

Wie auch ‚Physikalismus‘, hat ‚metaphysischer Realismus‘ eine nicht klar festgelegte Bedeutung. Crispin Wright, dessen Interpretation von Putnam selbst als entscheidende Klärung gelobt worden ist<sup>160</sup>, charakterisiert diese Position vorsichtig so:

The view of the world which Putnam calls metaphysical realism is certainly nothing very precise. It involves thinking of the world as set over against thought in such a way that it is only by courtesy of a deeply contingent harmony, or felicity, that we succeed, if we do, in forming an overall picture of the world which, at least in its basics, is correct. This is what commits the metaphysical realist to the possibility that even an ideal theory might be false or seriously incomplete. And the same kind of thinking surfaces in the idea that the world comes prejointed, as it were, into real kinds, quite independently of any classificatory activity of ours. (Wright [356], 238)

<sup>159</sup> Vgl. Müller [214].

<sup>160</sup> Vgl. Müller [214], 74.

Damit impliziert der metaphysische Realismus beispielsweise folgende These:

- (R) Es ist möglich, dass wir mit unserer Partitionierung der Welt (z.B. der Kategorisierung in natürliche Arten) an der Wirklichkeit total vorbeigehen.

Natürlich hofft auch der metaphysische Realist, dass dem nicht so ist. Möglicherweise glaubt er auch Gründe zu haben, die es zumindest rechtfertigen, zu glauben, dass dem nicht so ist (prognostischer Erfolg seiner Theorien, etc.) Aber der metaphysische Realist kann für die Richtigkeit seiner Partitionierung der Welt keine *Gewissheit* beanspruchen, weil die Wirklichkeit nun mal unabhängig davon ist, was wir über sie denken und wie wir sie in Kategorien einteilen. Dies soll durch (R) ausgedrückt werden. Vor diesem Hintergrund entwickelt Putnam in *Vernunft, Wahrheit und Geschichte* (Putnam [264]) sein „Gedankenexperiment“ (das wir hier etwas verkürzt wiedergeben).

Man stelle sich vor, dass wir durch einen kosmischen Zufall seit Anbeginn unserer Existenz irgendwo im Universum als Gehirne in einer Nährflüssigkeit liegen. Unsere Nervenenden sind an einen Riesenrechner angeschlossen, der uns „interaktiv“ mit Input versorgt. Die Welt, so wie wir sie durch unsere „Sinne“ „wahrnehmen“ ist nur eine große Computersimulation. Da wir alle an denselben Rechner angeschlossen sind, sind wir in der Lage zu kommunizieren.

Da das folgende Argument nicht unkompliziert ist, wollen wir sofort unser Analyseinstrumentarium verwenden. (R), die von Putnam attackierte Überzeugung, impliziert für diese Szenario, dass die Gehirne im Tank folgende Hypothese nicht a priori<sup>161</sup> ausschließen können:

- (H) Ich bin ein Gehirn im Tank.

Die Targetthese scheint also in der Aussage zu bestehen, dass (H) nicht a priori ausgeschlossen werden kann. (H) scheint aber durch folgendes Argument ausgeschlossen werden zu können<sup>162</sup>:

- (1) Entweder gibt es Gehirne, oder es gibt keine Gehirne.

Fall a: Es gibt Gehirne.

- (2) Aus meinem Mund denotiert das Wort ‚Gehirn‘ die Menge aller Gehirne.  
 (3) Wenn es Gehirne gibt, dann denotiert das Wort ‚Gehirn‘ aus dem Mund eingetanker Gehirne nicht die Menge aller Gehirne.

---

<sup>161</sup> Der metaphysische Realismus ist zu der stärkeren These verpflichtet, dass es sich auch a posteriori nicht ausschließen lässt, dass unsere Kategorisierungen unzutreffend sind (dass wir niemals Gewissheit bezüglich der Adäquatheit unserer Kategorisierungen erlangen können). Wenn gezeigt werden kann, dass die skeptische Hypothese (H) schon a priori ausgeschlossen werden kann, ist natürlich diese stärkere These ebenfalls widerlegt.

<sup>162</sup> Wir folgen der Rekonstruktion von Müller [214], 181.

(4) Also denotiert das Wort ‚Gehirn‘ aus dem Mund eingetankter Gehirne nicht die Menge aller Gehirne.

(5) Also bin ich kein Gehirn im Tank.

Fall b: Es gibt keine Gehirne.

(6) Ich bin kein Gehirn im Tank.

---

(7) Also bin ich kein Gehirn im Tank.

Prämisse (1) ist eine triviale Alternation, die dem Argument in dieser Rekonstruktion vorausgestellt wird, um nicht bereits annehmen zu müssen, dass es Gehirne gibt. (2) ist eine bloße Plattitüde über unsere Sprache. Die Ausdrücke unserer Sprache lassen sich homophon in Bezug auf Referenz, Erfüllungs- und Wahrheitsbedingungen charakterisieren. (3) folgt aus dem Gedankenexperiment und Putnams Überlegungen zur Referenz von Prädikaten, die sich auf natürliche Arten beziehen (dem so genannten „Externalismus“<sup>163</sup>). Wenn die eingetankten Gehirne keinen kausalen Kontakt mit der Außenwelt haben, sondern „Gehirne“ bisher nur *simuliert* obduziert, gewogen und bestrahlt haben, kann ihr Ausdruck ‚Gehirn‘ keine Gehirne bezeichnen. (4) folgt aus (3) und der Fallunterscheidung, (5) aus (4) und (2) (und trivialen Annahmen über die Identität von Sprachen). (6) folgt trivial aus der Tatsache, dass es keine Gehirne gibt. (7) aus den beiden Konklusionen der Fallunterscheidung. Damit ist (R) falsch.

Es ist offensichtlich, dass Putnam nicht der Auffassung ist, dass man Gehirne entsprechend eintanken müsse, um festzustellen, ob sie (H) a priori ausschließen können.

Wie die Diskussion dieses Arguments gezeigt hat, ist die Bezugnahme auf den hypothetischen Fall aus verschiedenen Gründen wesentlich. Es reicht für dieses Argument, dass es nur *einen* Fall gibt, der nach der Auffassung des metaphysischen Realismus eine Instanz von (R) darstellt, aber a priori ausgeschlossen werden kann. Es ist nicht nötig, dass *jedes* skeptische Szenario a priori ausgeschlossen werden kann. (In der Tat kann die Geschichte leicht so variiert werden, dass sie nicht mehr a priori auf diese Weise ausgeschlossen werden kann, indem man annimmt, dass alle letzte Nacht eingetankt wurden.) Es geht aber (in dieser Rekonstruktion) auch nicht um eine Widerlegung des Skeptizismus, sondern um die Kohärenz des

---

<sup>163</sup> Unter „Externalismus“ bzw. „semantischem Externalismus“ versteht man die Auffassung, dass die Intension eines Ausdruck nicht dadurch bestimmt ist, was die einzelne Sprecher einer Sprachgemeinschaft über die Bedeutung des Ausdrucks meinen, sondern durch externe Faktoren festgelegt ist, wie beispielsweise die Struktur der Umwelt, in der sich die Sprachgemeinschaft befindet, bzw. durch Strukturen der Sprachgemeinschaft selbst (wenn es in der Sprachgemeinschaft beispielsweise „Experten“ und „Laien“ in Bezug auf die Referenten bestimmter Ausdrücke gibt). Während der semantische Externalismus an der (Fregeschen) These festhält, dass die Intension die Extension bestimmt, leugnet er, dass die Intension etwas ist, was „im Kopf“ der Sprecher repräsentiert sein muss. Wir werden ‚Externalismus‘ in dieser Untersuchung zur Bezeichnung dieser Auffassung verwenden.

metaphysischen Realismus. Es wird nachzuweisen versucht, dass wir uns offenbar nicht *vollständig* über die richtige Partitionierung der Wirklichkeit täuschen können. In der Gehirn-im-Tank-Hypothese sollen solche radikalen Fehlkategorisierungen vorkommen: die Gehirne im Tank bezeichnen „eigentlich“ nur verschiedene Muster elektrischer Impulse, aber sie kategorisieren sie als ‚Tiger‘, ‚Katze‘, ‚Baum‘, ‚Gehirn‘, etc.

Weiterhin ist es für dieses Argument notwendig, dass in dem speziellen hypothetischen Szenario überhaupt natürliche Arten vorkommen, dass ‚Gehirn‘ eine solche bezeichnet, etc.<sup>164</sup>

Das Argument von den Gehirnen im Tank ist aber auch deswegen interessant, weil wir hier eine Instanz eines Possibility-Refuter-Gedankenexperiments haben, das sich nichtsdestotrotz im Häggqvist-Schema wiedergeben lässt. Der metaphysische Realismus impliziert, dass eine bestimmte Möglichkeit nicht a priori ausgeschlossen werden kann, das Argument soll nachweisen, dass dies sehr wohl geht. Was gezeigt werden soll, ist, dass es sich bei der Gehirn-im-Tank-Hypothese um eine *selbstwiderlegende* Annahme handelt. Hierbei scheint, wie im theoretischen Gedankenexperiment bei Hempel, das eigentlich als möglich behauptete Szenario aufgrund der Targetthese selbst als möglich etabliert werden zu können.

#### 4.1.6 ZOMBIES

Zombies sind in den letzten Jahren zu besonderer Berühmtheit in der Philosophie des Geistes gelangt. Dabei wird unter ‚Zombie‘ auch innerhalb der Philosophie des Geistes nicht immer dasselbe verstanden. Erschwerend kommt hinzu, dass Philosophen unter ‚Zombie‘ etwas völlig anderes verstehen, als der Rest der Welt. Sucht man im *Merriam Webster Online*<sup>165</sup> nach ‚zombie‘ erhält man folgende Auskunft:

Main Entry: **zom·bie**

Variant(s): also zom·bi / ˈzäm-bE/

Function: noun

Etymology: Louisiana Creole or Haitian Creole zōbi, of Bantu origin; akin to Kimbundu nzúmbe ghost

1 usually zombi a : the supernatural power that according to voodoo belief may enter into and reanimate a dead body b : a will-less and speechless human in the West Indies capable only of automatic movement who is held to have died and been supernaturally reanimated

2 a : a person held to resemble the so-called walking dead; especially : AUTOMATON b : a person markedly strange in appearance or behavior

3 : a mixed drink made of several kinds of rum, liqueur, and fruit juice

<sup>164</sup> Vgl. Müller [214], 182.

<sup>165</sup> <http://www.m-w.com/>.

Fast nichts davon trifft auf philosophische Zombies zu. Zombies im philosophischen Sinn haben nichts mit Zauberei und Hexerei zu tun, sondern sind Wesen, die in bestimmten Hinsichten normalen Menschen gleichen, denen aber gewisse mentale Eigenschaften fehlen.

Wie diese vorsichtige Formulierung bereits nahe legt, kommen Zombies nicht nur in einer Version und nicht nur in einem Argument vor. In der allgemeinsten Variante ist ein Zombie ein Wesen, das die wesentlichen kognitiven Fähigkeiten des Menschen perfekt nachahmt und in seinem Verhalten von einem wirklichen Menschen nicht zu unterscheiden ist, obwohl es überhaupt nichts zu fühlen oder zu empfinden vermag. Dabei könnte die Forderung der kognitiven Gleichwertigkeit insbesondere dahingehend präzisiert werden, dass das Wesen sich auch in seinem verbalen Verhalten von dem eines Menschen nicht unterscheidet. Ferner könnte man spezielle „physiologische Zombies“ betrachten, die dem Menschen nicht nur im (verbalen und nonverbalen) Verhalten, sondern auch in der körperlichen Konstitution völlig gleichen.<sup>166</sup> Letztere sind in etwa die, die in der Philosophie des Geistes die wohl größte Rolle zu spielen scheinen: physikalische Zwillinge von uns, die sowohl in ihrem Verhalten, als auch in ihrem mikrophysischen Aufbau uns exakt gleichen, und die sich nur darin von uns unterscheiden, dass sie nicht über Bewusstsein verfügen, insbesondere in dem Sinne, dass sie keine phänomenalen Bewusstseinszustände (sprich „Qualia“) besitzen, dass es also nichts gibt „wie es ist, ein Zombie zu sein“.

Man kann also zunächst philosophische Zombies nach ihren äußerlichen und epistemischen Fähigkeiten einordnen und erhält dabei im Grunde drei verschiedene Zombiesorten: zunächst behaviorale Zombies, die uns zwar im Verhalten (also auch im Sprachverhalten) völlig gleichen, die aber nicht auf Kohlenstoffbasis aufgebaut sind und keine funktional/computationale Struktur besitzen, auf deren Basis man sie in wahrer Weise mit psychologischem Vokabular beschreiben könnte. Eine zweite Sorte Zombies sind dann so genannte funktionale Zombies, die zwar immer noch nicht aus Fleisch und Blut bestehen, aber eine funktional/computationale Struktur besitzen, die zumindest einen Teil unseres psychologischen Vokabulars in wahrer Weise attribuierbar macht. Dieses psychologische Vokabular könnte allerdings unvollständig auf Zombies applizierbar sein, indem es beispielsweise in der Regel ausgenommen sein wird, dass funktionale Zombies auch phänomenale Zustände – also Qualia – haben (funktionale Zombies sind also auch behaviorale Zombies, aber nicht umgekehrt). Eine dritte Form von Zombies ist dann schließlich die bereits erwähnte Klasse uns völlig gleichender physikalischer Zwillinge, die sich von uns ausschließlich in der Abwesenheit von Qualia unterscheiden, die so genannten physiologischen oder physikalischen Zombies.<sup>167</sup>

Diese Dreiteilung kann man noch einmal zusätzlich verkomplizieren, indem man die verschiedenen Weisen aufführt, in denen solche Zombies dann außerdem noch möglich sein sollen, und dementsprechend Zombies in logisch, metaphy-

---

<sup>166</sup> Vgl. Polger [259], Güzeldere [133].

<sup>167</sup> Güzeldere [133].

sisch und natürlich mögliche unterteilt.<sup>168</sup> Logisch möglich sei ein Zombie-Szenario, wenn es widerspruchsfrei beschrieben werden kann. Metaphysisch möglich sei ein Zombie-Szenario, wenn es darüber hinaus keinen etwaigen metaphysischen Notwendigkeiten widerspricht (wobei es vielleicht metaphysische Notwendigkeiten geben mag, die nicht bereits in der Menge der logischen Notwendigkeiten enthalten sind). Natürlich möglich sei ein Zombie-Szenario, wenn es mit den Naturgesetzen kompatibel ist, wobei hierbei auch eventuelle psychophysikalische Naturgesetze eingeschlossen sind. Dieser Varianz an Zombies entspricht auch eine Varianz an Target-Thesen, gegen die die Zombie-Argumente gerichtet sind. In der Regel handelt es sich dabei um Spielarten des Physikalismus. Wir werden auf David Chalmers' Zombie-Argumente, die in der gegenwärtigen Philosophie des Geistes große Beachtung erfahren haben, im Detail in Kapitel 7.2 eingehen, da ihre Darstellung Einiges an Erläuterungen bedarf, die wir hier noch nicht leisten können. Aus diesem Grund sei an dieser Stelle als Beispiel nur ein Zombie-Argument gegen den *semantischen* Physikalismus erwähnt:

Der *semantische Physikalismus* (auch „logischer Behaviorismus“) zeichnet sich durch folgende Thesen aus<sup>169</sup>:

- (B<sub>1</sub>) Mentale Ereignisse sind physische Ereignisse
- (B<sub>2</sub>) Mentalistische Aussagen lassen sich in bedeutungsgleich nichtmentalistische Aussagen über Verhaltensdispositionen übersetzen.
- (B<sub>3</sub>) Alle physischen Ereignisse sind ausschließlich von physischen Ereignissen verursacht.

Durch den semantischen Physikalismus wären alle Zombiesorten ausgeschlossen. Insbesondere B<sub>2</sub> behauptet ja, dass behaviorale Zombies logisch unmöglich sind. Da damit die schwächste Zombievariante ausgeschlossen ist, sind alle ausgeschlossen. Gegen diese Version des Physikalismus ließe sich mit einem Zombie-Argument folgendermaßen argumentieren:

Es ist vorstellbar, dass es Wesen gibt, die dieselben Verhaltensdispositionen wie wir zeigen (beispielsweise ein Verhalten, das wir an den Tag legen, wenn wir Chinesisch verstehen), die im Innern aber durch einen Computer gesteuert werden, der nur syntaktisch Symbole verarbeitet (wie Searle im Chinesischen Zimmer). Der semantische Physikalismus impliziert für diese Situation, dass es logisch unmöglich ist, dass dieses Wesen Chinesisch *nicht* versteht. Das ist aber absurd (hier könnte man ähnliche Überlegungen wie in Searles Argument anführen). Folglich ist der semantische Physikalismus falsch.

In diesem Argument (es sei einmal dahingestellt, ob es jemanden überzeugen kann<sup>170</sup>) wird gegen den semantischen Physikalismus mit der logischen Möglich-

<sup>168</sup> Vgl. Polger [259].

<sup>169</sup> Pauen [251], 81; vgl. auch Beckermann [18], 65.

<sup>170</sup> Es gibt wesentlich stärkere Argumente gegen den semantischen Physikalismus. Hierzu gehört, dass mentale Prädikate in der Regel Cluster-Begriffe sind, die sich nicht ohne Weiteres

keit eines behavioristischen Zombies argumentiert. Das Argument haben wir schon in Häggqvists Schema formuliert. Dieses Argument erbt Hintergrundannahmen aus Searles Argument. Eine interessantere Variante des Zombie-Arguments werden wir – wie gesagt – in Kapitel 5 kennen lernen.

#### 4.1.7 DAS REDUPLIKATIONSARGUMENT

Als Abschluss der Sammlung sei noch einmal ein Gedankenexperiment aus der Metaphysik erwähnt. In diesem Fall ein Gedankenexperiment aus der Debatte um die Analyse der Bedingungen diachroner personaler Identität. Bei dieser Debatte geht es um die Frage, welche notwendigen und hinreichenden Bedingungen dafür bestehen, dass ein Personenstadium  $P_1$  zu einem Zeitpunkt  $t_1$  zur selben Person gehört wie ein Personenstadium  $P_2$  zu einem späteren Zeitpunkt  $t_2$ .<sup>171</sup>

So genannte „mentalistiche“ Theorien versuchen (in der Tradition von John Locke) die diachrone Identität von Personen mittels der Erinnerung bzw. der Kontinuität des Bewusstseins, oder der Erinnerung und der potenziellen Erinnerung zu analysieren. Eine solche Analyse könnte etwa folgendermaßen<sup>172</sup> lauten:

- (M) Ein Personenstadium  $P_2$  zu  $t_2$  gehört zur selben Person wie ein früheres Personenstadium  $P_1$  zu  $t_1$ , gdw.  $P_2$  mit  $P_1$  psychisch kontinuierlich ist.

Gegen solche Analysen ist das so genannte „Reduplikationsargument“ vorgebracht worden, dass wir uns kurz näher ansehen wollen.<sup>173</sup> Hier ist John Perrys Darstellung des zugrunde liegenden Gedankenexperiments:

Brown, Jones und Smith gehen in die Klinik, um sich einer Gehirnverjüngungskur zu unterziehen. (Bei einer Gehirnverjüngungskur wird das Gehirn entnommen, seine Verschaltungen werden von einer sagenhaften Maschine analysiert, und ein neues Gehirn, das in allen relevanten Hinsichten genauso beschaffen ist wie das alte, nur dass es aus einer frischen Hirnmasse besteht, wird in den Schädel zurückverpflanzt. Nach einer Gehirnverjüngungskur fühlt man sich besser, man kann klarer

---

durch die Angabe notwendiger und hinreichender Bedingungen definieren lassen; dass es außerdem schwierig sein wird, solche Definitionen so vollständig zu definieren, dass sie nicht mit Gegenbeispielen konfrontiert sind, und dass sich mentale Ausdrücke vermutlich nicht zirkelfrei in physikalischer Sprache definieren lassen. Vgl. Beckermann [18], 90.

<sup>171</sup> Aus technischen Gründen ist das Problem in „vierdimensionalistischer“ Redeweise über zeitliche Personenstadien formuliert. Personenstadien sind zeitliche Teile von Personen. Ob es solche gibt, ist kontrovers. (Vgl. etwa Sider [298] für einen Überblick.) Manche Philosophen finden es z.B. kontraintuitiv, dass jetzt nur ein zeitlicher Teil von mir vor dem Computer sitzt, und nicht ich als Ganzes. Für die oben skizzierte Problemstellung hat das allerdings keine Konsequenzen. Das Problem kann genauso gut „dreidimensionalistisch“ über „mögliche Geschichten einer Person“ formuliert werden. Vgl. Noonan [237], 122-127.

<sup>172</sup> Vgl. etwa Noonan [237], 13. Zu alternativen Varianten des Mentalismus, vgl. Perry [253], Perry [254], Perry [255], Perry [256].

<sup>173</sup> Vgl. Shoemaker [296]; Wiggins [341], Wiggins [342], Wiggins [343], Wiggins [344].

denken und sich deutlicher erinnern, wobei aber die Inhalte der Erinnerungen und Überzeugungen nicht verändert werden.) Die Gehirne der drei werden entnommen und auf das Gehirn-Wägelchen gelegt. Durch einen dummen Zufall schmeißt der Krankenpfleger das Wägelchen um; die Gehirne von Smith und Brown sind hinüber. Damit sein tragischer Missgriff nicht entdeckt wird, lässt der Krankenpfleger Jones' Gehirn dreimal durch die sagenhafte Maschine laufen und bringt dann die Duplikate zurück in den Operationssaal. Zwei dieser Duplikate werden in die ehemals Brown und Smith gehörenden Schädel verpflanzt. Jones' altes Herz hat versagt, und nach einer Weile wird er für tot erklärt.

Nach ein paar Stunden erwachen jedoch zwei Individuen, die beide behaupten, Jones zu sein, die beide froh sind, endlich ihre Kopfschmerzen los zu sein, die aber etwas bestürzt sind über die drastischen Veränderungen, die sich offenbar an ihren Körpern vollzogen haben. Wir werden diese Personen „Smith-Jones“ und „Brown-Jones“ nennen. Die Frage lautet: Wer sind sie? (Perry [254], 121)

Appliziert man (M) auf diesen als „möglich“ behaupteten Fall, sind Smith-Jones und Brown-Jones beide Jones, da sie beide mit Jones psychisch kontinuierlich sind. Das ist aber absurd, da Identität eine transitive Relation ist, und demnach, wenn Brown-Jones mit Jones, und Smith-Jones mit Jones identisch ist, auch Brown-Jones mit Smith-Jones identisch sein müsste. Brown-Jones und Smith-Jones sind aber offenbar nicht miteinander identisch:

Smith-Jones hat sich hingelegt, Brown-Jones hat sich aufgesetzt; Brown-Jones denkt gerade an seine Krankenschwester, Smith-Jones denkt gerade an Jones' Frau (Sie denken beide an die Krankenschwester, wenn sie im Zimmer ist, sonst an die Ehefrau; in diesem Augenblick ist aber die Krankenschwester gerade in Brown-Jones' Zimmer). Auf den einen trifft also alles Mögliche zu, was auf den anderen nicht zutrifft. (Perry [254], 122)

(M) ist also falsch.

Die Hintergrundannahmen, die hier eingehen und zunächst unproblematisch erscheinen, sind etwa, dass ‚synchrone Identität‘ und ‚diachrone Identität‘ denselben Sinn von ‚Identität‘ meinen; dass Jones zuvor nur eine Person war, etc.<sup>174</sup>

---

<sup>174</sup> Vgl. hierzu Perry [256], Noonan [237].

## 4.2 GEDANKENEXPERIMENTE IN DER PRAKTISCHEN PHILOSOPHIE

Auch in der Praktischen Philosophie gibt es eine ganze Reihe prominenter Gedankenexperimente. Zu den Bekanntesten gehören sicherlich die Gedankenexperimente von Judith Jarvis Thomson zu ethischen Fragen der Abtreibung aus ihrem einflussreichen Artikel ‚A Defense of Abortion‘ (Thomson [324]). Andere bekannte Gedankenexperimente sind der bereits erwähnte Ring des Gyges, Nozicks Erfahrungsmaschine, Moores zwei Welten, Foots Gasbeispiel, etc.<sup>175</sup> In diesem Kapitel wollen wir uns auf das bekannteste Gedankenexperiment Thomsons beschränken, den „weltberühmten Geiger“.

### 4.2.1 THOMSONS GEIGER

Thomson stellt zunächst fest, dass die Abtreibungsdebatte in der Regel um die Frage kreist, ob der Fötus bzw. Embryo, dessen Abtreibung zur Debatte steht, als *Person* gelten soll, oder nicht. Abtreibungsgegner versuchen nachzuweisen, dass er als Person zu gelten hat, Abtreibungsbefürworter versuchen nachzuweisen, dass es sich noch nicht um eine Person handeln kann. Dabei scheint von beiden Parteien davon ausgegangen zu werden, dass, wenn es sich bei einem Fötus bzw. Embryo bereits um eine Person handelt, diese Tatsache die Frage der moralischen Zulässigkeit einer Abtreibung entscheidet. Für diesen Zusammenhang wird aber nur selten argumentiert. Thomson vermutet, dass in etwa das Folgende vorausgesetzt wird:

- (A) Das Lebensrecht einer Person (qua Person) wiegt in jedem Fall moralisch schwerer als das Selbstbestimmungsrecht anderer Personen über ihren Körper.

D.h., dass Frauen zwar ein Selbstbestimmungsrecht darüber zugestanden wird, was mit ihrem Körper passiert, dieses Recht allerdings außer Kraft gesetzt werden kann von dem schwerwiegenderen Recht einer anderen Person auf Leben. Alle Personen haben ein solches Recht auf Leben, folglich auch der Embryo, so er denn eine Person ist.

But now let me ask you to imagine this. You wake up in the morning and find yourself back to back in bed with an unconscious violinist. A famous unconscious violinist. He has been found to have a fatal kidney ailment, and the Society of Music Lovers has canvassed all the available medical records and found that you alone have the right blood type to help. They have therefore kidnapped you, and last

---

<sup>175</sup> Für einen repräsentativen Überblick vgl. Tittle [329].

night the violinist's circulatory system was plugged into yours, so that your kidneys can be used to extract poisons from his blood as well as your own. The director of the hospital now tells you, "Look, we're sorry the Society of Music Lovers did this to you – we would never have permitted it if we had known. But still, they did it, and the violinist is now plugged into you. To unplug you would be to kill him. But never mind, it's only for nine months. By then he will have recovered from his ailment, and can safely be unplugged from you." (Thomson [324], 48-49)

Würde (A) auf einen solchen Fall appliziert, sollte man moralisch verpflichtet sein, die nächsten neun Monate an den Geiger angeschlossen zu bleiben. Er ist eine Person und hat ein Recht auf Leben, was, falls (A) uneingeschränkte Geltung hat, das Selbstbestimmungsrecht auch in diesem Fall übertrumpfen sollte. Diese Konsequenz ist aber *prima facie* absurd. Es wäre zwar unheimlich nett, bliebe man die nächsten neun Monate im Bett liegen und liehe man dem Geiger seine Nieren, moralisch verpflichtet dazu scheint man allerdings nicht zu sein.

Dass (A) in diesem Fall zu absurden Konsequenzen führt, ist aber nur dann für die Abtreibungsthematik von Relevanz, falls die Hintergrundannahme zutrifft, dass das Verhältnis zwischen dem Geiger und dem unfreiwilligen Nierenleiher in allen ethisch relevanten Hinsichten analog zu *zumindest manchen* Mutter/Embryo-Verhältnissen ist. Es ist häufig gegen dieses Gedankenexperiment eingewendet worden, dass der Fall nicht in allen ethisch relevanten Hinsichten analog zu *allen* Mutter/Embryo-Verhältnissen ist<sup>176</sup>, was für Thomsons Argument aber zunächst völlig unerheblich ist.<sup>177</sup> Eine entscheidende Disanalogie wäre vielmehr dann gefunden, wenn die Mutterschaft *an sich* moralische Verpflichtungen mit sich brächte, die zur Beurteilung des vorliegenden Falls relevant wären.

Eine andere Hintergrundannahme wäre, dass die moralische Intuition, dass es kein Unrecht darstellt, verweigerte man dem Violinisten die Hilfe, invariant ist bezüglich der metaethischen Position, die man vertritt. Ob diese Hintergrundannahme zutrifft, ist beispielsweise von Peter Singer bezweifelt worden, der der Meinung ist, dass Utilitaristen eine solche Entscheidung als Unrecht beurteilen würden (Singer [299], 191-194). Nach utilitaristischer Auffassung sollten wir unser Handeln danach richten, was die besten Konsequenzen hat, wobei in diesem Fall nach Abwägen aller Umstände die Konsequenzen für alle anderen (insbesondere den Geiger) schlimmer zu sein scheinen, als die Konsequenzen für den unfreiwilligen Nierenleiher, was bedeutet, dass er angekoppelt bleiben sollte:<sup>178</sup>

<sup>176</sup> Da (beispielsweise) in dem beschriebenen Szenario die betreffende Person entführt wurde, Schwangerschaften aber nicht immer durch Vergewaltigung entstehen, etc.

<sup>177</sup> Thomson will es durchaus zulassen, dass das Recht auf Leben des ungeborenen Kindes die Selbstbestimmungsrechte der Mutter unter bestimmten Umständen übertrumpfen. Es geht nur darum, dass sie dies nicht immer automatisch tun.

<sup>178</sup> Wir werden im nächsten Kapitel genauer auf Zirkularitätsvorwürfe gegen solche Gedankenexperimente eingehen. Es sollte aber hier schon festgehalten werden, dass die Debatte zwischen Abtreibungsbefürwortern und Abtreibungsgegnern keine Debatte zwischen Nicht-Utilitaristen und Utilitaristen ist. Im Gegenteil stehen viele Utilitaristen (u.A. Singer selbst) auf der Seite der Abtreibungsbefürworter. Dass Thomson für ihr Argument also eine nicht-

Indem sie Thomsons Rechtstheorie und damit ihr Urteil im Fall des Geigers zurückweisen, würden die Utilitaristen auch ihr Argument für die Abtreibung ablehnen. Thomson hat behauptet, daß ihr Argument die Abtreibung selbst dann legitimieren würde, wenn wir es zuließen, daß das Leben eines Fötus ebensoviel zählt, wie das Leben einer normalen Person. Utilitaristen würden hingegen behaupten, es sei falsch, sich zu weigern, das Leben einer Person neun Monate lang zu erhalten, wenn dies die einzige Überlebenschance der Person wäre. (Singer [299], 194)

Nach unserer Rekonstruktion soll das Gedankenexperiment von Thomson ein stillschweigend angenommenes moralisches Prinzip kritisieren. Das fragliche Prinzip besagt hierbei, dass Rechte einer bestimmten Art prinzipiell mehr wiegen als Rechte einer anderen Art. Das Gegenbeispiel von Thomson weist dann – so es denn überzeugen kann – nach, dass dieses Prinzip nicht immer gilt und dass es darüber hinaus in Fällen nicht gilt, die in allen ethisch relevanten Hinsichten analog zu bestimmten strittigen Fällen sind (unfreiwillige Schwangerschaft aufgrund einer Vergewaltigung).

Ausgeweitet auf die Verwendung von Gedankenexperimenten in der Ethik im Allgemeinen ist es nahe liegend anzunehmen, dass sich die ethische „Epistemologie“ etwa folgendermaßen rekonstruieren lässt: zur Bewertung problematischer Fälle werden stillschweigend oder explizit moralische Prinzipien (wie z.B. (A)) angenommen. Durch Gedankenexperimente können solche Prinzipien bereichsbeschränkt oder im Allgemeinen widerlegt werden. Es müssen dann neue Prinzipien gefunden werden, die mit der intuitiven Bewertung des imaginären Falles kompatibel sind, um den strittigen Fall zu bewerten. Gedankenexperimente in der Ethik hätten dann die Häggqvist-Standardstruktur und würden eine ähnliche epistemologische Rolle spielen wie die Gedankenexperimente in anderen Bereichen der Philosophie.

Wenn dem so wäre, könnten wir Gedankenexperimente in der Ethik im Allgemeinen analog zu allen anderen Gedankenexperimenten behandeln (u.U. mit Sonderproblemen, die gegenstandsspezifisch sind), zumindest ihrer Funktion nach wären sie analog und müssten dann auch analoge Gütekriterien erfüllen, etc. Da dies für unsere Untersuchung natürlich wünschenswert wäre, müssen wir an dieser Stelle bereits unsere Rekonstruktion gegen eine Alternative verteidigen, die ebenfalls von sich behauptet, eine rationale Rekonstruktion der Verwendung imaginärer Fälle in der Praktischen Philosophie darzustellen, diesen aber eine gänzlich andere Funktion in der ethischen Epistemologie zuweist.

---

utilitaristische Rechtstheorie voraussetzt macht ihr Argument problematisch, aber nicht *zirkulär*.

#### 4.2.2 FUNKTIONIEREN GEDANKENEXPERIMENTE IN DER PRAKTISCHEN PHILOSOPHIE ANDERS?

In ‚The Role of Imaginary Cases in Ethics‘ (Dancy [82]) und in *Moral Reasons* (Dancy [83]) hat Jonathan Dancy dafür argumentiert, dass eine an moralischen Prinzipien orientierte (oder auf moralische Prinzipien abzielende) Ethik imaginären Fällen keinen sinnvollen Platz in ihrer Methodologie zuweisen kann. Dies steht offenbar in Widerspruch mit unserer Diagnose. Wenn wir das Häggqvist-Schema auch auf Gedankenexperimente in der Ethik anwenden wollen, müssen wir auch dort „Target-Thesen“ identifizieren. *Prima facie*-Kandidaten für solche Target-Thesen sind aber gerade moralische Prinzipien.

The sort of appeal to imaginary cases I have primarily in mind has instances varying from some of the parables of the New Testament and Aesop’s fables to more direct attempts to determine our attitudes in the following sort of way. There is at present a controversy about whether Britain ought to return the Elgin Marbles to Greece.<sup>179</sup> An argument might run as follows. Suppose that an orphan’s estate is under the control of a not particularly conscientious trustee, who for reasons of his own allows a collector to remove part of that estate in exchange for a (not really sufficient) payment which the trustee then absorbs for his own purposes. The orphan eventually comes into his own estate and attempts to recover what was his property. Whatever be the legal situation, must we not admit that the collector has some moral obligation to return the property? And surely that obligation is not reduced or in any way counterbalanced by the fact that the property is now on view in a collection of objects many of which were acquired in similar dubious circumstances. So the collector ought to return the property. And similarly Britain ought to return the Elgin Marbles.

---

<sup>179</sup> Diese Diskussion ist immer noch aktuell und hat sogar im Zusammenhang mit der Austragung der Olympischen Spiele in Athen im Jahr 2004 wieder an Brisanz gewonnen. Es geht darum, dass Thomas Bruce (Lord Elgin) Anfang des 19. Jahrhunderts Teile des Frieses des Parthenons aus Griechenland (möglicherweise illegal) ausgeführt hat. Griechenland befand sich damals unter türkischer Herrschaft. Lord Elgin, der als britischer Botschafter in Konstantinopel eingesetzt war, hatte 1801 eine Abmachung mit dem Sultan treffen können, die es ihm gestattete, herabgestürzte Teile des Parthenons nach England zu schaffen. 1816 verkaufte er die Stücke an die britische Regierung; seitdem gehören sie dem British Museum. Die Stücke, die Lord Elgin wohl zum Teil schon bei der „Bergung“ beschädigte, mussten auch durch Transport und „Restaurierung“ weiter Schaden nehmen. Die griechische Regierung fordert die Rückgabe dieser Kulturgüter, stellt mittlerweile aber keine Besitzansprüche mehr. Zuletzt wurde von Griechenland vorgeschlagen, dass die Stücke im Besitz des British Museums bleiben, allerdings permanent in Athen ausgestellt werden (wobei eventuelle Eintrittspreise vom British Museum erhoben werden dürften). Im Gegenzug bietet die griechische Regierung zusätzlich an, unentgeltlich Prachtstücke griechischer antiker Kunst in einem Rotationsverfahren nach England zu entleihen. Die gegenwärtige britische Regierung lehnt eine Herausgabe der Elgin Marbles allerdings ab. Vgl. auch Hitchens [150].

This seems to be an instance of a recognisable form or moral argument. [...] [It] attempts to establish what we actually ought to do [...]. We suppose that our duty is not clear, but that it can be made clear by the consideration of a relevantly similar but imaginary case. (Dancy [82], 143)

Gegen Argumente dieser Art, bei denen ein imaginärer Fall herangezogen wird, um in einem tatsächlichen Fall zu einer Entscheidung zu kommen, bringt Dancy vier unabhängige Argumente vor. Die *kritisierte* methodologische Rolle imaginärer Fälle ist demnach, ein moralisches Urteil herbeizuführen, das dann auf einen analogen aktuellen Fall übertragen werden kann, gegeben, dass der imaginäre Fall Eigenschaften hat, die eine solche Übertragung gestatten. Was findet Dancy an dieser Verwendungsweise problematisch?

Dancys erster Einwand besteht darin, dass imaginäre Fälle moralisch unbestimmt sind. Dancys Idee scheint ungefähr folgende zu sein: in imaginären Fällen spezifizieren wir immer nur eine bestimmte Menge von Eigenschaften, die ein Fall hat. Die moralischen Eigenschaften einer Situation „resultieren“ aus den nicht-moralischen Eigenschaften, die ein Fall hat. Wenn nicht alle nicht-moralischen Eigenschaften gegeben sind, ist der Fall moralisch unbestimmt. Wenn wir in solchen Situationen dennoch behaupten, wir könnten ein moralisches Urteil fällen, täuschen wir uns:

The problem lies in the idea that the moral relevance of the properties so far described is not affected by any further properties; that there is a way of filling out the description which leaves untouched that moral relevance as it presently stands. But in the absence of further description there is no such thing as the moral relevance of properties so far given. [...] And if this is so, in responding to the case as so far described we are pretending (claiming) to make a decision about a matter which is indeterminate. But such a pretence does not make sense. (Dancy [82], 145)

Im Hintergrund dieses Arguments scheint eine ziemlich zweifelhafte Metaphysik zu stehen, nach der moralische Eigenschaften von Situationen von *absolut allen* nicht-moralischen Eigenschaften einer Situation ontologisch abhängen. Wenn dem so wäre, ließe sich Dancys Problem mit der Unbestimmtheit durch eine mikrophysikalische Beschreibung und eine Liste aller relevanten Naturgesetze beseitigen.<sup>180</sup> Wäre dadurch viel gewonnen? Man könnte zwar dafür argumentieren, dass es in einer so beschriebenen möglichen Welt nicht mehr unbestimmt sein kann, welche moralischen Eigenschaften in ihr auftreten, *ablesen* könnten wir diese von einer solchen Beschreibung allerdings viel weniger als von einer vergleichsweise oberflächlichen Beschreibung. Warum ist das so?

Wenn wir im Alltag moralische Urteile über aktuelle Fälle treffen, tun wir dies ebenfalls in Unkenntnis unzählbar vieler nicht-moralischer Eigenschaften. Im Alltag ist jeder Fall (natürlich) vollständig metaphysisch determiniert, und – zumindest prinzipiell – könnten wir weitere Eigenschaften des Falles aufdecken, die uns

---

<sup>180</sup> Wobei wir kurz eventuelle anti-physikalistische Probleme mit dieser Aussage ignorieren wollen.

zu einer Revision unseres moralischen Urteils führen könnten. *De facto* behandeln wir dabei einige nicht-moralische Eigenschaften als *moralisch relevante* Eigenschaften, andere nicht. Natürlich können wir uns dabei irren, so dass auf den ersten Blick moralisch irrelevante Eigenschaften sich nach genauerer Analyse als moralisch relevante Eigenschaften erweisen. Wenn es keine Trennung zwischen moralisch relevanten und moralisch irrelevanten Eigenschaften gäbe, müssten wir uns immer eines moralischen Urteils enthalten (da praktisch jedes moralische Urteil eine Anmaßung darstellen würde).

Dancy könnte nun einwenden, dass die erfolgreiche Verwendung imaginärer Fälle in der Ethik davon abhängt, dass tatsächlich alle moralisch relevanten Eigenschaften in einem beschriebenen Fall festgelegt worden sind. Andernfalls ist unser moralisches Urteil ohne Gegenstand und damit *simlos*. In Bezug auf obigen Fall zu antworten, dass es richtig wäre, dem Waisenkind seine Besitztümer zurückzugeben, ist eine (möglicherweise) sinnlose Aussage, da der beschriebene Fall (eventuell) noch gar keine moralische Eigenschaft besitzt, die von uns aufgedeckt werden könnte. Im aktuellen Fall gibt es aber eine solche moralische Tatsache, folglich ist eine moralische Beurteilung des Verhaltens des British Museums eine sinnvolle, wenn auch fehlbare Aussage.

Metaphysisch mag man dazu stehen wie man will, erkenntnistheoretisch scheint hier kein wirkliches Problem zu bestehen. Ein Irrtum in Bezug auf den imaginären Fall scheint genau in demselben Sinne fehlbar zu sein wie ein Urteil in Bezug auf einen tatsächlichen Fall. Es scheint auch in genau derselben Weise korrigierbar zu sein.

Angenommen, wir irren uns bei der Bewertung eines aktuellen Falles in Bezug auf die Menge der moralisch relevanten Eigenschaften. Der Fall hat noch eine weitere moralisch-relevante Eigenschaft R, die wir bisher nicht bedacht haben. Wenn wir sie in unsere Beurteilung mit einbeziehen, sehen wir uns allerdings gezwungen, unser ursprüngliches Urteil zu revidieren. In Bezug auf die Beurteilung des aktuellen Falls war unser Urteil einfach falsch.

Angenommen, wir irren uns bei der Bewertung eines imaginären Falles in Bezug auf die Menge der moralisch relevanten Eigenschaften. Wir haben nicht stipuliert, wie der Fall hinsichtlich R aussieht. Wenn wir den Fall hinsichtlich R aber spezifizieren, sehen wir uns gezwungen, unser ursprüngliches Urteil zu revidieren. In Bezug auf die Beurteilung des aktuellen Falls war unser Urteil laut Dancy „sinnlos“.

In der Praxis unserer Beurteilung aktueller bzw. imaginärer Fälle macht dies allerdings keinen Unterschied. Es mag sowohl so sein, dass wir durch weiteres Nachdenken über die beiden Fälle darauf gekommen sind, dass es noch eine weitere moralisch relevante Eigenschaft gibt, die wir zur Bewertung heranziehen müssen, wie auch, dass uns unsere Bekanntschaft mit anderen (aktuellen oder imaginären) Fällen dazu gebracht hat, auch Eigenschaften zur Bewertung heranzuziehen, die wir bislang für moralisch irrelevant gehalten haben.

Wir könnten genauso gut in beiden Fällen sagen, dass unser moralisches Urteil falsch war, und zwar – in beiden Fällen – weil wir uns über die Menge der mora-

lisch relevanten Eigenschaften getäuscht haben. Bei Dancys erstem Argument handelt es sich also um ein Scheinproblem.<sup>181</sup>

Dancys zweites Argument bezweifelt die Übertragbarkeit des Ergebnisses vom imaginären auf den aktualen Fall. Der imaginäre Fall ist unvollständig beschrieben, und wir bekommen suggeriert, dass es eine weitere Beschreibung der Details des Falles gibt, die die moralischen Eigenschaften des Falles nicht ändern würde. Für die Übertragbarkeit des Urteils müssen wir aber außerdem annehmen, dass diese weitere Beschreibung des imaginären Falles die Analogie zum aktualen Fall nicht nachträglich zerstört. Während wir für den imaginären Fall wissen können, wie er weiter auszubuchstabieren ist, ohne dass seine moralischen Eigenschaften verletzt werden (da wir ihn ja stipulieren), können wir dies nicht für den aktualen Fall wissen, ja nicht einmal gerechtfertigter Weise annehmen.

Zunächst ist es verwunderlich, dass wir plötzlich „Wissen“ von etwas haben können, was zwei Absätze zuvor noch unbestimmt war. Wie dem auch sei, auch hier scheint kein wirkliches Problem zu bestehen. Natürlich kann sich unsere *Vermutung*, dass der imaginäre Fall sich im Detail weiter beschreiben ließe, ohne dass die Analogie zum aktualen Fall verletzt würde, als falsch erweisen. Aber ein solcher Irrtum könnte natürlich korrigiert werden. Angenommen, die Amerikaner hätten während des zweiten Weltkrieges auch die Britischen Kronjuwelen zur Sicherung vor den Deutschen in Verwahrung genommen und nach Washington geschafft. Dem griechischen Botschafter gelingt es, für einige Flaschen Ouzo die britischen Kronjuwelen vom amerikanischen Präsidenten zu erwerben, und während einer großen Gala in Athen werden unter den Augen der schockierten Öffentlichkeit die britischen Herrschaftsinsignien vom griechischen Parlament als Rache für den Raub von Lord Elgin zertrampelt und eingeschmolzen. Wäre diese Tatsache Bestandteil der Geschichte des aktualen Falls, wäre der imaginäre Fall nicht mehr direkt analog. Auch die weitere Beschreibung des imaginären Falls müsste eine solche Weiterführung beinhalten, falls man *vermuten* würde, dass dieser Bestandteil der tatsächlichen Geschichte von moralischer Relevanz ist. *Ob* er es ist, könnte man dann sehr wohl an unserer Reaktion gegenüber dem so erweiterten imaginären Fall ablesen.

Dancys drittes Argument scheint darauf hinaus zu laufen, dass sich die Rechtfertigung für die Verwendung von imaginären Fällen in der Ethik selbst unterminiert. Folgt man Dancys Rekonstruktion der Rolle, die imaginäre Fälle in der Ethik spielen, geht es dabei – wie wir oben schon sagten – um die Übertragung eines Urteils, das am imaginären Fall gewonnen wurde, auf einen aktualen Fall. Laut Dancy liegt der Sinn dieser Übung darin, dass der aktuelle Fall eine unmittelbare Beurteilung nicht zulässt, oder nicht sicher zulässt, weshalb man überhaupt den imaginären Fall zu Rate zieht. Dieser imaginäre Fall soll uns ein sichereres oder einfacheres moralisches Urteil gestatten als der aktuelle, und dieses Urteil soll dann übertragbar sein, weil der imaginäre Fall nichtsdestotrotz in allen moralisch relevanten Hinsichten dem aktualen Fall hinreichend ähnlich ist, um so eine Übertragung zuzulassen.

---

<sup>181</sup> Eine ähnliche Argumentation findet sich auch bei Häggqvist [137], 41-43.

Nun ist der imaginäre Fall vom aktuellen Fall in einer Hinsicht verschieden: er ist einfacher zu beurteilen als der aktuelle Fall.

But this could only be so if there were other relevant differences between the two cases, in which the one is hard and the other easy. The one that is easy can be so because it does not contain factors which complicate the issue in the hard case. And, this being so, it is odd to suppose that our decision in the easy case should help when we come to the actual, hard case. (Dancy [82], 146)

Es könnte sehr wohl der Fall sein, dass diejenigen Unterschiede, die den aktuellen Fall komplizieren, selbst moralisch relevant sind, das Urteil bezüglich des imaginären Falles entsprechend auch nicht übertragbar. Damit ergibt sich – nach Dancy – folgendes Dilemma: entweder die Fälle sind sich sehr ähnlich, dann ist der imaginäre Fall genauso schwer zu beurteilen wie der aktuelle Fall, und es ist nur ablenkend, sich mit dem imaginären Fall zu beschäftigen, oder die Fälle sind sich nicht sehr ähnlich, dann ist der imaginäre Fall zwar leichter zu beurteilen, aber dieses Urteil ist nicht relevant für den anders gelagerten aktuellen Fall.<sup>182</sup>

Natürlich ist dies kein echtes Dilemma, wenn wir in der Lage sein sollten, zwischen ethisch relevanten Faktoren einerseits und moralisch irrelevanten, bloß komplizierenden Faktoren andererseits zu unterscheiden. Hieran schließt Dancy sein viertes Argument:

The suggestion here is that we are able to determine the similarity of A and B in respects relevant to F-ness before we have determined their similarity (or otherwise) in respect to F-ness; and that we should work from the first decision to the second. Does this make good sense? The first point to make is that if we are to work in this way we need to determine, not just that A and B are similar in respects  $R^1$ - $R^n$ , but that  $R^1$ - $R^n$  are all the properties relevant to F-ness (or relevant to the F-ness of these two cases). Can we do this in advance of determining whether B is F or the degree to which it is F? (Dancy [82], 147)

Nach Dancy müssten wir eigentlich zuerst feststellen, dass A und B unter dieselbe Beurteilung fallen, um von da aus Grund zu der Annahme zu haben, dass sich A und B in allen moralisch relevanten Eigenschaften hinreichend ähnlich sind. Dazu müssten wir wissen, wie A und B zu beurteilen sind; wenn A der aktuelle Fall ist, wissen wir dies aber von A auf jeden Fall nicht. Dancys erster Vorwurf besteht also darin, dass das Verfahren, aktuelle Fälle an in allen moralisch relevanten Hinsich-

---

<sup>182</sup> Dieses Argument erinnert an ein Argument aus der Tierethik, bei dem die Rechtfertigbarkeit von Tierversuchen in Zweifel gezogen wird. Entweder seien die Tiere, an denen Versuche ausgeführt werden, dem Menschen sehr ähnlich, dann ist nicht klar, warum ihr Leiden weniger schlimm sein sollte (und warum sie dann an Stelle von Menschen mit den Versuchen gepeinigt werden), oder sie sind vom Menschen verschieden, dann ist aber die Übertragbarkeit der Resultate auf den Menschen in Zweifel zu ziehen, weshalb der Nutzen der Versuche in keinem Verhältnis zu den verursachten Qualen mehr steht. Auch dieses Argument ist so zunächst unzureichend und kein echtes Dilemma, wenn die medizinisch relevante Ähnlichkeit von der moralisch relevanten Ähnlichkeit unabhängig ist. (In Dancys Fall geht es um relevante epistemische und moralische Ähnlichkeit.)

ten ähnlichen imaginären Fällen zu entscheiden, bereits eine Entscheidung des aktuellen Falls voraussetzt, um einen imaginären Fall überhaupt gerechtfertigt als in allen moralisch relevanten Eigenschaften hinreichend ähnlich behandeln zu dürfen.

Ein weiterer Einwand Dancys gegen diese Vorgehensweise speist sich außerdem aus der von Wittgenstein stammenden Vermutung, dass die moralischen Eigenschaften eines Falles zwar von den nicht-moralischen Eigenschaften dieses Falles bestimmt sind, dies aber nicht in gesetzesartiger Weise sind. Es gibt also keine wahren Verallgemeinerungen der Art ‚Normalerweise, wenn ein Fall die nicht-moralischen Eigenschaften  $R^1$ - $R^n$  hat, hat er die moralische Eigenschaft  $F$ ‘.

Da uns letzteres Problem noch im nächsten Kapitel etwas länger beschäftigen wird, wollen wir hier zunächst davon absehen und davon ausgehen, dass es keinen Grund gibt, anzunehmen, dass es die fraglichen wahren Verallgemeinerungen nicht gibt.<sup>183</sup>

Aus diesen vier Argumenten folgert Dancy, dass die Verwendung imaginärer Fälle in der Ethik unter der Prämisse eines Generalismus (oder auch ‚Universalismus‘) keinen guten Sinn macht. Unter ‚Generalismus‘ versteht Dancy dabei genau die Auffassung, die mit dem Verweis auf Wittgenstein bestritten wurde, nämlich dass es sehr wohl Verallgemeinerungen der oben genannten Art gibt, also allgemeine ethische Prinzipien, die für ganze Klassen von Entscheidungsfällen bzw. Situationstypen implizieren, wie die Handlungsoptionen ethisch zu beurteilen sind, bzw. wie in einer solchen Situation zu handeln ist.

Die dazu gegensätzliche Auffassung des „Partikularismus“, die davon ausgeht, dass wir moralische Urteile von Fall zu Fall neu treffen (ohne den Fall als von einem bestimmten Situationstyp zu klassifizieren und unter Prinzipien zu subsumieren), könne dieser Methode eine viel natürlichere Interpretation geben, wobei der Partikularist imaginäre Fälle allerdings nicht heranzieht, um an ihnen ein moralisches Urteil zu fällen, das er auf einen aktuellen Fall übertragen kann (denn dafür bräuchte er ja allgemeine Prinzipien, die eine solche Übertragung gestatten), sondern er nutzt imaginäre Fälle, um festzustellen, welche Eigenschaften von moralischer Relevanz sein *können*, und verbessert damit sein moralisches Urteil in aktuellen Fällen (da er eine vollständigere Checkliste hat, welche Eigenschaften eines Falles von moralischer Relevanz sein können und deshalb im vorliegenden Fall zu prüfen sind). (Verwirrender Weise bezeichnet Dancy auch die Aussage, dass eine bestimmte Eigenschaft eines Falles von moralischer Relevanz sein *kann* als „moralisches Prinzip“.)<sup>184</sup>

Nach Dancy hat diese Auffassung folgende Vorteile: (i.) sie erklärt, weshalb wir uns tatsächlich auf die nicht-moralischen Eigenschaften beziehen, wenn wir unser Verhalten in einem Fall moralisch rechtfertigen wollen – bestimmte nicht-morali-

---

<sup>183</sup> Es ist auch ohnehin nicht klar, ob Dancy diesen Punkt als eigenständiges Argument betrachtet, zumal er ja nachweisen möchte, dass die Methode, imaginäre Fälle in der Ethik heranzuziehen unter der Prämisse des Generalismus keinen guten Sinn macht. Der Generalismus geht ja davon aus, dass es solche Verallgemeinerungen gibt.

<sup>184</sup> Vgl. Gesang [122], Kapitel 5.

sche Eigenschaften können in bestimmten Fällen die moralischen Eigenschaften des Falls determinieren; (ii.) wir können erklären, warum wir imaginäre Fälle heranziehen, da aus imaginären Fällen – trotz aller von Dancy erwähnten Probleme – zumindest gelernt werden kann, welche nicht-moralischen Eigenschaften von Relevanz sein *können*; (iii.) wir können erklären, wie wir überhaupt zu Wissen von moralischen Prinzipien aus der Beobachtung von Einzelfällen gelangen – ein „moralisches Prinzip“ im Sinne des Partikularismus ist eben nur eine Möglichkeitsaussage (Eigenschaft X *kann* von moralischer Relevanz sein) und diese kann an Einzelfällen verifiziert werden; (iv.) wir können erklären, wieso „moralische Prinzipien“ im Sinne des Partikularismus notwendig sind, obwohl sie keine Verallgemeinerungen sind – es sind einfach Notwendigkeiten der Form  $\Box \diamond p$ , also nezesitierte Möglichkeitsaussagen; (v.) diese Auffassung ist kompatibel mit dem Partikularismus.

Nach Dancys Auffassung betrachten wir in der Ethik imaginäre Fälle also nicht zur Widerlegung allgemeiner Prinzipien, sondern zur Etablierung von Möglichkeitsaussagen über die moralische Relevanz nicht-moralischer Eigenschaften. Diese Analyse steht also zu unserer in Widerspruch.

Um zu beurteilen, ob Dancys Rekonstruktion erfolgreich gegen unsere Rekonstruktion durchgesetzt werden kann, ist zunächst festzustellen, dass Dancy die von uns als richtig erachtete Auffassung, nach der Gedankenexperimente in der Ethik hauptsächlich moralische Prinzipien kritisieren sollen, überhaupt nicht diskutiert, sondern in seiner Kritik einen ganz anders gearteten *Strohmann* in Schwierigkeiten bringt. Während unser Vorschlag einem falsifikationistischen Modell entspricht, ist Dancy Strohmann ein relativ naiver Abduktivist, der aus der Beurteilung eines Einzelfalls auf ein allgemeines Prinzip schließt. Insofern entspricht (iii.) nur dem Induktionsproblem, das Dancy auf verifikationistische Weise löst: da Einzelfälle allgemeine Aussagen nicht verifizieren können, haben imaginäre Fälle einen anderen Zweck: sie verifizieren Einzelaussagen. Unser falsifikationistisches Modell, bei dem wir davon ausgehen, dass imaginäre Fälle (in der Regel) herangezogen werden, um moralische Prinzipien zu *falsifizieren*<sup>185</sup>, erfüllt ebenfalls alle von Dancy genannten Punkte: (i.) wir können erklären, weshalb wir uns auf nicht-moralische Eigenschaften beziehen, wenn wir unser Verhalten rechtfertigen, da wir bewährte Prinzipien zur Rechtfertigung heranziehen können, die die nicht-moralischen Eigenschaften mit den moralischen in Zusammenhang bringen; (ii.) wir können erklären, warum wir imaginäre Fälle heranziehen, da moralische Prinzipien modaler Natur sind und auch auf imaginäre Fälle Anwendung finden; (iii.) wir können erklären, wie wir zu moralischem „Wissen“ gelangen, da wir ein Modell angeben können, wie wir moralische Prinzipien gegen moralische Intuitionen abwägen; (iv.) wir können erklären, wieso wahre moralische Prinzipien (falls es sie gibt) notwendig sind (nämlich *weil* sie strenge Verallgemeinerungen sind, die auch auf bloß mögliche, imaginäre Fälle zutreffen sollen); (v.) diese Auffassung ist sowohl mit dem Partikularismus, wie auch dem Generalismus kompatibel, da sie zwar hypo-

<sup>185</sup> Vgl. auch (Gähde [114], 194) zur Verifikations/Falsifikations-Asymmetrie bei Gedankenexperimenten in der Ethik. Vgl. auch Bunzl [46].

thetisch annimmt, dass es wahre Verallgemeinerungen in der Ethik gibt, jede davon aber einem Testverfahren unterzieht, das sehr wohl dazu führen kann, dass diese Annahme irgendwann aufgegeben wird.

Erläuterungsbedürftig ist (iii.). Zunächst einmal steht Dancys Einwand, dass wir die moralische Relevanz aller Eigenschaften eines Falles nicht in Absehung von seiner moralischen Beurteilung feststellen können, noch im Raum. Wenn wir uns an Thomsons Geiger erinnern, kann ein Abtreibungsgegner natürlich zunächst erwidern, dass ja gar nicht klar ist, dass tatsächlich (A) vertreten werden muss, um vom Personsein des Embryos darauf schließen zu können, dass die Interessen der Mutter keine entscheidende Rolle spielen, sondern nur so etwas wie

- (A\*) Das Lebensrecht einer Person (qua Person) wiegt in jedem Fall, der dem Fall einer ungewollten Schwangerschaft in allen moralisch relevanten Eigenschaften hinreichend ähnlich ist, moralisch schwerer als das Selbstbestimmungsrecht anderer Personen über ihren Körper.

Um (A\*) aber am Geiger-Fall zu prüfen, muss man zunächst sowohl den Geiger-Fall, wie auch den Fall der ungewollten Schwangerschaft hinsichtlich ihrer moralisch relevanten Eigenschaften beurteilen, um festzustellen, ob der Geiger-Fall unter (A\*) zu subsumieren ist. Hat man den Fall der ungewollten Schwangerschaft aber bezüglich seiner moralisch relevanten Eigenschaften beurteilt, scheint die Frage, was mit (A\*) und dem Geiger ist, völlig uninteressant. Um diesem Einwand angemessen zu begegnen, müssen wir unser falsifikationistisches Modell etwas weiter erläutern.

Angenommen, eine Theorie (T) über irgendeinen Gegenstandsbereich behauptet, dass zwei Eigenschaften F und G kovariieren. Um diese Theorie zu testen, führen wir die Testsituationen *a*, *b* und *c* herbei, bei denen eine der Eigenschaften (sagen wir F) instantiiert ist, und prüfen, ob die andere Eigenschaft dann ebenfalls auftritt. Damit ein solcher Test zuverlässig ist, müssen wir zunächst ausschließen, dass irgendwelche äußeren Umstände das Ergebnis verfälschen. Dies tun wir auf der Grundlage einer Reihe von Hintergrundannahmen darüber, was für das Ausbleiben oder Auftreten von G sonst noch relevant ist, darüber, wie die verwendeten Messinstrumente funktionieren, darüber, wie die Messergebnisse zu interpretieren sind, etc. Diese Annahmen sind im selben Sinne fehlbar, wie die zu testende Hypothese, weshalb das scheinbare Ausbleiben von G nicht logisch zwingend erfordert, dass wir T zurückweisen, sondern auch eine der Hintergrundannahmen zurückweisen können (unter Umständen auch die Hintergrundannahme, dass es außer den bekannten möglichen Störquellen für das Auftreten von G keine weiteren gibt.) Unser Urteil darüber, ob G nun in diesem Fall tatsächlich ausgeblieben ist, wie auch unser Urteil, ob damit T widersprochen wird, hängt also von einer Reihe von Hintergrundannahmen ab, die wir als gerechtfertigt annehmen müssen, wenn wir überhaupt zu einem Urteil über den Testfall kommen wollen. Da die Logik uns zunächst nicht vorschreibt, an welchen Hypothesen festzuhalten ist und welche aufgegeben werden müssen, ist das Beste, was wir in solchen Fällen tun können, unsere bisherigen Erfahrungen, unsere Hintergrundannahmen und

unsere Überzeugungen bezüglich T in ein reflektiertes Gleichgewicht zu bringen, um ein möglichst kohärentes Überzeugungssystem zu erreichen.<sup>186</sup>

Ein Überzeugungssystem ist umso kohärenter, je mehr die in diesem System repräsentierten Überzeugungen logisch miteinander verknüpft sind. Entsprechend ist ein Überzeugungssystem *mit* Prinzipien oder Gesetzesaussagen ein System, das in der Regel kohärenter ist als eines ohne diese, da Prinzipien und Gesetzesaussagen Verbindungen zwischen logisch unabhängigen Überzeugungen schaffen können, etc. Kohärenz ist dabei eine Tugend von Überzeugungssystemen, die nicht nur für unsere wissenschaftlichen Überzeugungen wünschenswert ist, sondern auch bezüglich unserer moralischen Überzeugungen in der Regel angestrebt wird.<sup>187</sup> So, wie es in den Testfällen *a*, *b* und *c* uns nicht genügt, zu *irgendeinem* Urteil zu kommen, sondern es darauf ankommt, Urteile zu fällen, die unser Überzeugungssystem maximal kohärent machen, geht es uns auch in moralischen Entscheidungen nicht bloß darum, zu *irgendeiner* Entscheidung zu kommen. In Ermangelung eines narrensicheren Indikators, was in einer Entscheidungssituation moralisch richtig ist, streben wir danach, zu möglichst *kohärenten* Entscheidungen zu kommen. Dies ist durchaus kompatibel mit der Auffassung, dass es wenige oder gar keine moralischen Prinzipien gibt, die über viele Entscheidungssituationen hinweg Anwendung finden (also durchaus kompatibel damit, dass der Partikularist im Recht ist). Diese Auffassung ist aber *nicht* kompatibel damit, dass wir in einem aktualen und einem imaginären Fall zu unterschiedlichen Urteilen kommen, ohne der gerechtfertigten Meinung zu sein, dass es zwischen den beiden Fällen einen moralisch relevanten Unterschied gibt. Deshalb ist unser Urteil bezüglich des imaginären Falles so lange relevant, wie sich kein moralisch relevanter Unterschied zwischen den beiden Fällen aufzeigen lässt, oder gezeigt werden kann, dass unser Urteil im Geiger-Fall zum Wohle der Gesamtkohärenz unseres Überzeugungssystems revidiert werden muss. Die Suche nach solch einem reflektierten Gleichgewicht ist dabei – wie wir gesehen haben – kein Sonderproblem der Ethik und auch keines imaginärer Fälle. Zu diesem Ergebnis kommt auch Häggqvist:

Dancy suggests that we use thought experiments in ethics mainly to decide whether a particular actual case has a certain property exhibited by the hypothetical case, and that we already presuppose some universalizable list of relevant properties. But a more realistic picture of what goes on in most ethical thought experiments is, I think, that the ethicist challenges some proposed list of relevant properties by concocting a hypothetical case satisfying the list but lacking the target property. Holism typically allows several alternative responses (which may of course vary widely in plausibility): rejecting the list, revising it, amending it, or insisting that the moral target property is, after all, present in the hypothetical case. But this holism is an in-

---

<sup>186</sup> Was dies im Einzelnen bedeutet, kann hier nicht erläutert werden. Stattdessen verweise ich auf Bartelborth [10].

<sup>187</sup> Vgl. Birnbacher [24]. Diese Bemerkung muss an dieser Stelle genügen. Warum dem so ist, und ob sich unser Kohärenzstreben erkenntnistheoretisch weiter rechtfertigen lässt, kann hier nicht diskutiert werden. Vgl. Bartelborth [10].

escapable condition both in testing by ordinary experiment and by thought experiment. It invalidates neither method. (Häggqvist [137], 45-46)

Auch Gähde kommt zu diesem Ergebnis:

[Die Intuitionen, die in ethischen Gedankenexperimenten bezüglich imaginärer Fälle abgefragt werden] fungieren, vereinfacht gesprochen, als *Testinstanzen*, an denen die Kompatibilität [eines zu prüfenden] Moralprinzips mit unserer *common sense*-Moral überprüft wird. Dabei sind weder das ethische Grundprinzip noch die moralischen Intuitionen sakrosankt. Im Fall eines Konflikts zwischen ethischem Grundprinzip und moralischer Intuition existieren vielmehr zwei Möglichkeiten: Zum einen kann ein solcher Konflikt die Modifikation oder Aufgabe des ethischen Grundprinzips erzwingen. Zum anderen können aber auch die moralischen Intuitionen im Licht der ethischen Theorie ihre *prima facie*-Plausibilität verlieren. Bekanntlich hat John Rawls [...] im Rahmen seiner Konzeption eines Überlegungsgleichgewichts dieses Wechselspiel zwischen ethischer Theorie und moralischer Intuition zu dem grundlegenden Prinzip der ethischen Theoriebildung erhoben. (Gähde [114], 193)

Dementsprechend scheint es auch keinen Grund zu geben, Dancys Rekonstruktion der Verwendungsweise imaginärer Fälle in der Ethik der unsrigen vorzuziehen. Unsere Rekonstruktion scheint deutlich fruchtbarer, da sie die Verwendung imaginärer Fälle in verschiedenen Wissensgebieten einheitlich behandeln kann, und hat außerdem den Vorteil, nicht von vornherein auf eine bestimmte meta-ethische Position inhaltlich festgelegt zu sein. Damit beschließen wir den Exkurs zu Dancys Rekonstruktion von Gedankenexperimenten in der Ethik.

#### 4.3 WEITERE ARTEN VON GEDANKENEXPERIMENTEN IN DER PHILOSOPHIE

Natürlich gibt es in der Philosophie auch Gedankenexperimente, die nicht als Kritik gedacht sind, sondern eher dazu dienen, die Konsequenzen einer Theorie deutlich zu machen. Wie Hempel schon beobachtet hatte, kann es zwischen solchen klärenden Gedankenexperimenten und Gedankenexperimenten zur Überzeugungsänderung fließende Übergänge geben. So kann man den Unterschied zwischen einer deskriptivistischen (traditionellen) Bedeutungstheorie und einer externalistischen Bedeutungstheorie an demselben<sup>188</sup> Gedankenexperiment veranschaulichen, welches Putnam gewählt hatte, um die traditionelle Bedeutungstheorie zu kritisieren. Verwendet man Gedankenexperimente in dieser Weise, so haben sie nicht die übliche Argumentform und treffen in der Regel auch keine Aussagen darüber, welche Beschreibung der imaginären Situation die plausible

---

<sup>188</sup> Bei dieser Redeweise von „demselben“ Gedankenexperiment handelt es sich um eine elliptische Formulierung, wie in Kapitel 3.1 erläutert wurde.

ist, ganz so, wie wir es in Kapitel 3.1 auch für naturwissenschaftliche Gedankenexperimente in klärender Absicht feststellen konnten.<sup>189</sup>

Manchmal werden auch *erläuternde Analogien* in der Philosophie als Gedankenexperiment bezeichnet, wie beispielsweise Platons Höhlengleichnis oder Leibniz' Uhrenanalogie. In diesen Fällen werden abstrakte Sachverhalte an tatsächlichen oder imaginären Fällen erläutert, die in gewissen Hinsichten Analogien zu diesen Sachverhalten aufweisen. So erläutert Platon unsere epistemische Situation in Bezug auf die Ideenwelt am analogen Fall von Gefangenen in einer Höhle, die nur die Schatten von Abbildern der wirklichen Gegenstände sehen; und Leibniz erläutert z.B. den Leib-Seele Parallelismus an der Analogie zweier unabhängig, aber parallel laufender Uhren, die irgendwann gleichzeitig in Gang gebracht wurden. Solche erläuternden Analogien finden sich auch in den Naturwissenschaften, wie beispielsweise beim „Schlüssel/Schloss-Prinzip“ der Biologie (z.B. zur Erläuterung der Wirkungsweise von Enzymen). Hier wie dort handelt es sich eigentlich nicht um Gedankenexperimente, auch nicht um Gedankenexperimente in klärender Absicht, da der imaginäre Sachverhalt nicht zum intendierten Gegenstandsbereich der fraglichen Theorie gehört (der Parallelismus ist keine Theorie über Uhren, die Biochemie redet nicht über Schlösser und Schlüssel). Sofern trennscharf angegeben kann, was in den Anwendungsbereich einer Theorie fällt, kann auch trennscharf zwischen klärenden Gedankenexperimenten und illustrierenden Analogien unterschieden werden (natürlich hängt für unsere Diskussion nichts davon ab, ob eine solche begriffliche Trennung möglich ist oder nicht).<sup>190</sup>

#### 4.3.1 EINE WEITERE ART VON GEDANKENEXPERIMENTEN IN DER PRAKTI-SCHEN PHILOSOPHIE?

Neben den bisher angeführten „Gedankenexperimenten“ der Philosophie findet sich noch eine weitere Variante in der analytischen Ethik, die nicht ohne Weiteres unter eine der bisher aufgewiesenen Funktionsweisen subsumiert werden kann. Worum es sich bei diesen „Gedankenexperimenten“ handelt, und welche methodologischen Probleme mit ihrer Funktionsweise verbunden sind, werden wir kurz zu klären haben.

Wie Birnbacher (vgl. Birnbacher [24]) erläutert, ist einer der Gründe, der bei einer „schwachen“<sup>191</sup> Begründung einer moralischen Norm zur Geltung kommen

<sup>189</sup> Für weitere Beispiele aus der Philosophie, vgl. Smart und Williams [300], in dem die Konsequenzen verschiedener utilitaristischer Positionen an imaginären Fällen erläutert werden.

<sup>190</sup> Ein Grenzfall von einem klärenden Gedankenexperiment und einer erläuternden Analogie scheint die Verwendungsweise hypothetischer bzw. idealisierter Verfahrensnormen in der politischen Philosophie zu sein. Vgl. Birnbacher [24], S. 86-87. Birnbacher bezeichnet auch solche Erläuterungen als „Gedankenexperimente“. Vgl. hierzu aber auch den nächsten Abschnitt.

<sup>191</sup> Eine „schwache Begründung“ unterscheidet sich von einer „starken Begründung“ dadurch, dass sie nur Plausibilitätsgründe liefert, keine zwingenden Gründe. Schwache Begründungen

kann, ihr *Anspruch auf Allgemeingültigkeit*. Dieser Anspruch fordert von moralischen Normen, dass sie prinzipiell jedem Verständigen gegenüber rechtfertigbar sind, wobei insbesondere die von der Norm in negativer Weise Betroffenen von Relevanz sind.

Inhaltlich eng verbunden mit diesem Allgemeingültigkeitsanspruch moralischer Normen ist das *Prinzip der umfassenden Interessenberücksichtigung*. Nach diesem Prinzip kann „die Wahl einer moralischen Norm nicht davon absehen, wie sich die Annahme und Befolgung der Norm auf das Wohlbefinden der von ihr – als Akteure oder Beteiligte – Betroffenen auswirkt“ (Birnbacher [24], 417). Bedauerlicherweise stehen uns keine gut bestätigten Theorien darüber zur Verfügung, wie das Wohlbefinden der Betroffenen sich bei Implementierung einer fraglichen Norm verändern würde. Daher greift man in verschiedenen Modellen der idealen Prinzipienwahl auf „Gedankenexperimente“ zurück, die solche Prognosen aus gut bestätigten Theorien ersetzen sollen. Beispiele für solche Modelle sind der „Schleier des Nichtwissens“ bei Harsanyi und Rawls, wobei der „Gedankenexperimentierer“ sich fragt, ob er eine bestimmte Norm für sich selbst wollen kann, gleichgültig in welcher der möglichen Betroffenen-Positionen er sich befindet. Weitere solche Modelle erfordern die hypothetische Identifikation mit allen von einer implementierten Norm faktisch Betroffenen (wie etwa bei Hare und C. I. Lewis).<sup>192</sup>

Hierbei handelt es sich offenbar um Verfahrensweisen, deren Funktion darin besteht, eine bestimmte Qualität einer vorgeschlagenen Norm (nämlich den Grad, zu dem sie in umfassender Weise die Interessen der Betroffenen berücksichtigt) auszuloten. Es geht hier offenbar nicht um ein Verfahren, das hauptsächlich der Überzeugungsänderung bzw. der Kritik vorgeschlagener Normen dienen soll.<sup>193</sup>

Auch diese „Gedankenexperimente“ können selbstverständlich methodologisch evaluiert werden. Ein Einwand methodologischer Art, der insbesondere gegen die zuletzt erwähnten Identifikationsmodelle vorgebracht wird, bezieht sich auf die logische Möglichkeit solcher Gedankenexperimente:

Im Gegensatz zum Modell des „Schleier des Nichtwissens“ sind die Identifikationsmodelle allerdings mit einer grundsätzlichen Schwierigkeit konfrontiert: der Schwierigkeit, dass eine vollständige Identifikation eines Wählenden mit den von einer bestimmten Norm Betroffenen an Grenzen der Denkbareit stößt. Verliert das Ich des Wählenden durch eine *vollständige* Identifikation mit den Betroffenen nicht jede Substantialität? Wie kann es es selbst bleiben, wenn es nicht nur – im Sinne eines „Rollentauschmodells“ – die *Rollen* der jeweils Betroffenen übernehmen soll, sondern auch deren Präferenzen, Werte, Normen und Ideale? (Birnbacher [24], 420)

---

lassen es daher zu, dass auch ein rationaler Akteur sich gegen eine schwach begründete Norm entscheiden kann. Vgl. Birnbacher [24].

<sup>192</sup> Vgl. Birnbacher [24], 419.

<sup>193</sup> Dieser Punkt wird von Gähde [114] völlig übersehen, weshalb seine methodologischen Bemerkungen zu Gedankenexperimenten in der Ethik an dieser Klasse von Gedankenexperimenten vorbeigehen.

Hierbei handelt es sich allerdings um ein Scheinproblem. Es ist zwar zutreffend, dass – gegeben eine bestimmte reduktionistische Auffassung personaler Identität<sup>194</sup> – eine Person nicht alle empirisch feststellbaren Attribute einer anderen Person *annehmen* kann, ohne dabei die andere Person zu werden, das hat aber zunächst *nichts* damit zu tun, dass man sich in eine andere Person nicht vollständig hineinversetzen kann.<sup>195</sup> Birnbacher unterscheidet hier nicht zwischen dem Umstand, dass eine Person alle empirisch feststellbaren Attribute einer anderen Person *hat*, was denkmöglich sein könnte<sup>196</sup>, und dem Umstand, dass eine Person überlegt, wie sie sich fühlen würde, wäre sie von den Konsequenzen einer bestimmten Norm so betroffen, wie die Person, deren Überzeugungen und Werthaltungen sie sich hypothetisch zu eigen macht, was keine Denkmöglichkeit darzustellen scheint. Es entsteht dabei auch kein weitergehendes (essentialistisch konstruiertes) Problem, nach dem es darum ginge, dass eine Person *unmöglich* eine andere Person *sein* kann, also auch *hypothetisch* keine andere Person *sein* kann. Obwohl es „oberflächengrammatisch“ so scheint, als verlange das Identifikationsmodell die hypothetische Vorstellung einer metaphysischen Unmöglichkeit, ist sie schlimmstenfalls missverständlich formuliert. Williams, der sich in ‚Vorstellungskraft und Selbst‘ (Williams [347], 47-77) mit ganz ähnlich gelagerten Problem beschäftigt, scheint mir genau dasselbe sagen zu wollen:

Doch wo das Problem darin besteht, dass ich mir beispielsweise vorstelle, Napoleon zu sein, führt die Wendung „sich vorstellen, selbst Napoleon zu sein“ möglicherweise in die Irre. Sie führt uns nahe an eine Wendung heran, die man ebenfalls gebrauchen und die noch irreführender sein kann, obgleich sie es natürlich nur dann ist, wenn man über sie nachzudenken beginnt. Dies ist die Wendung, dass ich Napoleon bin (oder war)“. Denn mit Bezug auf diese Wendung hat man vielleicht das Gefühl fragen zu müssen, was dieses „ich“ ist, das innerhalb des Ausdrucks meiner Vorstellung auftaucht. [...] Wenn diesen Phantasievorstellungen kein Einhalt geboten wird, so wird der ihnen angemessene Vorstellungsmodus am wenigsten irreführend durch die Worte „sich vorstellen, Napoleon zu sein“ ausgedrückt. Was dies ausdrückt, nämlich das vorgestellte Spielen der Rolle Napoleons, ist der einzige Modus, der überhaupt dazu fähig ist, die Spekulationen, die wir erörtert haben, zu tragen. Und wenn man diesen Modus richtig versteht, dann führt er kein weiteres „ich“ ein, das jene Schwierigkeiten erzeugt: wie gesagt, hier sind nur zwei Personen im Spiel, die wirkliche Person Ich und Napoleon. (Williams [347], 76-77)

Mehr als dieser (absolut kohärente) Modus ist für die Identifikationsmodelle aber auch nicht erforderlich. Es geht schließlich nur darum, abzuschätzen, wie sich eine

---

<sup>194</sup> Dies ist der Unterschied zwischen dem so genannten „Simple View“ und dem „Empiricist View“, den wir in Kapitel 5.3.1.2 noch genauer besprechen werden.

<sup>195</sup> Hier geht es um logische Möglichkeit. Es kann natürlich aus irgendwelchen psychologischen (oder anderen kontingenten) Gründen so sein, dass man sich nicht vollständig in eine andere Person hineinversetzen kann.

<sup>196</sup> Vgl. aber hierzu das letzte Kapitel dieser Arbeit.

Person in einer in Bezug auf eine Normkonsequenz bestimmten Rolle (gegeben ein bestimmtes Hintergrundwissen und bestimmte Werte) fühlen würde.<sup>197</sup>

Ein tatsächlich methodologisches Problem, das für das Modell des „Schleiers des Nichtwissens“, wie auch für die angesprochenen Identifikationsmodelle gleichermaßen zu bestehen scheint, hat eher mit den psychologischen Hintergrundannahmen dieser Verfahren zu tun als mit eventuellen metaphysischen Problemen. Wie wir gesagt haben, geht es eigentlich darum, eine bestimmte Qualität einer Norm zu beurteilen. Die eigentliche Frage ist: wie würden sich die Betroffenen Individuen fühlen, wäre die Norm implementiert (bzw., wie würde ich mich als Betroffener fühlen, befände ich mich in dieser oder jener Rolle)? Es geht also um eine Abschätzung des zukünftigen Wohlbefindens bestimmter Personen unter bestimmten hypothetischen Umständen. Da wir keine Theorien haben, die uns sichere Prognosen über zukünftiges Wohlbefinden bei variierenden Randbedingungen erlauben, greifen wir zu einem anderen Instrument. Gegeben, dass wir unsere eigenen Entscheidungen bei der Wahl zwischen verschiedenen Handlungsoptionen häufig dadurch beurteilen, dass wir uns überlegen, wie unser zukünftiges Ich sich fühlen würde, würden diese oder jene Konsequenzen einer solchen Entscheidung eintreten, und gegeben, dass wir das Wohlbefinden anderer häufig dadurch beurteilen, dass wir uns in sie hineinversetzen, macht es *prima facie* Sinn, diese beiden Verfahren zu kombinieren.<sup>198</sup>

Allerdings – und das *ist* von methodologischer Relevanz – sollte man beachten, dass diese Prognosen in hohem Maße unzuverlässig sind. Insbesondere in den letzten beiden Jahrzehnten sind zum Thema „affective forecasting“, also zu der Fähigkeit, die Dauer und Intensität des *eigenen* zukünftigen Wohlbefindens für den Fall des Eintreffens eines bestimmten Ereignisses oder Umstandes<sup>199</sup> vorherzusagen, eine Reihe psychologischer Untersuchungen gemacht worden, die sämtlich darauf hinweisen, dass wir relativ schlecht darin sind, uns in unser zukünftiges Selbst hineinzuversetzen. Obwohl wir in der Regel die Wertigkeit eines zukünftigen Gemütszustandes korrekt prognostizieren, also richtig liegen, wenn es um die Frage geht, ob ein Ereignis oder ein Umstand eher positive oder eher negative Gefühle verursachen wird, sind wir weniger gut, wenn es darum geht, abzuschätzen, welche Gefühlsmischung genau ein Ereignis auslösen wird, und sehr schlecht, wenn es darum geht, die *Dauer* und *Intensität* eines zukünftigen Gefühlszustandes vorherzusagen. Es kommt dabei zum sogenannten „impact bias“ (vgl. Wilson und Gilbert [349]), einer Überschätzung des andauernden Einflusses, den zukünftige

---

<sup>197</sup> Falls eine andere Interpretation – wie Williams meint – tatsächlich in logische Unmöglichkeiten führt, sollte es schon das *principle of charity* gebieten, die Identifikationsmodelle in diesem unproblematischen Modus zu interpretieren. Vgl. zu diesem Problem auch Reynolds [279].

<sup>198</sup> Natürlich tun wir dies im Alltag ohnehin ständig.

<sup>199</sup> Dabei geht es nicht nur um die Einschätzung zukünftiger punktueller Ereignisse und deren Nachwirkung, sondern auch um die Beurteilung von zukünftig eintretenden Dauerzuständen (Arbeitslosigkeit, Behinderung, etc.) So konnte gezeigt werden, dass amerikanische Assistenzprofessoren den Zustand nach der *tenure* gemäß dem (gleich zu erläuternden) *impact bias* völlig überschätzen. Vgl. Gilbert und Wilson [124], Wilson und Gilbert [349].

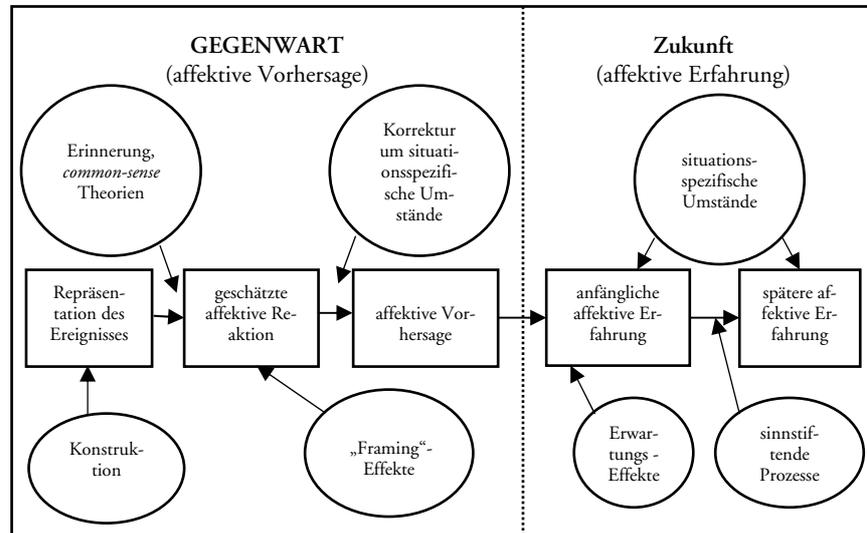


Abbildung 4.3-1

Ereignisse auf unser Gefühlsleben haben, sowie zu einer Reihe unsystematischer Fehlleistungen.

In Abbildung 4.3-1 sind die von Psychologen aufgewiesenen möglichen Fehlerquellen im Prognoseprozess in Kreisen dargestellt (Darstellung nach Wilson und Gilbert [349]). Der Prognoseprozess ist dabei in drei Komponenten aufgeteilt, (i.) die Repräsentation des Ereignisses oder Umstandes oder Zustandes, bezüglich dessen man eine Prognose abgeben soll, (ii.) die Abschätzung des eigenen Wohlbefindens in den so vorgestellten Umständen, (iii.) die eigentliche Prognose. Die Erfahrungssituation ist ihrerseits in zwei Komponenten aufgebrochen, (i.) die anfängliche affektive Erfahrung, wenn das Ereignis oder der Umstand eintritt, (ii.) die Erfahrung des Umstandes über einen bestimmten Zeitraum, bzw. die Nachwirkung eines bestimmten Ereignisses auf das Gefühlsleben. Zwischen diesen fünf Komponenten können verschiedene Einflüsse dafür sorgen, dass man sein zukünftiges Wohlbefinden in unzutreffender Weise prognostiziert.

Schon bei der mentalen Repräsentation des zukünftigen Ereignisses kann man das Ereignis schlicht unzutreffend vorstellen. Man stellt sich den Besuch eines Openair Konzertes im kommenden August als einen warmen Sommertag vor, während es dann tatsächlich ein fürchterlich verregneter Tag wird. Diese Fehlerquelle ist unsystematisch, da sie weder zum „impact bias“ beiträgt noch mit signifikanter Häufigkeit einen Fehler in einer bestimmten Richtung produziert. Es kann sowohl sein, dass man das Ereignis im positiven Sinne fehlerhaft konstruiert, wie auch im negativen.

Eine andere Fehlerquelle, die bereits Einfluss auf die Repräsentation des Ereignisses hat, subsumieren Gilbert und Wilson unter den „framing-effect“<sup>200</sup>. Befragt man Testpersonen danach, wie sie sich unter bestimmten geänderten Umständen fühlen würden, werden die Befragten ihre Aufmerksamkeit bei der Konstruktion des zukünftigen Umstandes auf die genannten Änderungen richten und weniger Rücksicht auf all diejenigen Dinge legen, die gleich bleiben. Dies ist eine *systematische* Fehlerquelle, da sie zum „impact bias“ beiträgt.

Weitere Fehlerquellen sind natürlich unzutreffende Theorien darüber, wie man sich unter bestimmten Umständen fühlt. Eine statistisch gesehen stark verbreitete Überzeugung scheint z.B. zu sein, dass Frauen während ihrer Menstruationsphasen übelläunig sind (wobei Frauen auch aus der Erinnerung berichten, dass dem in der Regel so war), während empirische Untersuchungen zeigen, dass sie während der Menstruationsphasen nicht übelläuniger sind als sonst (Wilson und Gilbert [349], McFarland, Ross und DeCourville [205]).

Während falsche Theorien eine unsystematische Fehlerquelle darstellen, trägt die Heranziehung von Erinnerungen an vergangene Erlebnisse ähnlicher Umstände zur Prognose zum *impact bias* bei, da sich in der Erinnerung die Intensität einer emotionalen Reaktion stärker durchsetzt als die tatsächliche Dauer der Reaktion.

Ein weiterer Einflussfaktor ist der Zustand, in dem man sich selbst zum Zeitpunkt der Prognose befindet. Hat man gerade gut gefrühstückt, dann *unterschätzt* man leicht die Menge an Nahrungsmitteln, die man in den nächsten Tagen brauchen wird, um das Wohlbefinden sicher zu stellen. Knurrt der Magen beim Supermarktbesuch, wird man die Menge *überschätzen*. Eine genaue Beurteilung, in welchem Ausmaß der momentane Zustand die Intuition in die Irre führt, scheint schwierig. Diese Fehlerquelle kann sogar zur Fehleinschätzung der Wertigkeit eines zukünftigen Ereignisses führen, wenn keine Korrektur vorgenommen wird (z.B. wenn man schwer an Grippe erkrankt abschätzt, wie man sich in zwei Wochen bei einer Party fühlen wird).

Tritt der Umstand bzw. das Ereignis dann ein, wird die emotionale Wahrnehmung in nicht unerheblichem Ausmaß durch unsere Erwartung beeinflusst. Erwartet man einen lustigen Film im Kino, wird man den Film lustiger finden, als wenn man keine solchen Erwartungen hat. Diese Fehlerquelle ist ebenfalls unsystematisch, da sich solche Erwartungshaltungen aus allen möglichen Gründen bilden können (eine systematische Fehlerquelle dieser Art ist dabei die *self-fulfilling prophecy*).

Zwei weniger offensichtliche Fehlerquellen sind eine Überschätzung der Relevanz, die ein Ereignis in Relation zu allen anderen Ereignissen des Alltags hat, und eine Unterschätzung des „psychischen Immunsystems“, das uns hilft, negative Erlebnisse schnell zu verarbeiten. Beide Faktoren tragen zum *impact bias* bei.

Wenn wir prognostizieren sollen, welchen Einfluss ein zukünftiges Ereignis auf unser Wohlbefinden hat, beachten wir in der Regel nicht, dass das Ereignis in einen Alltag eingebettet ist, in dem noch viele andere Dinge passieren, die unser

---

<sup>200</sup> Vgl. Tversky und Kahneman [331].

Wohlbefinden ebenfalls beeinflussen. „Focalism“ bezeichnet diese Art von systematischem Fehler, bei dem wir bei der Prognose das zu bewertende Ereignis nur in einem Vakuum betrachten. Dieser Fehler wirkt sich insbesondere auf die Fehleinschätzung der Dauer eines Emotionszustandes aus.

Unser psychisches Immunsystem – andererseits – bewirkt, dass wir negative (wie positive) Erlebnisse rasch verarbeiten, indem wir Ereignisse rationalisieren oder relativieren. Der positive wie negative Einfluss von Ereignissen oder Umständen auf unser Wohlbefinden wird dadurch verkürzt, ohne dass wir uns dieser Mechanismen bewusst wären:

Our basic argument is that people make sense of their worlds in a way that speeds recovery from emotional events, and that this sense-making process is largely automatic and nonconscious. Humans inexorably explain and understand events that were initially surprising and unpredictable, and this process lowers the intensity of emotional reactions to the events. Human beings are adept at orienting to important, novel events in their environment and then transforming these events psychologically in order to understand them better. (Wilson und Gilbert [349], 371)

Dies soll als Übersicht zunächst genügen. Der Befund, dass wir in gewissen Hinsichten mit hoher Wahrscheinlichkeit Fehler machen, wenn wir unser eigenes zukünftiges Wohlbefinden für bestimmte hypothetische Umstände prognostizieren sollen, ist von direkter Relevanz für den „Schleier des Nichtwissens“, da bei diesem Modell ja nach dem prognostizierten *eigenen* Wohlbefinden unter bestimmten hypothetischen Umständen gefragt wird, und indirekt relevant für die erwähnten Identifikationsmodelle, da ein schlechtes Ergebnis bei der Prognose des eigenen zukünftigen Wohlbefindens darauf schließen lässt, dass es uns sicherlich nicht leichter fallen wird, das zukünftige Wohlbefinden anderer zu prognostizieren – insbesondere, wenn es explizit nach derselben Methode prognostiziert werden soll.

Insofern keine alternative Methode zur Verfügung steht, den Grad der umfassenden Interessenberücksichtigung einer vorgeschlagenen Norm zu evaluieren, sollte dieses Ergebnis allerdings nicht dazu veranlassen die genannten Modelle aufzugeben, sondern eher dazu, sich zu fragen, wie diese Methode verbessert werden kann. Dies scheint durchaus möglich zu sein.

Wie wir gesagt haben, ist ein hervorstechendes Phänomen der so genannte *impact bias*, d.h. unabhängig davon, welche Fehler uns sonst noch bei der Prognose unterlaufen, in der Regel werden wir die Intensität und Dauer des Einflusses eines Ereignisses oder Umstandes auf unser zukünftiges Wohlbefinden überschätzen. Wenn sich die Vermutungen von Wilson et al. bestätigen lassen, würden wir einen bestimmten systematischen Fehler machen, den wir dann natürlich im Nachhinein systematisch korrigieren können. (Etwa durch eine methodische Regel, die uns anweist, die prognostizierte Dauer und Intensität einer emotionalen Reaktion einer Gruppe von Betroffenen auf einen bestimmten Umstand oder ein bestimmtes Ereignis erst um ein bestimmtes Verhältnis nach unten zu korrigieren, bevor man den Grad der umfassenden Interessenberücksichtigung der fraglichen Norm auf dieser Grundlage festlegt.)

Die psychologischen Experimente haben außerdem gezeigt, dass Individuen, die vor der Abgabe einer Prognose darüber, wie ein bestimmtes Ereignis ihr Wohlbefinden in einer bestimmten Zeitspanne beeinflussen würde, zuerst gebeten wurden, aufzulisten, was sie meinen, das in dieser Zeitspanne ihre Aufmerksamkeit beanspruchen wird, deutlich bessere Prognosen abgeben (Wilson und Gilbert [349]). Legt man vorher ein hypothetisches Tagebuch an, wie der Tag nach dem großen Länderspiel im Einzelnen ablaufen wird, wird einem selber klar, dass ein großer Erfolg des eigenen Teams am Vortag die Gesamtstimmung weitaus weniger positiv beeinflussen wird, als man dies meinen könnte, wenn man sich den großen Erfolg in einem Vakuum vorstellt. D.h., insofern unsere Fehler systematisch sind, können wir sie auch systematisch verbessern.

In den anderen Fällen, in denen unsere Prognosefehler von unsystematischen Einflüssen verursacht werden, kann es helfen diese Fehlerquellen zu kennen, um sie einzudämmen. So, wie eine über mehrere Tage zustande gekommene Einkaufsliste einen vor den Tücken plötzlicher Hungerattacken beim Wocheneinkauf bewahrt, mag man auch bei der Beurteilung von Normen auf strategische Weise seine Prognosefähigkeit verbessern können. *Wie* man dies tun kann, wird die empirische Psychologie beantworten müssen. Hier ging es nur darum, aufzuzeigen, dass *diese* methodologischen Fragen psychologische Fragen sind, weil es bei diesen „Gedankenexperimenten“ um psychologische Prognosen geht. Ob man hierbei tatsächlich von „Gedankenexperimenten“ reden mag, sei jedem selbst überlassen.

Wie wir gesehen haben, scheint es sich jedenfalls bei den paradigmatischen<sup>201</sup> „Gedankenexperimenten“ in der Philosophie um Gedankenexperimente zu handeln, wie wir sie in den Naturwissenschaften identifizieren konnten. D.h. gemäß der vorgenommenen Präzisierung des Ausdrucks ‚Gedankenexperiment‘ aus Kapitel 3, kann man auch in der Philosophie von „Gedankenexperimenten“ reden, ohne sich einer begrifflichen Verwirrung schuldig zu machen. Die nächste Frage wäre nun, inwiefern sich methodologische Einsichten aus der Wissenschaftstheorie der Naturwissenschaften vernünftigerweise auf philosophische Gedankenexperimente übertragen lassen.

Wie wir in Kapitel 3 gesehen haben, ist die Frage der Rechtfertigung bestimmter Urteile in naturwissenschaftlichen Gedankenexperimenten nicht hinreichend beantwortet. Da es bei philosophischen Gedankenexperimenten um andere Gegenstände geht, könnte sich dieses Problem (trotz der zugestandenen Ähnlichkeit in anderen Hinsichten) für philosophische Gedankenexperimente auf andere Weise stellen. Dabei ist sowohl möglich, dass es sich bei philosophischen Gedankenexperimenten plausibler lösen lässt, wie auch, dass es sich noch schlechter lösen lässt. Aus diesem Grunde werden wir uns in den nächsten Kapiteln verschiedene Arten der Kritik an der Methode des Gedankenexperiments in der Philosophie ansehen.

---

<sup>201</sup> Wir hätten diese Auswahl noch beliebig auf weitere systematische Bereiche der Philosophie erweitern können (wie etwa der Ästhetik), haben uns aber auf die bekannteren Fälle beschränkt. Für eine eindrucksvolle Sammlung philosophischer Gedankenexperimente aus allen Bereichen der Philosophie vgl. Tittle [329].

## 5. Kritik an philosophischen Gedankenexperimenten im Allgemeinen

Gedankenexperimente in der Philosophie haben im Großen und Ganzen eine schlechte Presse, insbesondere im Vergleich mit ihren Gegenstücken in den Naturwissenschaften:

Thought experiments are found both in science and philosophy. In science [...] they can be helpful; and they tend to be relatively unproblematic. [...] In philosophy on the other hand, and in particular in the domain of the philosophy of mind, they can be – in fact they usually are – both problematic and positively misleading. (Wilkes [345], 1-2)

We have elucidated two grounds for suspecting a particular thought experiment to be a poor one. The grounds are relevant to philosophical and scientific thought experiments alike. However, our examples suggest that in science the damage is restricted [...]. (Peijnenburg und Atkinson [252], 317)

Im Folgenden wollen wir genauer betrachten, worauf diese Einschätzung beruht. Warum zweifelt man an der Brauchbarkeit philosophischer Gedankenexperimente? Warum glaubt man, ihre Überzeugungskraft sei fragwürdig?

In den Teilen 2 und 3 dieser Untersuchung haben wir Einiges über Gedankenexperimente in den Naturwissenschaften erfahren. In Teil 4 haben wir feststellen können, dass Gedankenexperimente in Philosophie und Naturwissenschaften ähnliche Zwecke mit ähnlichen Mitteln verfolgen. Schneiden philosophische Gedankenexperimente dabei *grosso modo* schlechter ab? Wenn ja, warum könnte das so sein? (Wohlgemerkt ist dies alles andere als offensichtlich. Mit den Naturwissenschaften assoziiert man eher eine empirische Vorgehensweise, mit der Philosophie eher eine nicht-empirische. Dass ausgerechnet *Gedankenexperimente* dann in den Naturwissenschaften besser funktionieren sollen, ist erstmal überraschend!)

Als Vorgehensweise bietet es sich an, vom Allgemeineren zum Speziellen fortzuschreiten. D.h., wir werden die möglichen Gründe, deretwegen man philosophischen Gedankenexperimenten misstrauen könnte, in eine systematische Ordnung bringen – falls es einen sehr allgemeinen Grund gibt, warum man philosophischen Gedankenexperimenten nicht trauen sollte, erübrigt es sich, speziellere genauer unter die Lupe zu nehmen. In diesem Teil der Untersuchung werden wir uns daher zunächst allgemeinen Kritiken an der Methode des Gedankenexperiments in der Philosophie zuwenden.

### 5.1 ES GIBT KEINE ERKLÄRUNG, WESHALB GEDANKENEXPERIMENTE IN DER PHILOSOPHIE FUNKTIONIEREN SOLLTEN

Einer der allgemeinsten Vorwürfe gegen die Verwendungsweise von Gedankenexperimenten in der Philosophie besteht in dem Hinweis, dass es keine allgemein akzeptierte Theorie darüber gibt, wie Gedankenexperimente funktionieren. Solange es keine Erklärung dafür gibt, weshalb man den in Gedankenexperimenten abgefragten Intuitionen irgendein Vertrauen entgegenbringen sollte, hätten diese Intuitionen überhaupt keine methodologische Rolle zu spielen. Bei Verena Mayer, insbesondere<sup>202</sup> aber bei Jaakko Hintikka<sup>203</sup> findet sich genau dieser, allgemeinste Vorwurf:

Analytische Philosophen [...] verwenden Gedankenexperimente in der Regel, ohne sich auf eine Diskussion der Beziehung zwischen Sprachgebrauch, Denken und Wirklichkeit einzulassen. So können wir zwar annehmen, dass Gettier die Ergebnisse seiner Experimente als allgemeingültige, empirisch nicht überprüfungsbedürftige Fakten verstanden hat, die aussagekräftig für die Wirklichkeit sind – denn er schloss aus ihnen, dass eine wahre, gerechtfertigte Überzeugung keine hinreichende Bedingung für Wissen darstellt. Aber es fehlt eine Theorie, die diesen Schluss rechtfertigt [...]. Wer den Schritt vom Sprachgebrauch zu den Fakten nicht theoretisch durchdacht und explizit gemacht hat, sollte auf Gedankenexperimente besser verzichten. (Mayer [201], 376)

I find [appeals to intuitions in support of the writer's views] scandalous. In the past, every major philosopher who appealed to intuitions had a theory or at least an explanation of why it is that we can obtain new knowledge or insight by reflecting on our own ideas. Aristotle found a basis for such appeals to intuition in his theory of thinking as a genuine realization of forms in the thinker's soul. Descartes found it in the theory of innate ideas, and Kant found it in his transcendental theory of mathematical relations as having been imposed on objects by ourselves in the act of sense-perception, which makes them intuitively knowable, that is, recoverable by means of what we would call intuition. But contemporary uses of intuition in philosophy are seldom backed up by any such justification. This is enough to make them highly suspect. [...] I am tempted to suggest that there should be a long moratorium on unanalyzed appeals to intuition in philosophical argumentation, except when the writer can provide some further reasons for us to believe in his or her intuitions. (Hintikka [148], 257-258)

Eine solche Argumentation ist hanebüchen. Die Tatsache, dass eine Erkenntnis-methode keine allgemein akzeptierte Rechtfertigung besitzt, ist *für sich genommen*

---

<sup>202</sup> Bei Verena Mayer ist der Vorwurf insofern weniger allgemein, als sie nachweist, dass gewisse Theorien über die Funktionsweise des Gedankenexperiments nicht sonderlich plausibel sind.

<sup>203</sup> Vgl. zu Hintikkas Position auch Hintikka [149].

kein hinreichender Grund, um von der Verwendung dieser Methode abzuraten. Ein Defizit in der Wissenschaftstheorie bzw. Erkenntnistheorie ist irrelevant für die Funktionstüchtigkeit von Forschungsverfahren in anderen Bereichen. Man erinnere sich an Hume und die Schlussfolgerung, die er aus der Beobachtung zog, dass die Methode der Induktion nicht befriedigend gerechtfertigt werden kann:<sup>204</sup>

Though we should conclude [...], that, in all reasoning from experience, there is a step taken by the mind which is not supported by any argument or process of the understanding; there is no danger that these reasonings, on which almost all knowledge depends, will ever be affected by such a discovery. (Hume [155], 41)

Die Tatsache, dass wir für den Erkenntnisgewinn aus Erfahrung keine Rechtfertigung besitzen, mag ein „major scandal“<sup>205</sup> für die Philosophie sein, aber keiner, der die empirischen Wissenschaften betrifft. Diese sind erst dann alarmiert, wenn es Gründe gibt, die *gegen* die Verwendung einer Methode sprechen. Wenn man z.B. *nachweisen* kann, dass eine Methode unzuverlässig ist.

Aus diesem Grund kann man auch nicht mit gutem Recht von jedem Philosophen verlangen, dass er *seine* Partizipation am Methodensortiment der Disziplin nochmals eigens durchdacht und gerechtfertigt haben muss. Man muss weder den Modus Ponens rechtfertigen können, noch zuvor eine Fundierung der logischen Konstanten leisten, wenn man ein Argument mit dem Anspruch auf Gültigkeit vorträgt, warum sollte das beim Gebrauch des Gedankenexperiments anders sein?

Hier kommt außerdem eine Asymmetrie zum Tragen, die wir schon angesprochen haben: Bei Popper haben wir die Auffassung kennen gelernt, dass kritische Methoden unter liberaleren methodologischen Standards stehen als apologetische Methoden. Dies ist – wie wir noch sehen werden – hier tatsächlich der Fall. Gedankenexperimente sind – wie wir schon gesehen haben – Instrumente zur Kritik philosophischer Positionen. Wenn es hier also eine Asymmetrie im Sinne Poppers gibt, dann sollte die Beweislast auf den Schultern desjenigen liegen, der die Zuverlässigkeit von Gedankenexperimenten in Abrede stellt (und sich damit vor einem Instrument der Kritik immunisiert), nicht aber auf den Schultern desjenigen, der sie verwendet.

Wenn man davon ausgehen darf, dass die Philosophie auch in denjenigen Bereichen, in denen sie mit Gedankenexperimenten forscht, Fortschritte gemacht hat<sup>206</sup>, dann besteht die Aufgabe eines Kritikers der Methode des Gedankenexperiments darin, nachzuweisen, dass es sich bei den Erkenntnisfortschritten um pures Glück gehandelt haben muss. Dieser Aufgabe kann er dadurch nachkommen,

---

<sup>204</sup> Diese Analogie hinkt insofern, als Hume sogar der Meinung war, dass es prinzipielle Gründe gibt, weshalb man die Induktion niemals rechtfertigen können. Der Einwand Hintikkas besagt bisher nur, dass ihm keine Rechtfertigung bekannt ist. (Allerdings spricht das auch nur *gegen* Hintikka.)

<sup>205</sup> Broad [36], 152.

<sup>206</sup> Ich denke, es ist offensichtlich, dass die analytische Philosophie in allen Bereichen in den letzten hundert Jahren signifikante Fortschritte gemacht hat. Ich werde an dieser Stelle aber nicht dafür argumentieren.

dass er nachweist, dass es Gründe gibt der Methode zu misstrauen, nicht aber indem er bloß konstatiert, dass es keine allgemein geteilten Gründe gibt, ihr zu trauen. Solche Gründe (der Methode zu misstrauen) wollen wir uns nun ansehen.

## 5.2 WIR WISSEN EINFACH NICHT, WAS WIR SAGEN WÜRDEN

Ein häufiger Einwand gegen die Methode des Gedankenexperiments ist, dass wir „einfach nicht wissen, was wir sagen würden“, wenn wir mit einem kontrafaktischen Fall konfrontiert werden. Ist Jones dieselbe Person wie Brown-Jones? Hat Mary neues Tatsachenwissen erworben? Versteht Searle Chinesisch?

Dieser Einwand kann auf zwei Weisen gemeint sein: (a) es gibt keine Tatsache in Bezug darauf, was wir sagen würden, und (b): obschon es eine Tatsache gibt, was wir sagen würden, können wir durch die Methode des Gedankenexperiments nichts darüber herausfinden. Ersteres bedeutet, dass bloß mögliche Sachverhalte außerhalb des Anwendungsbereichs unserer Sprache liegen, Letzteres, dass sie zwar nicht außerhalb ihres Anwendungsbereichs liegen, dass wir aber keinen Zugang dazu haben, wie die Sprache auf die bloß möglichen Fälle angewendet würde.

### 5.2.1 ES GIBT KEINE TATSACHE IN BEZUG DARAUF, WAS WIR SAGEN WÜRDEN

Dieser Vorwurf scheint insbesondere von Quine und Wittgenstein erhoben worden zu sein:

The method of imaginary cases has its uses in philosophy, but at points [...] I wonder whether the limits of the method are properly heeded. To seek what is “logically required” for sameness of person under unprecedented circumstances is to suggest that words have some logical force beyond what our past needs have invested them with. (Quine [267], 490)

„Es ist als wären unsere Begriffe bedingt durch ein Gerüst von Tatsachen.“

Das hieße doch: Wenn du dir gewisse Tatsachen anders denkst, sie anders beschreibst, als sie sind, dann kannst du die Anwendung gewisser Begriffe dir nicht mehr vorstellen, weil die Regeln ihrer Anwendung kein Analogon unter den neuen Umständen haben. – Was ich sage, kommt also darauf hinaus: Ein Gesetz wird für Menschen gegeben, und ein Jurist mag wohl fähig sein, Konsequenzen für jeden Fall zu ziehen, der ihm gewöhnlich vorkommt, das Gesetz hat also offenbar seine Verwendung, einen Sinn. Trotzdem aber setzt seine Gültigkeit allerlei voraus; und wenn das Wesen, welches er zu richten hat, ganz vom gewöhnlichen Menschen abweicht, dann wird z.B. die Entscheidung, ob er eine Tat mit böser Absicht begangen hat, nicht etwa schwer, sondern (einfach) unmöglich werden. (Wittgenstein [351], Zettel, § 350)

Diese kurzen Bemerkungen legen die These nahe, dass es für Situationen, die kontingenter- oder notwendigerweise bisher noch nicht aufgetreten sind, keine festen Regeln zu ihrer korrekten Beschreibung *gibt*. Nach dieser Auffassung ist die Sprache ein Instrument, das entwickelt worden ist (bzw. sich entwickelt hat), um bestimmte Zwecke zu erfüllen, insbesondere, um bestimmte tatsächlich auftretende Situationen beschreiben zu können. Für diese Situationen gibt es Regeln, die bestimmen, ob eine Situation korrekt oder inkorrekt beschrieben ist. Für nicht-auftretende Situationen gibt es keine solchen Regeln, folglich gibt es keine Tatsache, was wir in solchen Situationen sagen würden.

Wie Sören Häggqvist argumentiert hat, ist diese Auffassung *prima facie* nicht sonderlich plausibel:

Clearly “past needs” suffice to invest words with enough “logical force” to allow a fairly determinate answer to the question whether they apply to mundane actual situations. Encountering a new cat, I call it a “cat.” And the same goes for the counterfactual question whether the word would have applied in certain mundane situations that are never actualized. Pondering the question what I would say, were I to meet an animal looking just like my friend’s cat except for being brown rather than black, I don’t have to hesitate. (Häggqvist [137], 37)

Eine solche Erwiderung mag Manchen allerdings als unfair erscheinen, da doch zumindest Quine eine Theorie darüber zu haben scheint, weshalb für noch nicht eingetretene Umstände eine Prognose unseres Sprachverhaltens unmöglich sein sollte.

Quine hatte dafür argumentiert, dass eine Unterscheidung zwischen analytischen und synthetischen Sätzen nicht klar gezogen werden kann. Im Gegenteil bilden unsere „synthetischen“ und unsere „analytischen“ Überzeugungen ein gemeinsames Überzeugungsnetz, so dass die „Bedeutung“ von Ausdrücken nichts ist, was unabhängig von Überzeugungsänderungen an irgendeiner Stelle unseres Überzeugungsnetzes als fixiert oder unveränderbar aufgefasst werden kann.<sup>207</sup>

Kontrafaktische Situationen bringen es nun mit sich, dass in ihnen Dinge anders sind, als wir glauben, dass sie tatsächlich sind. Insbesondere werden in diesen kontrafaktischen Situationen Sätze, die in der aktuellen Welt zur Bedeutung der relevanten Wörter (‘Überleben’, ‘Tatsachenwissen’, ‘Verstehen’) möglicherweise beitragen, einen anderen Wahrheitswert haben. Folgt daraus nicht, dass unsere Wörter in kontrafaktischen Situationen ihre Bedeutung verlieren?

Dies folgt zunächst einmal nicht. Es folgt lediglich, dass sich die Bedeutung unserer Wörter *in* diesen Situationen *ändert*, und es gibt keinen Grund anzunehmen, dass dies auf unbestimmte und unvorhersehbare Weise geschieht. Falls wir in der Lage sind, genügend Hintergrundinformationen zu liefern, wie die kontrafak-

---

<sup>207</sup> Es geht hier wohlgerne nicht um bloßen Überzeugungsholismus, also die (vermutlich wahre) These, dass man bestimmte Überzeugungen nur haben kann, wenn man eine Reihe anderer Überzeugungen ebenfalls besitzt, sondern um *Bedeutungsholismus*. Vgl. Pagin [247].

tische Situation beschaffen sein soll, wird es auch eine Tatsache geben, wie wir in<sup>208</sup> einer solchen Situation bestimmte Wörter verwenden würden. Es ist ebenfalls anzunehmen, dass wir halbwegs zuverlässige Vorhersagen darüber treffen können, wie wir in diesen Situationen sprachlich reagieren würden. Dem sollte zumindest so sein, wenn eine Änderung in unserem Überzeugungssystem zu bloß *lokalem* Bedeutungswandel führt, also z.B. die Annahme, dass Teletransporter möglich sind (was möglicherweise eine Verletzung der Quantenmechanik darstellt<sup>209</sup>), *nicht* dazu führt, dass sich unser Begriff personaler Identität ändert.

Falls eine Änderung im Überzeugungssystem allerdings immer (oder nahezu immer) zu *globalem* Bedeutungswandel führen sollte, der unkontrolliert alle inferentiellen Beziehungen in Mitleidenschaft ziehen würde, gäbe es vermutlich tatsächlich keine Tatsache in Bezug darauf, was wir sagen würden. Solch eine Form von globalem Bedeutungsholismus ist allerdings *mad dog*-Holismus. Falls die Bedeutung eines Ausdrucks durch die *gesamte* inferentielle Rolle im Überzeugungssystem bestimmt wird, dann unterscheiden sich alle zwei Überzeugungssysteme, die sich in einer Überzeugung unterscheiden, in allen Überzeugungen.<sup>210</sup>

Wie beispielsweise Jerry Fodor argumentiert hat, hätte dies die unliebsame Konsequenz, dass in der aktualen Welt keine zwei Personen jemals eine Überzeugung teilen können, keine zwei Personen jemals dasselbe meinen mit dem, was sie sagen, dass keine zwei Personenstadien derselben Person jemals dasselbe meinen mit dem, was sie sagen, etc. Ein Bedeutungsholismus, der die Unbestimmtheit dessen, was wir in kontrafaktischen Situationen sagen würden, plausibel und wahrscheinlich machen könnte, ist also seinerseits unplausibel und unwahrscheinlich.<sup>211</sup>

Einen näher liegenden Einwand hat Derek Parfit gegen Quines und Wittgensteins Kritik vorgetragen:

[Quine's and Wittgenstein's] criticism might be justified if, when considering such imagined cases, we had no reactions. But these cases arouse in most of us strong beliefs. (Parfit [250], 200)

<sup>208</sup> Inwiefern es tatsächlich um die Frage geht, wie wir „in“ solchen Situationen sprachlich handeln, wird uns im nächsten Abschnitt beschäftigen. Wir folgen hier zunächst den Darstellungen Wittgensteins, Quines und Fodors, demzufolge es bei Gedankenexperimenten in der Philosophie darum geht, zu prognostizieren, wie wir in kontrafaktischen Situationen sprachlich reagieren würden.

<sup>209</sup> Bartelborth [10], 206. Bartelborth meint, dass sich eine kohärentistische Theorie des Gedankenexperiments entwickeln lässt. Dieses (Bartelborths) Beispiel scheint ein gutes Gegenbeispiel gegen Bartelborths Hypothese abzugeben, dass eine kohärentistische Theorie des Gedankenexperiments sonderlich erhellend wäre.

<sup>210</sup> Diese These wird auch als „Total Change Thesis“ bezeichnet. Vgl. Pagin [247]. Erst eine solch radikale Auffassung scheint die erforderliche Instabilität zu garantieren, die die hier diskutierte Kritik am Gedankenexperiment plausibilisieren kann.

<sup>211</sup> Zu verschiedenen radikalen Formen des Holismus und ihre Bewertung, vgl. Pagin [247]. Zur Kritik am Holismus vgl. Fodor und Lepore [108], Fodor und LePore [109].

In der Regel scheinen unsere Wörter zumindest hinreichend mit „logical force“ ausgestattet zu sein um festzulegen, was wir sagen würden. In den meisten Fällen (insbesondere in denjenigen, die es in die philosophische Fachliteratur geschafft haben) können wir eher feststellen, dass unsere Prognosen konvergieren, als dass sie wild auseinander laufen. Dieser Eindruck wäre zunächst durch den Holisten wegzu erklären (wobei ich sozialen Druck (peer group pressure) und Massenhalluzination für unwahrscheinliche Erklärungen halte<sup>212</sup>).

Es ist demnach alles andere als offensichtlich, dass es keine Tatsachen gibt, die festlegen, was wir sagen würden. Es ist vielmehr wahrscheinlich, dass unsere Sprache für diese Situationen mit hinreichender „logical force“ ausgestattet ist. Wenden wir uns also (b) zu.

#### 5.2.2 OBSCHON ES TATSACHEN GIBT, DIE FESTLEGEN, WAS WIR IN BEZUG AUF EINE KONTRAFAKTISCHE SITUATION SAGEN WÜRDEN, KÖNNEN WIR SIE DURCH DIE METHODE DES GEDANKENEXPERIMENTS NICHT ERGRÜNDEN

Dieser Vorwurf ist von Jerry Fodor erhoben worden.<sup>213</sup> Fodor will nachweisen, dass die Methode des Gedankenexperiments in der Philosophie deswegen dubios ist, weil sie per definitionem nicht als zuverlässig erwiesen werden kann, und es Gründe gibt, sie a priori für unzuverlässig zu halten.

Fodor unterscheidet zunächst zwei Fälle, in welchen Philosophen eine Antwort auf die Frage suchen, was wir in einer Situation sagen würden. Einen der beiden Fälle hält Fodor für prinzipiell unproblematisch, den anderen für höchst zweifelhaft.

Der unproblematische Fall ist der Versuch, eine Sprache zu beschreiben; also diejenigen internalisierten Regeln explizit zu machen, die Sprecher erwerben, wenn sie eine Sprache lernen. Eine Möglichkeit dies zu tun, besteht darin, das verbale Verhalten von Sprechern in verschiedenen Situationen zu beobachten (in etwa so, wie Quines Feldlinguistin dies tun würde<sup>214</sup>).

Ein anderer Weg könnte darin bestehen, dass man Sprecher einfach fragt, was sie *denken*, wie sie sich in bestimmten vorgestellten Situationen sprachlich verhalten *würden*. Die so erworbenen Daten mögen im Vergleich mit der ersten Strategie (der Beobachtung in tatsächlichen Situationen) weniger zuverlässig sein, aber diese Methode ist dennoch im Großen und Ganzen unproblematisch. Fehlerhafte Urteile über das eigene Sprachverhalten kommen ans Tageslicht, wenn die Sprecher mit tatsächlichen Situationen konfrontiert werden und es sich herausstellt, dass das tatsächliche Sprachverhalten von dem prognostizierten abweicht. Die Zuverlässigkeit dieser Methode ist also empirisch evaluierbar, und es scheint plausibel

---

<sup>212</sup> Obwohl Van Inwagen [333] und Fuhrmann [112] diese Erklärungen ernsthaft in Betracht zu ziehen scheinen.

<sup>213</sup> Fodor [106].

<sup>214</sup> Quine [268], 63.

anzunehmen, dass das Sprachvermögen Sprechern eine gewisse Grundlage dafür gibt, solche Prognosen über ihr eigenes sprachliches Verhalten in neuen Situationen in zuverlässiger Weise zu tätigen. Die Plausibilität dieser Annahme ist genau, worauf sich Häggqvist in obigem Zitat bezogen hat: „Encountering a new cat, I call it a ‚cat‘.“

Der *problematische* Fall, in dem Philosophen fragen, was wir sagen würden, beruht auf einer Unterscheidung zwischen *logischen* und *empirischen Charakteristika* eines Wortes. Diese Begriffe sind bei Fodor folgendermaßen definiert<sup>215</sup>:

- (F1) Ein Merkmal F ist logisch charakteristisch für ein Wort *w*, gdw. »Wenn *w* korrekt angewendet wird, ist F instantiiert.« eine logische Wahrheit ist.
- (F2) Ein Merkmal F\* ist empirisch charakteristisch für ein Wort *w*, gdw. (i.) es ein Merkmal F gibt, das für *w* logisch charakteristisch ist; und (ii.) »Wenn F\* instanziiert ist, ist auch F instanziiert« eine wahre empirische Verallgemeinerung oder ein Naturgesetz darstellt (aber keine logische Wahrheit).

(F1) drückt die Auffassung aus, dass einige der Merkmale, die regelmäßig die korrekte Anwendung eines Wortes begleiten, zu den Anwendungskriterien des Wortes gehören, d.h., dass ihre Präsenz zu den notwendigen Bedingungen dafür gehört, dass das Wort in einer Situation korrekt verwendet wird (wie die Präsenz eines männlichen Geschwisters eine notwendige Bedingung für die korrekte Anwendung von ‚Bruder‘ ist).

(F2) hingegen weist darauf hin, dass einige der charakteristisch auftretenden Merkmale bloß symptomatisch sind – sie sind zwar mit dem Auftreten logisch charakteristischer Merkmale korreliert, aber nicht logisch korreliert (wie die Präsenz eines männlichen Geschwisters empirisch korreliert ist mit der Präsenz eines Lebewesens mit Nieren). Die Frage ist nun, wie man zwischen logisch und den bloß empirisch charakteristischen (aber absolut zuverlässigen) Merkmalen eines Wortes unterscheiden soll. Offensichtlich können wir nicht das Verfahren des Quineschen Sprachforschers anwenden. So wie wir ‚empirisch charakteristisch‘ definiert haben, gibt es keine Situationen, in die ein Sprecher auf natürliche Weise gelangen könnte, so dass in diesen nur die empirisch charakteristischen Merkmale abwesend wären.

Philosophen, die die empirisch charakteristischen von den logisch charakteristischen Merkmalen eines Wortes trennen wollen, lösen dieses Problem, indem sie fragen, ob wir das fragliche Wort auch dann noch zur Beschreibung einer Situation verwenden würden, wenn gewisse Merkmale, die in der aktuellen Welt zuverlässig die korrekte Anwendung des Wortes begleiten, fehlen würden. Falls das, was wir zu sagen behaupten, nicht davon abweicht, was wir in Situationen *mit* diesen Merkmalen sagen, ist gezeigt, dass die weggelassenen Merkmale bloß empirisch charakteristische Merkmale des Wortes waren.

---

<sup>215</sup> Der genaue Wortlaut, bei dem ontologisch einiges durcheinander geht, findet sich in Fodor [106].

Fodor glaubt, dass man die Zuverlässigkeit dieser Methode mit einem a priori Argument in Zweifel ziehen könnte. Sein Argument ist Folgendes:

- (1) Die Methode, Sprecher nach ihrem Sprachverhalten in kontrafaktischen Situationen zu befragen, kann nicht durch eine unabhängige empirische Untersuchung gerechtfertigt werden.
  - (2) Wir haben (daher) keine Gründe, dieser Methode zu vertrauen.
  - (3) Es gibt a priori Gründe, dieser Methode zu misstrauen.
  - (4) Wenn es keine Gründe für, aber Gründe gegen eine Methode gibt, sollte man sie aufgeben.
- 

- (5) Folglich sollte diese Methode aufgegeben werden.

(1) hält Fodor für offensichtlich, da die intuitiven Behauptungen der Sprecher über ihr Sprachverhalten in kontrafaktischen Situationen in der aktuellen Welt schließlich nicht getestet werden können. Es ist ja *per definitionem* so, dass in der aktuellen Welt Situationen nicht auftreten können, in denen ein Wort korrekt verwendet wird, die zuverlässig korrelierten empirischen Merkmale aber ausbleiben.

Obwohl man Sprecher nicht in unmögliche Situationen bringen kann, um die Zuverlässigkeit ihrer Prognosen zu testen, kann man sich leicht andere Testverfahren ausdenken, die dies könnten. Zunächst könnte man auf Ergebnisse der Neurowissenschaften hoffen. Falls man jemals zu einer neurowissenschaftlichen Theorie über Sprachkompetenz gelangt, die in der Lage ist, anzugeben, welche Mechanismen in unseren Gehirnen unser sprachliches Wissen repräsentieren, bräuchte man „nur“ überprüfen, ob diese Mechanismen auf zuverlässige Weise zum Einsatz kommen, wenn wir Prognosen über unser Sprachverhalten in kontrafaktischen Situationen abgeben.<sup>216</sup>

Andere Testverfahren liegen aber näher. So scheint es überhaupt nicht erforderlich, Sprecher in unmögliche Situationen zu bringen, um ihre prognostische Zuverlässigkeit durch Beobachtung zu testen. Es reicht völlig aus, wenn die Sprecher *denken*, sie befänden sich in solchen Situationen. Dann könnte man tatsächlich

---

<sup>216</sup> Die (zugegebenermaßen bisher bescheidenen) Ergebnisse der Simulationstheorie, die wir in Kapitel 3 diskutiert haben, scheinen dafür zu sprechen, dass dem so sein könnte. Wenn wir Erzählungen mental simulieren, könnten unsere sprachlichen Reaktionen auf solche Erzählungen auf analogen Prozessen und Mechanismen beruhen wie unser sprachliches Verhalten in aktuellen Situationen und deshalb ein zuverlässiger Indikator sein. Dies sind aber empirische Spekulationen. Vgl. zu diesen Überlegungen auch verwandte Überlegungen bei Carnap [53], 233-247. Eine neurowissenschaftliche Lösung des Problems entspräche Carnaps Vorstellungen zum Begriff der Intension bei einem Roboter. Falls wir auf der Basis neurowissenschaftlicher Untersuchungen in die Situation kämen, die Carnap für den Roboter beschreibt (dass wir über eine Blaupause seiner Konstruktion verfügen und seine sprachlichen Dispositionen daraus ermitteln können), könnte man sehr wohl unabhängige Evidenz für oder gegen die Prognosefähigkeit von Sprechern gewinnen. Vgl. insbesondere Carnap [53], 244-247.

durch Beobachtung feststellen, ob ihr tatsächliches Sprachverhalten dem Prognostizierten entspricht. Solche Situationen können hergestellt werden, ohne dass man die Naturgesetze ändern müsste – man braucht die Probanden nur darüber zu täuschen, in was für Situationen sie sich befinden (Magier tun dies jeden Tag).

Was die Ergebnisse solcher Tests sein werden ist reine Spekulation. Fodors Behauptung, (1) sei a priori, ist aber klarerweise falsch. Falls die fragliche Methode nicht empirisch gerechtfertigt werden kann, ist dies a posteriori so. Dies vermindert die Überzeugungskraft von Fodors Argument.

Dennoch ist das Argument damit keineswegs ausgeräumt. Falls wir keine positive empirische Rechtfertigung vorlegen können, kann Fodors Argumentation immer noch beweiskräftig sein, da (2) immer noch wahr sein kann. (2) ist *für sich genommen* kein Grund eine Methode aufzugeben – dafür haben wir oben bereits argumentiert. Falls Fodor Recht hat, und die Methode, Prognosen zu kontrafaktischen Fällen abzufragen, die einzige Methode ist, logisch und empirisch charakteristische Merkmale voneinander zu trennen (und wir aus irgendeinem Grund Interesse an einer solchen Trennung haben), wäre es außerdem *unverantwortlich*, diese Methode ohne Gründe aufzugeben. Die interessante Frage ist also, warum Fodor glaubt, dass der Methode a priori misstraut werden sollte.

Laut Fodor ist diese Methode *a priori* dubios, weil es keinen guten Grund gibt zu glauben, dass unsere sprachlichen Intuitionen uns bei der Vorhersage, was wir sagen würden, helfen könnten, falls unsere gegenwärtigen Überzeugungen sich in diesen Situationen als gravierend falsch herausstellen sollten:

[T]o ask what we would say should certain of our current beliefs prove false involves asking what new beliefs we would then adopt. But to answer this question we would now have to be able to predict what theories would be devised were our current theories to prove untenable [...] Since there is no general way to determine [on the basis of our linguistic intuitions] how many of our beliefs may need to be altered as the result of such discovery [...] there can be no general way to determine how much of our talking such a discovery may require us to revise. (Fodor [106], 207-208)

In gewisser Weise ist dies eine bloße Wiederholung des Quine/Wittgenstein-Einwandes den wir im vorhergehenden Abschnitt zurückgewiesen haben. Wir können diesen Einwand allerdings zum Anlass nehmen, genauer zu klären, worum es bei philosophischen Gedankenexperimenten eigentlich geht.

Fragen danach, was wir in bestimmten Fällen sagen würden, kann man auf mehr Weisen stellen, als die, die von Fodor auseinander gehalten werden. Man betrachte die Frage, was wir sagen würden, wenn wir alle Quasi-Erinnerungen<sup>217</sup> an die Vergangenheit unserer Eltern besäßen, die sich phänomenologisch in nichts von den Erinnerungen an unsere eigene Vergangenheit unterscheiden würden. Eine Möglichkeit diese Frage zu stellen, ist zu fragen, ob wir auch dann noch denselben Begriff vom Selbst hätten, *wären wir die Bewohner einer solchen möglichen Welt*. Dies wäre ähnlich der Frage, ob in einer Welt, in der amöbenhafte Teilung

<sup>217</sup> Zum Begriff der ‚Quasi-Erinnerung‘ vgl. Noonan [237], 169-191.

und Vereinigung alltägliche Phänomene sind, sich überhaupt ein Personenbegriff entwickelt hätte, ob wir in so einer Welt denselben Begriff von einem Leben hätten, etc.<sup>218</sup>

Diese Fragen sind in der Tat schwer intuitiv zu beantworten, da sie eigentlich eine komplexe sprachwissenschaftliche und psychologische Theorie erforderlich machen, die erklären und prognostizieren kann, wie sich Begriffe in einer Sprachgemeinschaft entwickeln, die eine bestimmte Nische bevölkert. Fodor hat völlig Recht, dass es unverantwortlich scheint, solche Fragen auf der Grundlage sprachlicher Intuitionen allein zu beantworten. *Unsere* sprachlichen Intuitionen machen uns vielleicht zu Experten in Hinblick auf *unsere* Wörter, aber sicherlich nicht notwendigerweise auch zu Experten für die Bedeutung von Wörtern, die von Sprachgemeinschaften verwendet werden, die andere mögliche Welten bevölkern. Wenn wir uns solche Fragen vorlegen, dann betrachten wir mögliche Welten als Welten, die mindestens auch die deskriptiven Sinne und Bedeutungen unserer Ausdrücke geprägt haben, als Welten, deren Beschaffenheit einen kausalen Einfluss auf die Entwicklung der in dieser Welt verwendeten Begriffe hatte.<sup>219</sup> Obwohl eine solche Betrachtungsweise interessant und erhellend sein mag, hat es nichts damit zu tun, woran Philosophen interessiert sind, wenn sie empirische und logische Charakteristika von Wörtern voneinander trennen wollen.<sup>220</sup>

Philosophen wollen die Bedeutung unserer Ausdrücke und den Gehalt *unserer* Begriffe erforschen, so wie diese in der *aktualen* Welt geformt wurden und gebraucht werden. *Deshalb* fragen wir uns, wie unsere Begriffe auf kontrafaktische Situationen angewendet würden. Wir fragen dann danach, ob wir ‚Selbst‘ auch dann auf das „bewusste, denkende Ding, das Schmerz und Freude empfinden kann und soweit wie sein Bewusstsein reicht“<sup>221</sup> anwenden würden, wenn wir in einer Welt mit Quasi-Erinnerungen zu Besuch wären, oder ob es „das Überleben von Personen“ ist, das wir in einer Parfit-Welt beobachten würden. Es ist plausibel anzunehmen, dass wir solche Fragen auf der Grundlage unserer Sprachkompetenz beantworten können, weil die Begriffe, die wir anzuwenden versuchen, *unsere eigenen* sind.<sup>222</sup>

Fodor selbst scheint dies fast zuzugeben:

It may still be claimed that the speaker's intuitions suffice to determine when a revision in our ways of talking is tantamount to a change in the meaning of some word. I do not deny this is so, but I deny that it is a claim we are entitled to without argument. (Fodor [106], 201)

---

<sup>218</sup> Vgl. hierzu Parfit [250], 303.

<sup>219</sup> Was „deskriptive Sinne“ sind und ob unsere Ausdrücke solche haben, wird in Kapitel 7.2 problematisiert. Bis auf weiteres kann man sie einfach als Fregesche Sinne betrachten.

<sup>220</sup> Außer von Fodor, Wittgenstein und Quine wird dieser Irrtum auch von Wilkes [345] und Gale [115] wiederholt, wie Kannuck [161] aufgewiesen hat.

<sup>221</sup> Locke [188], 39.

<sup>222</sup> Wohlgemerkt setzt dies nicht die Fixierung der Referenz rigider Designatoren voraus. Letzteres kann nicht auf der Grundlage der Sprachkompetenz alleine geleistet werden, wie wir noch sehen werden.

Seine Skepsis beruht dann darauf, dass – gegeben, dass wir keine unabhängig bestätigte Theorie über Bedeutungswandel haben – wir nicht sicher sein können, dass wir immer erfolgreich eine kontrafaktische mögliche Welt beschreiben, ohne dabei die deskriptiven Sinne und Bedeutungen unserer Ausdrücke zu ändern. Fodor befürchtet, dass es sein kann, dass immer (oder häufig) wenn wir eine mögliche Welt betrachten, die in bestimmter Hinsicht von der aktuellen Welt abweicht, unsere Begriffe sich auf unkontrollierbare Weise ändern.

Es bleibt aber völlig unklar, warum dem so sein sollte. Wenn Fodor Recht hat, und die Betrachtung kontrafaktischer Fälle der einzige Weg ist, empirisch und logisch charakteristische Merkmale von Wörtern voneinander zu trennen, dann scheint die Auffassung, dass es *überhaupt* so eine Trennung gibt, Resultat eben dieser von Fodor angezweifelte Intuitionen zu sein. Aber hätten wir eine solche Unterscheidung getroffen, wenn unsere Intuitionen ein sehr unstabiles Phänomen wären? Es scheint zumindest nicht so, dass Bedeutungswandel bei diesen modalen Überlegungen häufig vorkommt. Wir sind in der Lage, die Frage, welche Sprache die Bewohner der Parfit-Welt sprechen, von der Frage zu trennen, wie unsere Begriffe auf die Parfit-Welt angewendet würden. Dies führt im Allgemeinen nicht zu Verwirrungen und Konfusionen, weshalb es durchaus sinnvoll erscheint, zu glauben, dass wir zuverlässig Auskunft darüber geben können, wie wir eine solche Welt beschreiben würden. Außerdem sind wir (offenbar) an dieser Unterscheidung zwischen empirischen und logischen Charakteristika interessiert. Fodor sagt, dass wir keine alternative Methode haben, dass er sie aber nicht rechtfertigen kann. Dies ist wieder nur der allgemeinste Einwand gegen das Gedankenexperimentieren. Fodor gibt offenbar *gar kein* a priori Argument gegen *diese* Methode.

### 5.3 PHILOSOPHISCHE GEDANKENEXPERIMENTE SIND ZIRKULÄR ODER WIDERSPRÜCHLICH

In der Anfangsphase der Analytischen Philosophie, insbesondere in der Auffassung der Mitglieder des Wiener Kreises, hielt man es für eines der Hauptergebnisse der Philosophiegeschichte – das es nun abzustellen galt –, dass philosophische Probleme nie einer Lösung zugeführt werden, sondern philosophische Großkonzeptionen bloß aufeinanderprallen, ohne irgendeinen Schritt vorwärts zu machen. Moritz Schlick<sup>223</sup> war dabei recht zuversichtlich, dass dieser Missstand durch die analytische Philosophie behoben werden könne:

Ich bin nämlich überzeugt, daß wir in einer durchaus endgültigen Wendung der Philosophie mitten darin stehen und daß wir sachlich berechtigt sind, den unfruchtbaren Streit der Systeme als beendet anzusehen. Die Gegenwart ist, so behaupte ich, bereits im Besitz jener Mittel, die jeden Streit im Prinzip unnötig machen; es kommt nur darauf an, sie entschlossen anzuwenden. (Schlick [287], 5)

---

<sup>223</sup> Einen ähnlichen Optimismus verbreitet aber auch Carnap im Vorwort zum *Aufbau* Carnap [54].

Liest man moderne Kritiker der Methode des Gedankenexperiments, gewinnt man den Eindruck, dass Schlicks Optimismus durch die Geschichte widerlegt wurde. Obwohl die analytische Philosophie, also diejenige moderne Richtung der Philosophie, die sich historisch auf den Wiener Kreis (einschließlich des späten Wittgenstein) zurückführen lässt, heutzutage die akademische Philosophie dominiert, scheint es als hätten sich innerhalb dieser Richtung Lager gebildet, die ihre unterschiedlichen inhaltlichen Auffassungen bloß austauschen, ohne über Instrumente zu verfügen, die die Gegenseite überzeugen könnten. An erster Stelle taue dabei der Einsatz von Gedankenexperimenten weniger zur Problemlösung in solchen Streitfällen als zur weiteren Zementierung der jeweiligen Positionen. Dabei sei es nicht das Gedankenexperiment als solches, das dieses Problem verursacht, sondern der besondere Gebrauch, den Philosophen von ihm machen.

Die Position, die wir nun diskutieren wollen, ist spezifischer als die Positionen, die wir im letzten Kapitel betrachtet haben. Gedankenexperimente in der Philosophie können nicht überzeugen, weil sie entweder voraussetzen, was eigentlich zu zeigen ist, oder weil sie zu widersprüchlichen Resultaten führen. Beides scheinen nachvollziehbare Vorwürfe zu sein. Wenn sich zeigen ließe, dass philosophische Gedankenexperimente häufig oder fast immer diese Defekte aufweisen, wäre diese Methode doch schlechter als bisher angenommen. Schließlich haben wir Gedankenexperimente als Instrumente zur *rationalen* Überzeugungsänderung bestimmt. Argumente, die voraussetzen, was zu zeigen ist, sowie ungültige Argumente führen sicher nicht zu einer „rationalen“ Überzeugungsänderungen, falls sie überhaupt zu einer Überzeugungsänderung führen.

### 5.3.1 ZWEI INDIKATOREN FÜR SCHLECHTE GEDANKENEXPERIMENTE

Jeanne Peijnenburg und David Atkinson haben versucht, für mehrere der paradigmatischen Gedankenexperimente, die wir in Kapitel 4 vorgeführt haben (unter anderem *Mary, die Neurophysiologin, Searles Chinesisches Zimmer*, das Zombiargument), nachzuweisen, dass diese „widersprechende Konklusionen“ haben („conclusions contradict one another“ (Peijnenburg und Atkinson [252], 308)), oder „die Konklusion eine *petitio principii* begeht“ („conclusions beg the question“ (Peijnenburg und Atkinson [252], 310)). Beides sind nach Peijnenburgs und Atkinsons Auffassung Indikatoren<sup>224</sup> dafür, dass es sich um ein schlechtes („poor“) Gedankenexperiment handelt. Da diese Formulierungen nicht selbsterklärend sind, wollen wir sie zunächst rational rekonstruieren.

---

<sup>224</sup> Bei Atkinson und Peijnenburg sind beides nicht automatisch *Mängel* eines Gedankenexperiments, sondern eben zunächst nur *Indikatoren* für einen Mangel. Auch ein Gedankenexperiment, das beide Indikatoren aufweist (wie – nach Peijnenburg und Atkinson – das EPR-Gedankenexperiment) kann sich u.U. noch als gutes Gedankenexperiment herausstellen.

### 5.3.1.1 *Sich widersprechende Konklusionen*

Vermutlich hat es soweit kein analytischer Philosoph geschafft, ein Argument mit sich widersprechenden Konklusionen zu publizieren. Was Peijnenburg und Atkinson stattdessen meinen, ist offenbar, dass das intendierte Publikum (also die Philosophen selber) in Anbetracht der Beschreibung des Gedankenexperiments mit unterschiedlichen Intuitionen reagiert. Fragt man, ob Zombies vorstellbar sind, oder ob Mary etwas Neues gelernt hat, würde ein Teil der Philosophenwelt zustimmen, der andere sich ablehnend äußern, wobei Zustimmung und Ablehnung mutmaßlich mit den jeweiligen philosophischen Verpflichtungen korreliert sind. Ein Physikaliker, beispielsweise, würde in beiden Fällen eine ablehnende Antwort geben, ein Dualist eine zustimmende.

Die Folgen dieser Tatsache sind „disaströs“, da nicht nur prinzipiell beide Seiten mit demselben Gedankenexperiment für ihre Position argumentieren *könnten*, sondern weil beide dies auch noch *tun*:

This debate (es geht um Jacksons „Mary“) is doomed continually to go round on a merry go-round, since the participants more or less tacitly endorse two totally different starting points. [...] But of course this controversy cannot be resolved, since there is, at present, no way to know whether what Mary learns is or is not relevant for explaining consciousness. And as long as we do not know what is relevant, the same Mary-story can be happily taken in two opposite ways. (Peijnenburg und Atkinson [252], 310)

Ein Indiz für ein schlechtes Gedankenexperiment ist es also, wenn das Gedankenexperiment keine einheitlichen Intuitionen im intendierten Publikum auslöst, sondern zu unterschiedlichen Intuitionen führt, die sich genau in dem Punkt widersprechen, der das Thema des Gedankenexperiments ist (nämlich die Beurteilung der Target-These).

### 5.3.1.2 *Konklusionen begehen eine petitio principii*

Peijnenburgs und Atkinsons Redeweise von „question begging conclusions“ ist ebenfalls nicht klar. Einer Konklusion kann schließlich nicht in gutem Sinne der Vorwurf gemacht werden, *sie* würde etwas behaupten, was zu zeigen wäre, weil man von Konklusionen eines Arguments ja gerade *erwartet*, dass sie das behaupten, was zu zeigen war. Was gemeint ist, ist, dass in die Beschreibung des Gedankenexperiments bereits *Annahmen* eingehen, die erst zu zeigen sind:

[The second indication for poor thought experiments is] that the conclusions drawn from thought experiments beg the question: they hinge on intuitions of which the truth or falsity was supposed to be demonstrated by those very thought experiments. In other words, not only are the conclusions contradictory, they also include the intuitions for the sake of whose elucidation the thought experiment was constructed. (Peijnenburg und Atkinson [252], 310)

Für beide Indikatoren ist es schwer, unkontroverse Beispiele zu finden.<sup>225</sup> Ein Beispiel für beide Indikatoren ist wohl das „König-von-China“-Gedankenexperiment von Leibniz bzw. Swinburne (wobei es sich etwa im Sinne von  $I_3$  um „dasselbe“ Gedankenexperiment handelt). Hier ist die Version von Leibniz:

Nehmen wir an, dass irgendein Individuum plötzlich König von China werden sollte, unter der Bedingung jedoch, das zu vergessen, was es gewesen ist, so als ob es ganz von neuem geboren worden wäre – ist das nicht in der Praxis oder hinsichtlich der Wirkungen, die man wahrnehmen kann, genau dasselbe, als ob es vernichtet werden sollte und ein König von China sollte an seiner Stelle im gleichen Augenblick geschaffen werden? Dieses Individuum hat aber keinen Grund, dies zu wünschen. (Leibniz [183], 145-157)

Leibniz suggeriert hier, dass es ein gehaltloser bzw. inkohärenter Wunsch wäre, sich an die Stelle des Königs von China zu wünschen, falls dies den Verlust der eigenen Erinnerung bedeuten würde. Dieselbe Geschichte führte Richard Swinburne zu einer anderen Lehre:

Many people wish they were somebody else in the sense that they wish they were in his shoes with his body, position, relationships, appearance, memory and character. Perhaps my body is withered, my own position and relationships are unsatisfactory, my looks are ugly, my memories give me no joy, and I am profoundly dissatisfied with my own character. You on the other hand seem very satisfactory in these ways. So I wish I were you (in the above sense). Is the wish coherent? Am I, that is, wishing for the existence of a logically possible state of affairs different from the present state? Superficially, yes. (Swinburne [323], 245)

Leibniz und Swinburne scheinen bezüglich dieses Falls völlig widersprüchliche Intuitionen zu haben. Die Intuitionen, die beide dabei in Anschlag bringen, scheinen außerdem davon abzuhängen, welche Theorie personaler Identität sie vertreten. Leibniz war zum damaligen Zeitpunkt Anhänger einer Lockeanischen, also reduktionistischen Theorie personaler Identität. Swinburne andererseits ist der Überzeugung, dass personale Identität etwas Unanalysierbares und Primitives ist. Zu dieser Position gelangte Leibniz in seinen späteren Schriften ebenfalls, hauptsächlich aus theoretischen Gründen.<sup>226</sup> Daraufhin wandelten sich auch seine Intuitionen bezüglich des Gedankenexperiments:

Und wenn ich alle vergangenen Dinge vergessen hätte und gezwungen wäre, mich von neuem bis auf meinen Namen und bis aufs Lesen und Schreiben belehren zu lassen, so könnte ich immer von anderen mein vergangenes Leben in meinem vorhergehenden Zustand erfahren, wie ich meine Rechte bewahrt habe, ohne dass es notwendig wäre, mich in zwei Personen zu teilen und mich zum Erben meiner selbst zu machen. (Leibniz [182])

---

<sup>225</sup> Wie wir in Kapitel 7.3 argumentieren werden, weist Williams Gedankenexperiment zur personalen Identität vermutlich den zweiten Indikator ebenfalls auf.

<sup>226</sup> Vgl. Noonan [237], 57-64.

Falls die Intuitionen, die durch Gedankenexperimente abgefragt werden, im Großen und Ganzen auf diese Weise von den Theorien abhängen würden, die wir sozusagen „professionell“ vertreten<sup>227</sup>, könnten sie keine kritische Funktion erfüllen: sie würden unter diesen Umständen nicht zu rationalen Überzeugungsänderungen führen.<sup>228</sup>

### 5.3.2 SCHLECHTE GEDANKENEXPERIMENTE IN DER PHYSIK

Im Gegensatz zu vielen Kritikern des Gedankenexperiments, die wir in diesem Kapitel kennen gelernt haben, sind Peijnenburg und Atkinson der Meinung, dass schlechte Gedankenexperimente auch in der Naturwissenschaft häufig vorkommen, dass also auch naturwissenschaftliche Gedankenexperimente häufig die beiden soeben aufgeführten Indikatoren für schlechte Gedankenexperimente aufweisen. Allerdings sind Peijnenburg und Atkinson der Meinung, dass dies bei naturwissenschaftlichen Gedankenexperimenten weit weniger problematisch ist als bei philosophischen Gedankenexperimenten, da in den Naturwissenschaften andauernde Streitigkeiten früher oder später durch die empirische Überprüfung des Gedankenexperiments aufgelöst würden.

Zum Beleg ihrer Thesen diskutieren Peijnenburg und Atkinson zwei naturwissenschaftliche Gedankenexperimente, die beide auch in dieser Arbeit schon vorgekommen sind, Newtons Eimer-Experiment (bzw. Einsteins Sphären-Beispiel) und EPR.

In ‚The Foundation of General Theory of Relativity‘ (Einstein [99]) habe Einstein ein Gedankenexperiment vorgebracht, das belegen soll, dass beobachtbare Effekte durch beobachtbare Fakten zu erklären sind und nicht durch Rekurs auf einen absoluten Raum. Newtons Gedankenexperiment hingegen sei eine Beschreibung genau desselben Falls gewesen, nur damals in der Absicht, zu belegen, dass es einen absoluten Raum gibt (also das genaue Gegenteil). Einstein und Newton betrachten also dasselbe Gedankenexperiment, kommen dabei aber zu unterschiedlichen Ergebnissen. Dieses Beispiel scheint sowohl den ersten, wie den zweiten Indikator aufzuweisen. Newton und Einstein haben in Bezug auf denselben Fall unterschiedliche Intuitionen und beide Intuitionen sind Folgen ihrer explizit vertretenen Theorie, zu deren kritischer Diskussion das Gedankenexperiment aber gerade vorgetragen wurde. Glücklicherweise weiß man, welcher der beiden Versionen man zugeneigt sein sollte: Einsteins Theorie ist in vielen Hin-

---

<sup>227</sup> Es wäre allerdings – wie wir in 7.3 noch sehen werden – unproblematisch, wenn unsere Intuitionen uns darüber Auskunft geben würden, welche Theorie wir implizit vertreten. Die Überzeugungsänderung, die (in der Diskussion in 7.3) durch ein Gedankenexperiment initiiert werden soll, ist eine Überzeugungsänderung in Bezug darauf, welche explizite Theorie unsere impliziten Intuitionen am Besten rekonstruiert. Die Trennung zwischen beidem ist vermutlich nicht scharf, wenn es darum geht ein reflektiertes Gleichgewicht zwischen Intuitionen und expliziten Rekonstruktionen herzustellen.

<sup>228</sup> Ähnliche Vorwürfe finden sich bei Aydede und Güzeldere [8] und Güzeldere [133].

sichten (empirische Adäquatheit, Erklärungsstärke) viel besser als Newtons. Wenn wir also die Theorien betrachten, die zu den unterschiedlichen Intuitionen führen, wissen wir, welchen Intuitionen wir trauen sollten. In der Philosophie hingegen haben wir keinen Grund, die Intuitionen der einen Lobby den Intuitionen einer anderen Lobby vorzuziehen:

In philosophy, however, the turn to theories is of little help. How should we decide between, say, the theories of Searle and Dennett on understanding, meaning and consciousness? It looks as though, at the moment, we have no more than thought experiments here, and these thought experiments leave much to be desired. (Peijnenburg und Atkinson [252], 315)

Das zweite Beispiel, das von Peijnenburg und Atkinson betrachtet wird, ist das EPR-Gedankenexperiment. EPR versuchten zu zeigen, dass die Quantenmechanik unvollständig ist, da es „hidden variables“ geben muss, die in der quantenmechanischen Beschreibung nicht vorkommen. Wieder soll es so sein, dass kontradiktorische Folgerungen aus ein und demselben Gedankenexperiment gezogen wurden. Während EPR schlossen, dass die Quantenmechanik unvollständig sei, da sie Vollständigkeit daran banden, welche Messungen gemacht werden *könnten*, schloss die Gegenfraktion, dass die Quantenmechanik vollständig sei, weil Vollständigkeit davon abhängt, welche Messungen *tatsächlich* gemacht werden:

We have a thought experiment with contradictory conclusions: quantum mechanics is complete versus quantum mechanics is not complete, or something exists when you have in fact measured it versus something exists when you can infer it in principle. Moreover, the conclusions beg the question, for they are embodiments of those intuitions for the sake of which the entire thought experiment was conceived. What was at stake at the beginning of the debate was precisely the question what is or is not an element of physical reality, and it is inappropriate to present those initial intuitions as final conclusions. (Peijnenburg und Atkinson [252], 317)

Obwohl auch dieses Gedankenexperiment beide Indikatoren für ein schlechtes Gedankenexperiment erfüllt, stellte es sich laut Peijnenburg und Atkinson später insofern als ein gutes Gedankenexperiment heraus, als auf der Grundlage des empirischen Test des Gedankenexperiments durch Alain Aspect in den 80er Jahren des letzten Jahrhunderts dieser Streit entschieden werden konnte.<sup>229</sup> Obwohl die Ursprungssituation keine Entscheidung zugunsten der einen oder anderen Partei zuließ, konnte man später ein tatsächliches Experiment zur Klärung der Frage durchführen. Auch dies sei in der Philosophie so nicht möglich:

The EPR-experiment has thus been given a testable format, but it is unclear how we ever could put the Chinese Room or the Mary experiment to the test. To be sure, both the Chinese Room and the Mary experiment can be carried out, ethical

---

<sup>229</sup> Aspects eigene Auffassung von seinen Experimenten, ausgedrückt durch Titel wie ‚Fifty Years Later: When Gedanken Experiments Become Real Experiments‘ (Aspect und Grangier [5]) scheint mit der Interpretation von Peijnenburg und Atkinson konform zu sein.

considerations aside, but that would not resolve the philosophical conundrum. (Peijnenburg und Atkinson [252], 317)

Peijnenburg und Atkinson vertreten zusammengefasst also die folgenden Positionen:

- (PA1) Es gibt viele philosophische Gedankenexperimente, die beim intendierten Publikum divergierende Intuitionen auslösen, wobei beide Seiten eines philosophischen Disputs sich derselben Gedankenexperimente für die jeweils eigene Position bedienen.
- (PA2) Es gibt viele philosophische Gedankenexperimente, die eine *petitio principii* begehen.
- (PA3) Beides trifft auch auf manche Beispiele aus der Physik zu.
- (PA4) In der Philosophie können wir zur Lösung eines solchen Streits über ein Gedankenexperiment nicht auf die Theorien zurückgreifen, was in den Naturwissenschaften möglich ist.
- (PA5) In der Philosophie können wir zur Lösung eines Theorienkonflikts nicht auf ein *experimentum crucis* zurückgreifen, was in den Naturwissenschaften manchmal möglich ist, wenn wir dort das Gedankenexperiment real durchführen können.

### 5.3.3 SIND GEDANKENEXPERIMENTE IN DER PHYSIK SO SCHLECHT?

Zunächst einige wenige<sup>230</sup> Bemerkungen zu (PA<sub>3</sub>) und (PA<sub>5</sub>): Peijnenburg und Atkinson<sub>3</sub> behauptet, dass im Newton-Fall und im EPR-Fall unterschiedliche Intuitionen bezüglich des Ausgangs eines auf bestimmte Weise beschriebenen Experiments bei den jeweiligen, am fraglichen Disput beteiligten Parteien ausgelöst wurden. (PA<sub>5</sub>) behauptet dann, dass diese missliche Situation (zumindest beim EPR-Fall) dadurch aufgelöst werden konnte, dass man das Gedankenexperiment als reales Experiment durchführte.

Wie wir schon bei unserer Rekonstruktion des Eimer-Experiments und Machs Replik auf dieses Experiment gesehen haben, handelt es sich *erstens* überhaupt nicht um ein Gedankenexperiment (Newton behauptet, dass das Experiment tatsächlich durchgeführt wurde) und *zweitens* besteht der Streit nicht bezüglich der Frage, wie sich das Wasser im Eimer verhalten würde, sondern bezüglich der Frage, was als beste Erklärung für das so beschriebene Ereignis gilt. Bei Einstein ist das genauso. Auch Einstein diskutiert einen Fall, für den die Annahme eines absoluten Raumes, wie auch die Allgemeine Relativitätstheorie empirisch äquivalente Prog-

---

<sup>230</sup> Für eine ausführlichere Diskussion vgl. Cohnitz [73].

nosen machen, letztere aber epistemisch befriedigender ist (weil sie keine unbeobachtbaren Entitäten als Kausalkräfte auszeichnet). In Bezug darauf, was in Newtons Eimer passiert und was mit Einsteins Sphären geschieht, sind sich die beiden Theorien einig. (Sie sind sich uneinig darüber, was geschehen würde, gäbe es sonst keine Massen im ganzen Universum, dieser Fall wird aber weder von Newton, noch von Mach oder Einstein mit einem Gedankenexperiment diskutiert).

Bei EPR sieht die Sache ganz ähnlich aus. Auch hier ist der Experimentausgang überhaupt nicht strittig. Die Bohr-Interpretation und die EPR-Interpretation stimmen völlig darüber überein, welche Messungen im Experiment vorgenommen werden und was die jeweiligen Resultate sind. Wie auch im Fall von Newtons Eimer ist dieses Urteil nicht intuitiv, sondern wird in diesem Fall aus der Theorie (orthodoxe Quantenmechanik) abgeleitet. Was wiederum Anlass zum Disput gibt, ist die Frage, ob die Quantenmechanik als vollständige Erklärung eines so beschriebenen Phänomens zählen sollte, oder ob eine angenommene rivalisierende *hidden-variables* Theorie, die für diesen Fall empirisch äquivalent wäre, nicht vorzuziehen sei. Aus diesem Grund bringen EPR ihr „Realitätskriterium“ in Anschlag, das Bohr in seiner Replik dann kritisiert. PA<sub>3</sub> trifft auf beide Fälle also schlicht nicht zu.

PA<sub>3</sub> bezieht sich zunächst nur auf den EPR Fall. Wir sagten bereits, dass der eigentliche Experimentausgang gar nicht der Streitpunkt war, aber was sonst wurde denn dann in den 80er Jahren von Aspect so erfolgreich getestet? Was getestet wurde, war die Annahme, die in der EPR-Argumentation eine entscheidende Rolle spielt, nämlich, dass eine empirisch äquivalente *hidden-variables* Theorie möglich ist. Aspect gelang es, eine gewisse Familie<sup>231</sup> solcher Theorien als Möglichkeiten auszuschließen, weshalb diese Hintergrundannahme der Argumentation von EPR (in dieser Beziehung) in sich zusammenbrach. Diese Annahme war zum Zeitpunkt des EPR Gedankenexperiments gar nicht strittig (jedenfalls bezieht sich Bohr nicht darauf), es war damals auch nicht klar, wie bzw. dass diese Annahme empirisch testbar sein könnte. Dies ist insbesondere ein Verdienst von John Bell, 30 Jahre nach dem Gedankenexperiment.<sup>232</sup> PA<sub>3</sub> trifft auf diesen Fall also ebenfalls nicht zu.

#### 5.3.4 SIND GEDANKENEXPERIMENTE IN DER PHILOSOPHIE SO SCHLECHT?

Ob Peijnenburg und Atkinson die Naturwissenschaften und ihre Geschichte richtig darstellen ist aber auch nicht hauptsächlich unser Thema. Immerhin sind Peijnenburg und Atkinson sehr zurückhaltend, wenn es darum geht, naturwissenschaftliche Methoden zurückzuweisen. Wenn es zur Philosophie kommt, sind sie da schon weniger zurückhaltend. Aber stimmt denn ihre Charakterisierung der Diskussionssituation in der Philosophie?

<sup>231</sup> ... die so genannten „lokalen“ *hidden-variables* Theorien.

<sup>232</sup> Vgl. Clauser, Horne, Shimony und Holt [66], 880.

Erinnern wir uns an *Mary* (aus Kapitel 4.1): Mary ist eine junge Frau, die in einem schwarz-weißen Raum aufwuchs und nie eine andere Farbe zu Gesicht bekam. In ihrer Zeit im Gefängnis lernte sie alles, was sich aus neurowissenschaftlicher Sicht zum Farbsehen sagen lässt (da angenommen wird, dass Mary in einer Zukunft lebt, in der die Neurowissenschaft in diesem Sinne abgeschlossen ist). Mary weiß, was in ihren Mitmenschen passiert, wenn diese eine Farbe wahrnehmen, insbesondere weiß sie auch, in welchem Gehirnzustand man sich befindet, wenn man die Farbe Rot wahrnimmt. Nun darf sie zum ersten Mal ihr Gefängnis verlassen und ihr Blick fällt auf einen typischerweise roten Gegenstand. Mary bildet nun einen Gedanken, den sie vorher noch nie hatte: ‚So ist es also, wie es ist, Rot zu sehen.‘

Das intuitive Urteil, das hierbei angesprochen werden soll, ist das Urteil, dass es sich bei diesem Gedanken von Mary um neues Wissen handelt – sie hat etwas gelernt, was sie aus den Lehrbüchern nicht hätte lernen können. Obwohl eine Expertin für Farbwahrnehmung, gab es dennoch etwas über Farbwahrnehmung, was ihrer Kenntnis bisher entgangen sein musste und was durch ihren neuen Gedanken ausgedrückt wird. Weiterhin handelt es sich um Wissen von einer Tatsache, die auch schon vorher bestanden hat, die also nicht erst durch Marys Wahrnehmungsakt entstanden ist. Wie aber kann dann die Auffassung des Physikalisten wahr sein, dass alle Tatsachen bezüglich Wahrnehmung letztlich physikalische Tatsachen sind, wenn es Tatsachen bezüglich Wahrnehmung gibt, die jemandem, der perfekt über alle physikalischen Tatsachen informiert ist, entgangen sein können? Der Physikalismus muss also falsch sein.

Letzteres ist eine Folgerung aus dem Ergebnis des Gedankenexperiments und Annahmen darüber, was für Phänomene eigentlich in den Erklärungsbereich des Physikalismus fallen. Die intuitive Beurteilung des Gedankenexperiments beschränkt sich auf die Frage, ob man Marys Überzeugung ‚So ist es also, wie es ist, Rot zu sehen.‘ als „neues Wissen“ beschreiben soll oder nicht. Die Intuition, die hier zur Debatte steht, bezieht sich *nicht* darauf, ob der Physikalismus falsch ist. Dass er falsch sei, folgt überhaupt erst, wenn man zusätzliche weitere Annahmen macht.

In der Tat gibt es Philosophen, die angezweifelt haben, dass das Gedankenexperiment unproblematisch etablieren kann, dass Mary hier neues Wissen erwirbt. Schließlich sei unklar, was es heißen soll, dass die Neurowissenschaft, in der Mary ausgebildet wird, „vollständig“ ist. Entweder, ‚vollständig‘ heißt soviel wie ‚alles, was man über Farbwahrnehmung wissen kann‘, dann *kann* Mary ja *per definitionem* nichts Neues gelernt haben. Oder es heißt soviel wie ‚alles, was man über Farbwahrnehmung wissen kann, bis auf das, was Mary nun neu hinzulernt‘ (oder: ‚... bis auf den qualitativen Aspekt phänomenaler Erfahrungen‘), dann würde in die Beschreibung des Gedankenexperiments aber eine Annahme eingehen, die erst zu zeigen wäre, es handelte sich also um eine *petitio*. Hierbei werden keine anders gelagerten Intuitionen herbeizitiert, sondern die Kritik bezieht sich darauf, ob das Gedankenexperiment durch seine Beschreibung nicht an Aussagekraft verliert (al-

so wenn diese Kritik zutrifft, dann handelt es sich um den zweiten Indikator, nicht um den ersten).<sup>233</sup>

Andere Philosophen haben dieses Gedankenexperiment kritisiert, indem sie darauf hinwiesen, dass der Übergang von ‚Mary hat neues Wissen erworben.‘ auf ‚Es gibt eine weitere Tatsache über Farbwahrnehmung, die in der neurowissenschaftlichen Beschreibung nicht enthalten war.‘ unzulässig ist. Diese Philosophen analysieren Marys neues Wissen als *knowing how*, anstatt als *knowing that*, also als eine Form nicht-propositionalen Wissens, dem keine eigene Tatsache entsprechen muss.<sup>234</sup> Oder sie analysieren Marys neues Wissen als Wissen von einem neuen Aspekt bezüglich einer Tatsache, die ihr unter anderen Hinsichten bereits bekannt war.<sup>235</sup> Auch diese Strategien warten nicht mit anderen Intuitionen auf. Die Intuition, dass Mary etwas Neues gelernt hat, wird auch von diesen Philosophen geteilt; was bezweifelt wird, ist die Frage, ob es sich bei diesem neuen Wissen um Wissen von einer neuen Tatsache handelt, es geht also darum, welche Folgerungen man aus der Intuition ziehen darf. Diese Debatte wird geführt, indem man verschiedene Theorien über *knowing how* und *knowing that* miteinander vergleicht, bzw. verschiedene Theorien über propositionale Einstellungen heranzieht, diese im Lichte unabhängiger Daten miteinander vergleicht, die plausiblere auf den Fall appliziert und zuseht, was dabei herauskommt. Natürlich benutzen Philosophen auch dabei wieder Gedankenexperimente und befragen wiederum ihre Intuitionen, allerdings sind es diesmal keine Intuitionen über Mary und ihr Wissen und es sind auch keine Intuitionen, die sich in dem intendierten Sinne widersprechen. Die Frage, ob es sich bei Marys neuem Wissen um *knowing how* oder *knowing that* handelt, ist von der Physikalismus-Frage unabhängig. PA<sub>1</sub> und PA<sub>4</sub> treffen auf diesen Fall schlicht nicht zu!

Was ist mit PA<sub>2</sub>? Haben wir nicht gesagt, dass manche Philosophen tatsächlich meinen, dass das Gedankenexperiment in seiner Beschreibung voraussetzt, was eigentlich zu zeigen ist? Ja, das ist so, aber das ist keineswegs in irgendeinem problematischen Sinne so. Alle an der philosophischen Diskussion Beteiligten sind sich einig darüber, dass das Gedankenexperiment schlecht wäre, wenn es in seiner Beschreibung voraussetzen würde, was eigentlich zu zeigen ist. Deshalb findet man solche Vorwürfe auch *innerhalb* der philosophischen Debatte. Wenn es in der Philosophie – wie Peijnenburg und Atkinson es darstellen – zum normalen Geschäft gehörte, dass Gedankenexperimente eine *petitio principii* begehen, gäbe es kein solches Kontrollinstrument innerhalb der Philosophie des Geistes. Gedankenexperimente wären doch nur dann ein problematisches Instrument in der Philosophie, wenn es ausschließlich den Wissenschaftstheoretikern auffiele, wenn sie eine *petitio* begehen, alle (oder die Meisten) an der Debatte Beteiligten dies aber für normal und untadelig hielten.

---

<sup>233</sup> Vgl. hierzu auch die Diskussion der Definition des Physikalismus in Jackson [158].

<sup>234</sup> Vgl. Lewis [186] und Perry [255] zu den Ansätzen von Lewis und Nemirow, die so zu argumentieren scheinen.

<sup>235</sup> Perry [255].

Manchen mag dieses Argument als Sophisterei erscheinen, ich halte es aber nicht für Sophisterei, sondern für einen relativ wichtigen Punkt. Angenommen, ein Wissenschaftstheoretiker würde die statistischen Verfahren in den Sozialwissenschaften insgesamt in Zweifel ziehen und als ‚schlecht‘ bezeichnen, weil es sein kann, und manchmal vorkommt, dass scheinbare statistische Korrelationen keine tatsächlichen statistischen Korrelationen sind, *obwohl* es innerhalb der Sozialwissenschaften bekannt ist, dass dem so sein kann und manchmal so ist und es Testverfahren gibt, die untersuchen, ob es sich um eine scheinbare oder eine tatsächliche Korrelation handelt. Ich denke, dieser Wissenschaftstheoretiker hätte nichts in der Hand, um diese Methode zu kritisieren! Und so ist es auch hier.

Die Frage, ob das Gedankenexperiment schließlich voraussetzt, was erst zu zeigen wäre, ist die Frage, ob ‚vollständig‘ in unabhängiger Weise expliziert werden kann. Dies ist zunächst eine Frage, die mit den Intuitionen, ob Mary etwas Neues lernt oder nicht, nichts zu tun hat. Es gibt keinen Anlass zu glauben, dass diese Debatte dazu verurteilt ist, sich für immer im Kreis zu drehen, wie Peijnenburg und Atkinson es befürchten.

Das Zombie-Beispiel, das Peijnenburg und Atkinson besprechen, ist sehr ähnlich gelagert. Wie wir bereits erwähnten, teilen moderne Physikalisten in der Regel nicht die Auffassung, dass der Physikalismus a priori als wahr erkannt werden kann, sondern sie vertreten einen so genannten „a posteriori Physikalismus“, demzufolge wir durch empirische Untersuchungen herausgefunden haben, dass unsere mentalen Zustände letztlich nichts anderes als bestimmte neuronale Zustände unseres Gehirns sind. Da wir dies nicht a priori wissen konnten, ist es zumindest logisch möglich, dass es auch anders hätte sein können. Es wird daher von vielen zugegeben, dass es sich bei Zombies um eine logische Möglichkeit<sup>236</sup> handelt. Die Frage ist, was aus so einer logischen Möglichkeit folgt (wir werden darauf in Kapitel 7.2 noch eingehen).

Die philosophische Debatte beschäftigt sich dabei mit der Frage, ob *Hintergrundtheorien*, die unsere Intuitionen erklären (oder wegerklären) sollen, plausibel sind. Die Debatte geht nicht in der Hauptsache darum, welche Intuitionen wer hat. Wiederum ist es so, dass die Debatte dabei von der eigentlichen Streitfrage (Ist der Physikalismus falsch?) zu einer anders gelagerten Frage übergeht (Erklärt der modale Rationalismus<sup>237</sup> auf richtige Weise, wie unser sprachliches Wissen mit der Referenz unserer Ausdrücke in Verbindung steht?).

Auch im Zombie-Fall ist es so, dass innerhalb der Philosophie des Geistes Probleme mit der Beschreibung des Gedankenexperiments gesehen wurden und diese geprüft wurden. Wiederum war es dabei so, dass nicht etwa Physikalisten das Zombie-Gedankenexperiment in Argumenten für den Physikalismus verwendeten (wie Peijnenburg und Atkinson suggerieren), sondern man findet es *problematisch*, wenn das Gedankenexperiment in seiner Beschreibung das zu Etablierende bereits voraussetzt.

---

<sup>236</sup> Vgl. Chalmers [59], Chalmers [60].

<sup>237</sup> Zum modalen Rationalismus vgl. das übernächste Kapitel, sowie Balog [9].

Die Argumente, die dabei vorgetragen wurden, sind ähnlich subtil wie im Mary-Fall. Im Zombie-Fall werden wir gebeten, uns eine mögliche Welt vorzustellen, die von der unsrigen physikalisch ununterscheidbar ist. In dieser Welt müssten wohl auch alle physikalischen Gesetzmäßigkeiten wie in unserer Welt sein. Angenommen, es gibt in unserer Welt mentale Verursachung. D.h., dass unsere qualitativen mentalen Zustände manchmal bestimmte physikalische Prozesse verursachen. Nehmen wir weiter an, dass Schmerz, den ich gestern empfand, als ich auf die heiße Herdplatte fasste, den physikalischen Prozess verursachte, dass ich meine Hand von der Platte wegzog. Sei ‚M‘ mein Schmerzzustand und ‚P‘ das physikalische Ereignis, dass meine Hand sich auf bestimmte Weise bewegte. In einem „bloßen physikalischen Duplikat“  $w_2$  unserer Welt  $w_1$ , in dem es kein Bewusstsein gibt, wird es M wohl auch nicht geben. Aber was ist dann mit P? P sollte zu  $w_2$  gehören, da es sich um einen physikalischen Teil von  $w_1$  handelt und alle physikalischen Teile von  $w_1$  in  $w_2$  sind. Wie kann aber P in  $w_2$  sein, wenn M P nicht verursacht hat? Kann P unverursacht in  $w_2$  auftauchen, oder in  $w_2$  durch etwas anderes verursacht werden? Wie soll das möglich sein, wenn in  $w_2$  alles physikalisch „beim Alten“ bleibt, also keine neuen Ursachen und keine neuen Kausalgesetze hinzukommen?

Falls diese Überlegungen aber richtig sind<sup>238</sup>, dann kann P nur dann Teil von  $w_2$  sein, wenn M in  $w_1$  nicht *die* Ursache für P war – d.h., wir müssen dafür annehmen, dass unsere Welt eine epiphänomenalistische Welt ist (oder alle mental verursachten Ereignisse überdeterminiert sind.). Falls der Physikalismus bereits impliziert, dass der Epiphänomenalismus falsch ist, wäre dieses Gedankenexperiment tatsächlich mit einer *petitio principii* belastet. Falls Physikalismus und Epiphänomenalismus vereinbar sind, ginge immerhin eine sehr starke Annahme über die kausale Struktur unserer Welt in das Gedankenexperiment mit ein.<sup>239</sup>

Bei dieser Prüfung geht es also darum, das Verhältnis zwischen Physikalismus und Epiphänomenalismus zu explizieren, was wiederum in keiner Weise voraussetzt, dass der Physikalismus falsch ist, oder der Dualismus wahr, und auch keine Intuitionen dieser Art bedarf. Auch auf diesen Fall treffen Peijnenburgs und Atkinsons Punkte einfach nicht zu.

### 5.3.5 PHILOSOPHISCHE GEDANKENEXPERIMENTE KÖNNEN NICHT EMPIRISCH ENTSCHIEDEN WERDEN

Gedankenexperimente sind – auch in der Philosophie – im Allgemeinen nicht zirkulär. Es wird von den an der Debatte Beteiligten genau darauf geachtet, ob ein Gedankenexperiment in seiner Beschreibung das zu Beweisende bereits vorwegnimmt, ob die fragliche Intuition tatsächlich allgemein geteilt wird und ob die Tatsache, dass das Gedankenexperiment diese Intuition hervorruft, auch irgend-

<sup>238</sup> Vgl. Chalmers [62], Perry [255].

<sup>239</sup> Vgl. zu dieser Debatte Perry [255], Cohnitz [71], Newen [228].

welche Aussagekraft für die Konklusion hat, die in dem Gedankenexperiment-Argument gezogen wird.

Diese Debatten werden (zum Teil) geführt, ohne dass empirische Versuche an- gestellt würden. Philosophische Gedankenexperimente – das haben Peijnenburg und Atkinson richtig beobachtet – werden normalerweise nicht empirisch reali- siert. Ob das in jedem Fall unproblematisch ist, muss noch untersucht werden.

Wie wir ebenfalls gesehen haben, drehen sich die Debatten in der Philosophie deshalb nicht im Kreis, weil Dispute über Gedankenexperimente – sofern sie eben nicht überzeugen – in der Regel keine Dispute über die Frage sind, die das Gedan- kenexperiment eigentlich beantworten sollte, sondern die Debatte in andere Berei- che der Philosophie verlagern. Hierbei handelt es sich um Problemverschiebungen einer besonderen Art, die – wie wir noch sehen werden – mit den Besonderheiten der modernen Philosophie zu tun haben.

## 6. Der modale Status philosophischer Targetthesen

Die Überzeugungen, die in der Philosophie im Sinne einer philosophischen These oder Theorie mit Argumenten kritisiert und verteidigt werden, können von sehr unterschiedlicher Art sein. Manche Überzeugungen, für die in der Philosophie argumentiert wird, haben den Status einer metaphysischen bzw. empirischen These, manche Überzeugungen eher den Status einer Bedeutungsanalyse oder den eines normativen Vorschlags, eine bestimmte Praxis auf bestimmte Weise zu regeln (sei es nun die Sprachpraxis im Fall einer Begriffsexplikation oder unser anderes soziales Handeln im Fall eines ethischen Normvorschlags). Gedankenexperimente zur Überzeugungsänderung können sich entsprechend gegen Überzeugungen verschiedenen Typs richten. Je nachdem, welcher Typ von Überzeugung von einem Gedankenexperiment kritisiert werden soll, werden sich unterschiedliche Anforderungen an ein adäquates Gedankenexperimentieren stellen. Entsprechend hängt auch die Rechtfertigung der Kritik an der Methode des Gedankenexperiments davon ab, welcher Überzeugungstyp dem Gedankenexperiment als Targetthese unterstellt wird.

Wenn dem so ist, erfordert dies eine weitere funktionale Ausdifferenzierung der Methode des Gedankenexperiments zur Überzeugungsänderung. Eine solche Ausdifferenzierung müsste darauf Rücksicht nehmen, dass die Überzeugungen, die in der Philosophie diskutiert werden, von unterschiedlichem Typ sein können und an ein adäquates Gegenbeispiel jeweils unterschiedliche Anforderungen stellen.

Welche Überzeugungstypen dabei in Frage kommen, ist der Gegenstand dieses Teils unserer Untersuchung. Da wir bereits gesehen haben, dass die Überzeugungen, die als Kandidaten für Targetthesen in Frage kommen, modalen Charakters sein müssen (es muss ich – da es beim Gedankenexperiment darum geht, einen bloß möglichen Fall als Gegenbeispiel zu verwenden, bei einer Targetthese um eine Notwendigkeitsbehauptung handeln), werden wir zunächst verschiedene Notwendigkeitsbegriffe voneinander unterscheiden und in einem zweiten Schritt untersuchen, inwiefern diese verschiedenen Notwendigkeitsbegriffe mit typischen philosophischen Thesen oder Theorien in Zusammenhang stehen.

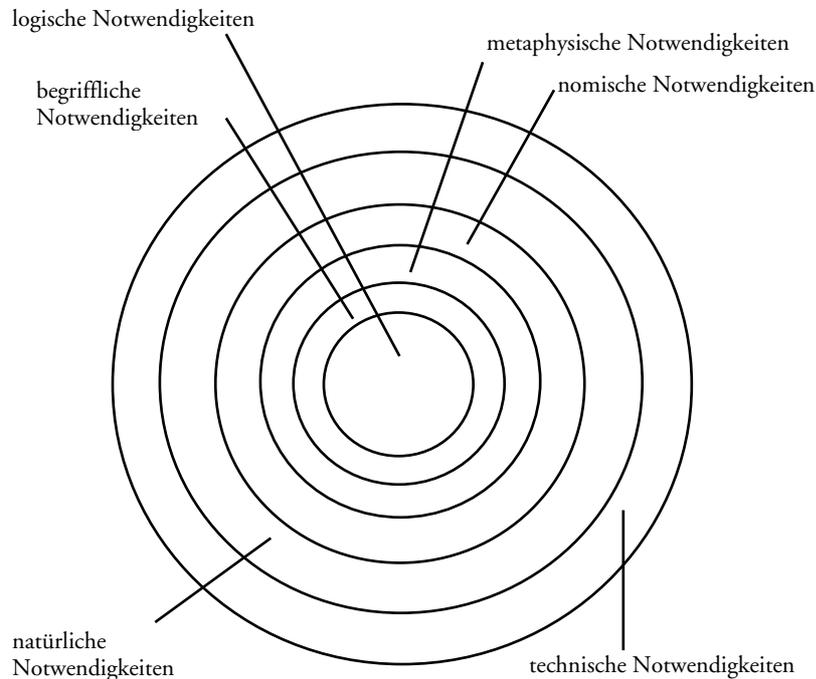


Abbildung 6.1-1

## 6.1 VERSCHIEDENE MODALITÄTEN

Wie wir sahen, war Fodor in seiner Diskussion philosophischer Gedankenexperimente davon ausgegangen, dass es bei den philosophischen Target-Thesen um Begriffsanalysen oder um Metaaussagen über Begriffsanalysen<sup>240</sup> handelt. Betrachten wir ein Beispiel aus unserer Sammlung paradigmatischer Fälle. Die These der traditionellen Epistemologie:

(TE) P weiß, dass q, gdw. P die gerechtfertigte Meinung hat, dass q.

würde nach dieser Auffassung als Begriffsanalyse verstanden. D.h., dass der Ausdruck links vom ‚gdw.‘, also das Analysandum, mit dem Ausdruck auf der rechten Seite, dem Analysans, bedeutungsgleich sein muss, wenn (TE) wahr ist. Wie wir in Kapitel 6.2 und 7.2 noch näher ausführen werden, stellt (TE) eine begriffliche Wahrheit dar, wenn es begrifflich notwendig ist, dass P weiß, dass q, gdw. P die

<sup>240</sup> So haben wir die Thesen des Physikalismus in dieser Interpretation eigentlich als solche Metaaussagen kennen gelernt, die im Grunde behaupten, dass eine Begriffsanalyse unseres mentalen Vokabulars vollständig in physikalischem Vokabular gegeben werden kann.

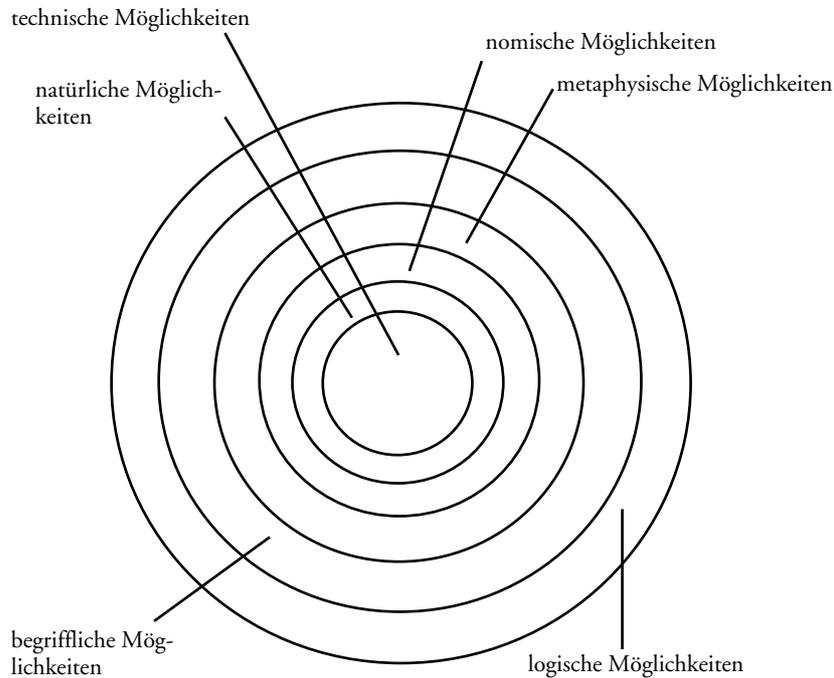


Abbildung 6.1-2

gerechtfertigte Meinung hat, dass  $q$ , d.h., dass es in allen begrifflich möglichen Welten, also in allen möglichen Welten, die keine begrifflichen Widersprüche enthalten (weil sie die Bedeutungspostulate des Deutschen erfüllen),  $P$  weiss, dass  $q$ , gdw.  $P$  die gerechtfertigte Meinung hat, dass  $q$ .

Zwischen Analysans und Analysandum muß also die besondere Beziehung der Bedeutungsgleichheit vorliegen. Eine solche Bedeutungsanalyse ist dann auch eine *analytische* Wahrheit. An Stelle dieser Verbindung von Analysans und Analysandum könnte man sich aber auch eine andere Verbindung vorstellen. In den Naturwissenschaften werden offensichtlich keine analytischen Wahrheiten untersucht, sondern Naturgesetze. Dort testen Gedankenexperimente also auch andere mögliche Welten als die bloß begrifflich möglichen Welten.

Betrachten wir zur Illustration die „Zwiebel der Notwendigkeit“, ein Modell, das wohl die meisten<sup>241</sup> Philosophen im Hinterkopf haben, wenn sie über verschiedene Notwendigkeitsbegriffe nachdenken. Laut der Zwiebel der Notwendigkeit ist die Menge der logischen<sup>242</sup> Notwendigkeiten eine echte Teilmenge der

<sup>241</sup> Die Zwiebel der Notwendigkeit ist zwar in prima facie Widerspruch zu einigen Thesen Kripkes, wir werden dies aber in diesem Kapitel nicht thematisieren. Eine kurze Diskussion erfolgt am Anfang von Kapitel 7.2.

<sup>242</sup> Abhängig davon, welche Auffassung man von Logik hat, könnte man die Menge der aussagenlogischen Notwendigkeiten als echte Teilmenge der (prädikaten-)logischen Notwendigkeiten in das Zentrum der Zwiebel setzen.

Menge der begrifflichen Notwendigkeiten, diese wiederum eine echte Teilmenge der metaphysischen Notwendigkeiten (wenn es solche gibt), diese eine Menge der nomischen Notwendigkeiten, diese eine Teilmenge der Naturnotwendigkeiten (wenn es einen Unterschied zwischen Anfangs- bzw. Randbedingungen und Naturgesetzen gibt)<sup>243</sup>, etc. Diese Ordnung ist in Abbildung 6.1-1 wiedergegeben. Ihr Pendant ist die „Zwiebel der Möglichkeiten“ (Abbildung 6.1-2). Eine solche Darstellungsweise erfordert, dass es eine Menge (primitiver) möglicher Welten gibt, die dementsprechend sortiert werden können.<sup>244</sup> Ob diese Notwendigkeitsbegriffe (bzw. die Rede von verschiedenen Mengen „möglicher Welten“) aufeinander oder auf nicht-modale Begriffe zurückgeführt werden können, ist nicht Gegenstand dieser Arbeit. Wenn hier von Möglichkeit und Notwendigkeit geredet wird, dann geht es jeweils um Weisen, wie unsere Welt sein kann, bzw. sein muss. Eine mögliche Welt ist eine vollständige Weise, wie unsere Welt sein kann. Eine – beispielsweise – *nomisch* mögliche Welt, ist – wie wir gesagt haben, eine vollständige Weise, wie unsere Welt sein kann, wenn sie sich in Bezug auf die Naturgesetze nicht von der aktualen Welt unterscheidet (vgl. Lewis [187], Hale [139]).

Leugnet man nun, dass es eine ausgezeichnete Klasse begrifflicher Wahrheiten gibt, hält ansonsten aber an der Vorstellung von Modalitäten fest<sup>245</sup>, so könnte man philosophische Thesen auch einfach als eine andere Art von Notwendigkeitsaussage uminterpretieren. Allgemein wäre eine philosophische These dann wahr, wenn sie in der relevanten Teilmenge der Menge der möglichen Welten wahr wäre, also in allen relevanten Analysandum-Welten auch das Analysans auftritt und umgekehrt. Welche Menge möglicher Welten dabei „relevant“ ist, ist dann die Frage, in welchem anderen Sinn philosophische Thesen zu interpretieren sind. Die verbreitete Auffassung, dass es sich um „metaphysische“ Notwendigkeiten handelt, wird uns dabei hauptsächlich beschäftigen (vgl. Kapitel 7.2).

Verschiebt man die Interpretation philosophischer Thesen weiter an den Rand der Zwiebel der Notwendigkeiten, scheint man das erkenntnistheoretische Problem, wieso *Gedankenexperimente* relevante Informationen enthalten sollen, dem Problem naturwissenschaftlicher Gedankenexperimente anzugleichen. Freilich

---

<sup>243</sup> Die Terminologie ist in der Fachliteratur uneinheitlich. Man vgl. etwa die abweichende Redeweise von „natürlichen“ Notwendigkeiten und „physikalischen“ Notwendigkeiten bei Brooks [37], der allerdings ebenfalls darauf hinweist, dass zwischen verschiedenen Modalitäten zunächst zu unterscheiden ist. Vgl. hierzu auch Ward [336]. Der Begriff der logischen Notwendigkeit wird manchmal synonym mit dem Begriff der begrifflichen Notwendigkeit verwendet, manchmal auch mit dem der metaphysischen Notwendigkeit Putnam [265]. ‚Epistemische Möglichkeit‘, die hier nicht extra aufgeführt ist, wird synonym mit ‚begriffliche Möglichkeit‘ verwendet (Gleiches gilt für die entsprechende Notwendigkeit).

<sup>244</sup> Für eine Interpretation der Modalbegriffe, die mit allem kompatibel scheint, was ich zu möglichen Welten sagen möchte, vgl. Bremer und Cohnitz [34], Kapitel 5.3.

<sup>245</sup> Quine (beispielsweise) findet die Redeweise von Modalitäten mindestens so unklar wie die Redeweise von Analytizität, fällt also vermutlich nicht in diese Gruppe. Putnam hingegen scheint (bezüglich bestimmter Ausdrücke) der Meinung zu sein, dass es zwar keine Bedeutungsanalysen dieser Ausdrücke im herkömmlichen Sinne gibt, sie aber sehr wohl dazu verwendet werden können metaphysische Notwendigkeiten auszudrücken. Diese Auffassung wird uns im Folgenden noch näher beschäftigen.

macht dies eine Lösung nicht unbedingt leichter (wie wir im Folgenden noch sehen werden).

*Eine normative Uminterpretation?*

Ein weiterer Vorschlag besteht darin, Gedankenexperimente weniger als Test für eine Behauptung zu verstehen, die darauf abzielt, eine objektiv bestehende Tatsache zu beschreiben, sondern als Vorschlag, eine normativ verstandene Regelung abzuändern. (Vgl. hierzu die Diskussion in Häggqvist [137], Sorensen [308] und das letzte Kapitel dieser Untersuchung).

In diese Richtung lässt sich beispielsweise der Unterschied zwischen *deskriptiver* und *revisionärer* Metaphysik interpretieren. Während die deskriptive Metaphysik versucht, metaphysische Zusammenhänge, die eine Sedimentierung in der Sprache erfahren haben, aufzuzeigen, versucht die revisionäre Metaphysik, solche überkommenen Zusammenhänge abzuändern.<sup>246</sup> Grundsätzlich sind beide Positionen zunächst unabhängig davon, ob sie solche metaphysischen Zusammenhänge als sprachabhängig betrachten. Für revisionäre Metaphysiker ist es allerdings typisch, dass sie dies tun. Metaphysik ist demnach ein System von Konventionen, das – innerhalb gewisser Beschränkungen – unseren Wünschen folgend angepasst werden kann. Entsprechende Positionen ließen sich dann auch für die anderen systematischen Teilbereiche der Philosophie skizzieren.

Wir werden auf diese Auffassung noch ausführlich eingehen. Sie scheint eine relativ natürliche Position zu sein, wenn es um Gedankenexperimente in der Ethik geht. Bezieht man diese Interpretation auf das Gebiet der Metaphysik oder der Erkenntnistheorie, ist sie in zweifacher Hinsicht klärungsbedürftig: Einerseits scheint es keine Frage von Konventionen zu sei, ob man ein Ereignis überlebt hat, oder ob man etwas weiß, weshalb der vorgeschlagene Konventionalismus zumindest erläuterungsbedürftig ist. Zum anderen machen viele prominente Gedankenexperimente unter dieser Interpretation keinen guten Sinn. Warum sollte die Konvention, die unsere Redeweise über diachrone personale Identität reglementiert, so abgeändert werden, dass sie mit dem, was wir gerne über Teletransporterunfälle sagen möchten, kohärent ist? Weil Teletransporter bestimmt irgendwann mal gebaut werden und wir für diese Eventualität vorbereitet sein wollen? Weil wir möchten, dass unsere Gegenstücke in diesen anderen möglichen Welten solche Ereignisse überleben? Warum sollen unsere Konventionen an Fälle angepasst werden, die uns nicht begegnen, anstatt an Fälle, die alltäglicher sind? Im nächsten Kapitel wollen wir zur systematischen Klärung dieser Zusammenhänge verschiedene Definitionstypen betrachten, die sämtlich in der Philosophie vorkommen und von denen fast alle in irgendeiner Hinsicht mit einem besonderen Typ von Notwendigkeit verbunden sind. Viele philosophische Debatten lassen sich als Streit um Definitionen der einen oder anderen Art auffassen und Gedankenexperimente spielen in diesen Streitigkeiten eine wichtige kritische Rolle. An der systematischen Ordnung, die wir durch eine Klassifikation von Definitionen und den ihnen zugeordneten Mo-

---

<sup>246</sup> Vgl. Parfit [250] als Vertreter einer revisionären, Strawson [322] als Vertreter einer deskriptiven Metaphysik.

dalitäten erhalten, können wir Gedankenexperimente, wie auch die Kritik an ihnen besser verstehen. Was uns für den Rest der Untersuchung interessieren soll, sind keine allgemeinen Kritiken an der Methode des Gedankenexperiments überhaupt, sondern spezifische Kritiken an der Verwendungsweise von Gedankenexperimenten zur Diskussion bzw. Kritik bestimmter Theorien bzw. bestimmter Typen von Definitionen.

## 6.2 DEFINITIONEN UND IHRE MODALITÄTEN

Philosophische ‚Was ist?‘-Fragen, wie ‚Was ist Kunst?‘, ‚Was ist Wahrheit?‘, ‚Was ist Bewusstsein?‘, etc. scheinen auf den ersten Blick Fragen nach einer Definition zu sein. Wir wollen z.B. erfahren, was das für Gegenstände sind, die wir „Kunstwerke“ oder „Personen“ nennen, was sie für besondere Eigenschaften haben, die sie von anderen Gegenständen unterscheiden. Des Weiteren suchen wir offenbar ebenfalls nach einer Definition, die es uns erlaubt, Begriffe wie ‚Wissen‘, ‚Bewusstsein‘, ‚Kunst‘, etc. in Theorien zu verwenden. Was aber ist eine Definition genau? Wann ist eine Definition gut (bzw. besser als eine rivalisierende, alternative Definition)? In diesem Unterkapitel werden wir kurz betrachten, was Definitionen eigentlich sind und mit welchen der oben besprochenen Modalitäten sie einhergehen, um danach besser sagen zu können, inwiefern Gedankenexperimente gegen Definitionen gerichtet sein können.

### *Typen von Definitionen*

Was ist eine Definition? Klassischerweise unterscheidet man Definitionen in so genannte „Nominaldefinitionen“ und so genannte „Realdefinitionen“. Eine *Realdefinition* versteht man dabei (im Anschluss an Aristoteles) als eine Angabe der „essentiellen Charakteristika“ einer Sache. Darunter fallen beispielsweise solche Definitionen, wie ‚Der Mensch ist ein vernünftiges Tier.‘, oder ‚Ein Stuhl ist eine bewegliche Sitzgelegenheit für eine Person.‘. Eine *Nominaldefinition* wird hingegen als eine *Stipulation* (Festsetzung) aufgefasst, die einen neuen Begriff als konventionelle Abkürzung für einen längeren sprachlichen Ausdruck vorgibt. Carl Gustav Hempel (vgl. Hempel [144], Hempel [145]) Beispiel für eine solche Stipulation ist die Definition des Neologismus ‚Tiglon‘:

- (D1) Sei das Wort ‚Tiglon‘ eine Abkürzung (ein synonyme Ausdruck) für ‚Nachwuchs eines männlichen Tigers und eines weiblichen Löwen‘.

Wir wollen uns zunächst die Besonderheiten von Nominal- und Realdefinitionen näher ansehen. Betrachten wir zunächst die Nominaldefinitionen.

## 6.2.1 NOMINALDEFINITIONEN

Nach Hempel kann man eine Nominaldefinition auffassen als „eine Festsetzung, dergestalt dass ein spezifizierter Ausdruck, das *Definiendum*, synonym sei mit einem bestimmten anderen Ausdruck, dem *Definiens*, dessen Bedeutung bereits festgelegt ist.“<sup>247</sup> Demnach kann eine Nominaldefinition in die folgende Form gebracht werden:

(D2) Sei der Ausdruck  $E_2$  synonym mit dem Ausdruck  $E_1$ .

Diese Form wird beispielsweise von (D1) exemplifiziert. Wenn eine Nominaldefinition in der Form von (D2) aufgeführt wird, spricht sie offenbar über zwei sprachliche Ausdrücke, die ihr *Definiendum* und ihr *Definiens* konstituieren. Da *über* die sprachlichen Ausdrücke gesprochen wird, brauchen wir *Namen* für diese. *Explizite Nominaldefinitionen* können immer in die folgende Form gebracht werden:

(D3)  $\text{---} =_{\text{Df}} \text{---}$

Hierbei steht das *Definiendum* auf der linken Seite, das *Definiens* rechts. Es ist hierbei nicht notwendig, dass der neu einzuführende Ausdruck *alleine* auf der linken Seite erscheint. In vielen Fällen hat man vielmehr mit *kontextuellen* Definitionen zu tun, in denen der neu einzuführende Ausdruck ein Teil des Ausdrucks auf der linken Seite der Nominaldefinition ist. Worauf es bei Nominaldefinitionen vielmehr ankommt, ist, dass folgende Adäquatheitsbedingung erfüllt ist:

**(AB1) Eliminierbarkeit:** Eine Nominaldefinition muss angeben, wie der neu eingeführte Ausdruck aus jedem Kontext, in dem er grammatisch auftauchen kann, eliminiert werden kann.

Dies kann natürlich auch dadurch geschehen, dass eine Nominaldefinition angibt, wie ein Ausdruck aus einem *Kontext* eliminiert werden kann. Ein Beispiel dafür ist

(D4)  $x$  ist härter als  $y =_{\text{Df}} x$  schneidet  $y$ , aber  $y$  schneidet nicht  $x$ .

Nominaldefinitionen finden sich auch in der Philosophie. Sie spielen immer dann eine Rolle, wenn neue Fachterminologie eingeführt wird. Ein ganzes Netz verschachtelter Nominaldefinitionen findet sich beispielsweise bei Dickie (Dickie [87]), der zur Analyse des Kunstbegriffs beispielsweise die Ausdrücke ‚Kunstwelt‘, und ‚Kunstweltssystem‘ definiert:

---

<sup>247</sup> Hempel [144], 654.

- (D5) Kunstwelt =<sub>Df</sub> Gesamtheit aller Kunstweltsysteme.
- (D6) Kunstweltsystem =<sub>Df</sub> Menge der Rahmenbedingungen, die alle Situationen gemeinsam haben, in denen ein Kunstwerk durch einen Künstler einem Publikum präsentiert wird.

Eine weitere Adäquatheitsbedingung fordert von Nominaldefinitionen, dass sie *konservativ* sind, d.h., dass sie die Menge der wahren Sätze einer Theorie nicht erweitern:

(AB2) **Konservativität:** Eine Nominaldefinition darf innerhalb einer Theorie nicht die Ableitung neuer Theoreme gestatten, die ohne die Nominaldefinition aus der Theorie nicht hätten abgeleitet werden können.<sup>248</sup>

Wir wollen nun die Nominaldefinitionen und ihre Analyse zunächst verlassen und uns dem zweiten klassischen Definitionstyp, der *Realdefinition* zuwenden. Obwohl Nominaldefinitionen in der Philosophie eine wichtige Rolle spielen, scheinen sie doch nicht für unser Hauptproblem adäquat zu sein. Wie bereits erwähnt, soll es uns um Antworten auf die klassischen ‚Was ist‘-Fragen der Philosophie gehen. Eine bloße Nominaldefinition ist dabei in der Regel nicht, was uns als mögliche Lösung vorschwebt. Wir wollen ja nicht bloß eine Bedeutung für den fraglichen Ausdruck (‚Kunst‘, ‚Bewusstsein‘, ‚Wissen‘, ‚Wahrheit‘ etc.) *stipulieren* (das scheint – auch mit den bisherigen Adäquatheitsbedingungen – eine recht einfache Sache zu sein). Stattdessen gehen wir ja davon aus, dass diese Begriffe bereits eine Bedeutung haben und dass es darauf ankommt, etwas über sie *herauszufinden*, nicht sie bloß *festzulegen*.

## 6.2.2 REALDEFINITIONEN

Darüber, was man überhaupt unter einer Realdefinition verstehen will, gehen die Meinungen etwas auseinander. Für unsere Zwecke halten wir uns an eine etwas modifizierte Version der Verwendung des Wortes ‚Realdefinition‘ bei Carl Gustav

---

<sup>248</sup> Da man Nominaldefinitionen gewöhnlich als bloße Festsetzungen bzw. Konventionen betrachtet, ist es vielleicht verwunderlich, dass die Bildung solcher Festsetzungen Adäquatheitsbedingungen unterliegt. Dass jedoch durch eine schlecht gewählte Nominaldefinition Widersprüche in einer Theorie auftreten können, kann man kurz an folgendem Beispiel deutlich machen: Angenommen wir definieren innerhalb der Arithmetik die Pseudo-Operation ‚ $\odot$ ‘ auf folgende Art und Weise: ‚ $x \odot y = z$  =<sub>Df</sub>  $x < z$  und  $y < z$ ‘. Dank dieser Nominaldefinition lässt sich nun innerhalb der Arithmetik ein Widerspruch leicht ableiten. Zunächst einmal gilt offensichtlich ‚ $1 \odot 2 = 3$ ‘, da ‚ $1 < 3$ ‘ und ‚ $2 < 3$ ‘. Es gilt aber auch ‚ $1 \odot 2 = 4$ ‘, da ‚ $1 < 4$ ‘ und ‚ $2 < 4$ ‘. Daraus können wir ableiten, dass ‚ $3 = 4$ ‘. Die Adäquatheitsbedingungen der *Eliminierbarkeit* und *Konservativität* sollen dafür sorgen, dass solche Widersprüche nicht auftreten können.

Hempel und Rudolf Carnap. Nach Hempel versteht man *traditionellerweise* unter einer Realdefinition eine *Wesensdefinition*, also eine Angabe der *wesentlichen Eigenschaften einer Sache*. Dabei geht man davon aus, dass es beispielsweise so etwas gibt wie „das Gute“, dessen Wesen radikal verschieden sein kann von dem, was die Sprecher des Deutschen mit dem Wort ‚gut‘ bezeichnen. Wir werden auf diese Problematik im Detail im nächsten Kapitel eingehen. Was man unter den „wesentlichen Eigenschaften“ einer Sache verstehen will, ist dabei bedauerlich unklar. Viel fruchtbarer scheint es für uns zunächst zu sein, Realdefinitionen entweder als *Sachanalysen* oder als *Bedeutungsanalysen* zu verstehen (und dabei davon auszugehen, dass Wesensanalysen im Sinne einer kausalen Referenztheorie unter den ersten Typ subsumiert werden können). Wir werden dieser Auffassung folgen und verwenden den Ausdruck ‚Realdefinition‘ entsprechend nicht gemäß der Tradition.

Hempels Beispiel für eine Realdefinition ist folgende Definition von ‚lebender Organismus‘ nach Hutchinson:

- (D7)  $x$  ist ein lebender Organismus gdw. (i)  $x$  aus einer diskreten Masse von Materie besteht, die eine bestimmte äußere Grenze besitzt, (ii)  $x$  kontinuierlich Materie mit seiner Umgebung austauscht, ohne innerhalb kurzer Zeiträume seine Eigenschaften stark zu verändern, und (iii)  $x$  durch Teilung oder Abtrennung aus einem oder aus zwei Objekten derselben Art entstanden ist.

Bei Hutchinson wird diese Definition offensichtlich nicht als Konvention eingeführt, wie der Ausdruck ‚lebender Organismus‘ von nun an zu benutzen sei. Bei dieser Definition scheint es sich vielmehr um einen Sprechakt zu handeln, der auf Wahrheit abzielt. (D7) – im Gegensatz zu den Nominaldefinitionen – kann offenbar wahr oder falsch sein.<sup>249</sup> Was aber sind die *Wahrheitsbedingungen* für solche Definitionen? Wie bereits angedeutet, können zwei unterschiedliche Interpretationen gegeben werden:

(D7) kann zunächst als *Synonymiebehauptung* verstanden werden, wobei der Unterschied zu einer Nominaldefinition eben darin bestünde, dass eine Synonymie *behauptet* wird, nicht stipuliert. Was dabei behauptet wird, ist, dass der Ausdruck links vom ‚gdw.‘ im *Deutschen* synonym ist mit dem Ausdruck rechts vom ‚gdw.‘. Eine solche Definition wollen wir im Folgenden mit den Wörtern ‚Bedeutungsanalyse‘, ‚Begriffsanalyse‘, oder ‚analytische Definition‘ bezeichnen. Da es hier und im Folgenden also um „Analysen“ geht, reden wir auch von ‚Analysans‘ und ‚Analysandum‘ statt von ‚Definiens‘ und ‚Definiendum‘, wobei wiederum ‚Analysandum‘ den Teil links vom ‚gdw.‘, ‚Analysans‘ den rechten Teil bezeichnet. Die Überprüfung einer solchen Bedeutungsanalyse erfordert offensichtlich keine em-

---

<sup>249</sup> Arthur Pap schlägt vor, auch Realdefinitionen als Stipulationen aufzufassen, die keinen Wahrheitswert besitzen. Er folgt damit der Intuition, dass nur, was wie eine Nominaldefinition funktioniert, auch eine Definition ist, wir folgen hier der Intuition, das alles, was wie eine Definition aussieht, auch eine Definition ist. Vgl. Pap [248], 447.

pirische Untersuchung lebender Organismen, sondern „lediglich“ eine Untersuchung der Bedeutungen der die Definition konstituierenden Ausdrücke im Deutschen.

(D7) kann hingegen aber auch so verstanden werden, dass es eine Tatsache ist, dass die Bedingungen (i) bis (iii) von allen und ausschließlich denjenigen Gegenständen erfüllt werden, die auch lebende Organismen sind. (D7) hätte dann eher den Charakter eines Gesetzes und die Beurteilung seines Wahrheitswertes würde eine Untersuchung der Charakteristika lebender Organismen erfordern. Definitionen dieser Art wollen wir im Folgenden als ‚Sachanalyse‘ bezeichnen. (D7) wäre dann eine Sachanalyse der Eigenschaft ein lebender Organismus zu sein.

Wenn hier von ‚Sachanalyse‘ gesprochen wird, dann weichen wir vom Sprachgebrauch Hempels ab, der diese Art der Definition mit dem Ausdruck ‚empirische Analyse‘ bezeichnet hat. Die Sachanalyse als Definitionstyp hat aber eigentlich zunächst nichts damit zu tun, dass sie nur durch „empirische“ Untersuchung (also a posteriori) als wahr erwiesen werden kann. Sachanalysen unterscheiden sich von Bedeutungsanalysen und Nominaldefinitionen zunächst einfach dadurch, dass sie nicht über die Ausdrücke reden, die rechts und links vom ‚gdw.‘ stehen, sondern über die durch diese Ausdrücke bezeichneten Gegenstände oder Sachverhalte.<sup>250</sup> Wir werden im Folgenden auch manchmal den Ausdruck ‚metaphysische Analyse‘ verwenden, um diesen Definitionstyp zu bezeichnen.

*Nominaldefinitionen*, *Bedeutungsanalysen* und *Sachanalysen* sind alle voneinander verschieden und unterliegen entsprechend auch unterschiedlichen Adäquatheitsbedingungen. Wir hatten bereits bemerkt, dass Realdefinition Sprechakte sind, die auf Wahrheit abzielen, weshalb sich trivialerweise die Adäquatheitsbedingung ergibt, *dass Realdefinitionen wahr sein sollen*. Eine Adäquatheitsbedingung, die man für Nominaldefinitionen nicht in sinnvoller Weise stellen konnte.

**(AB3) Wahrheit von Realdefinitionen:** Realdefinitionen sollen wahr sein.

Bei Sachanalysen wird man dieser Forderung dadurch gerecht, dass man angibt, welche notwendigen und hinreichenden Bedingungen vorliegen müssen, damit ein bestimmtes Phänomen (nämlich das durch das *Analysandum* bezeichnete Phänomen) auftritt. Nach Hempel besitzen solche Analysen gemeinhin den Charakter eines allgemeinen Gesetzes, das wir (nach herkömmlicher Auffassung) *a posteriori* durch die empirische Untersuchung des *Analysandums* entdecken. Wie wir schon sagten, werden wir diesen Ausdruck aber so verwenden, dass er es auch zulässt, dass (aus irgendwelchen Quellen) a priori von seiner Wahrheit Kenntnis erlangen können.

Wie wir im nächsten Kapitel noch näher erläutern werden, kann zwischen der *Referenz* eines Ausdrucks (wie z.B. H<sub>2</sub>O im Fall des Ausdrucks ‚Wasser‘) und den *referenzfixierenden Eigenschaften* eines Ausdrucks (z.B. „wässrig“ zu sein, also die

---

<sup>250</sup> Hempel war freilich der Meinung, dass sich Sätze, die sich auf die Wirklichkeit beziehen, nur durch empirische Untersuchung als wahr bzw. falsch erwiesen werden können.

phänomenalen Eigenschaften von Wasser zu besitzen) unterschieden werden. Mit Hilfe dieser Unterscheidung kann man unsere Begriffsverwendung folgendermaßen präzisieren: Es wird bei Sachanalysen davon ausgegangen, dass die *Referenz* der in ihnen vorkommenden Ausdrücke bereits festliegt und die *referenzfixierenden Eigenschaften* bekannt sind. Beides ist bei einer Nominaldefinition beispielsweise nicht der Fall, da hier der Sinn eines Ausdruckes durch die *Stipulation* von Synonymiebeziehungen die Referenz der Ausdrücke erst fixiert. Anders ausgedrückt: Nominaldefinitionen stipulieren die referenzfixierenden Eigenschaften eines Ausdruckes, Sachanalysen setzen sie als festgelegt *und* (zumindest implizit) bekannt voraus. Sachanalysen sind also wahr gdw. das durch das *Analysandum* bezeichnet Phänomen dann und nur dann auftritt, wenn die im *Analysans* genannten Bedingungen auftreten:

**(AB4) Wahrheit von Sachanalysen:** Sachanalysen sind wahr gdw. das durch das Analysandum bezeichnete Phänomen mit Notwendigkeit (Naturnotwendigkeit bzw. metaphysischer Notwendigkeit) dann und nur dann auftritt, wenn die im Analysans genannten Bedingungen auftreten.

Bei Bedeutungsanalysen wird ebenfalls davon ausgegangen, dass diese Fixierungen in einer Sprache bereits festliegen, die referenzfixierenden Eigenschaften aber durch die Aufdeckung bestehender Synonymiebeziehungen erst explizit gemacht werden müssen. Demnach ist eine Bedeutungsanalyse eine Behauptung, die wahr oder falsch ist, je nachdem ob das Analysans in einer gegebenen Sprache mit dem Analysandum synonym ist oder nicht.

**(AB5) Wahrheit von Bedeutungsanalysen:** Bedeutungsanalysen sind wahr für eine Sprache S, gdw. in S der Analysans- und Analysandum-Ausdruck synonym sind.

Dabei werden verschiedene Voraussetzungen gemacht, die allesamt nicht unproblematisch sind: Zunächst setzt diese Bestimmung für die Durchführbarkeit einer Bedeutungsanalyse eine Sprache voraus, die präzise bestimmte Bedeutungen besitzt (zumindest für die Ausdrücke, für die man sich interessiert) – man muss ja von zwei Ausdrücken immer angeben können, ob sie synonym sind oder nicht. Will man *die* Bedeutung eines Ausdruckes mit *der* Bedeutung eines anderen Ausdruckes in einer Sprache vergleichen, muss man außerdem annehmen, dass es so etwas überhaupt gibt, dass also die Anwendungsbedingungen für jeden Sprecher einer Sprache wohlbestimmt sind und darüber hinaus für alle Sprecher einer Sprache auf die gleiche Weise wohlbestimmt sind (zumindest für einen gegebenen Zeitraum und eine relevante Gruppe von Sprechern). Hempel nennt die erste Voraussetzung die Bedingung der „Determinanz“, die zweite die Bedingung der „Uniformität des Gebrauchs“. Beide Bedingungen werden im Folgenden (insbesondere in Kapitel 7.3) zu diskutieren sein.

Außerdem scheint die Bedeutungsanalyse, als philosophische Methode, aber auch ein übermäßiges Vertrauen in die Leistungsfähigkeit der Umgangssprache zu haben: warum sollte die tatsächliche Verwendung eines philosophisch interessan-

ten Ausdrucks in einer natürlichen Sprache philosophisch erhellend sein? Kann es nicht sein, dass sich in der Umgangssprache durch unreflektierten Umgang mit bestimmten Ausdrücken Inkonsistenzen oder Implausibilitäten verfestigt haben? Bedeutungsanalyse ist in diesem Sinne ein rein *deskriptives* Unterfangen, da durch sie nur die *tatsächliche* Verwendung von Ausdrücken in einer natürlichen Sprache zu Tage gefördert wird. Stößt man darauf, dass diese tatsächliche Verwendung nicht plausibel oder widersprüchlich oder vage ist, könnte man als Philosoph seine Aufgabe ja auch darin sehen, Präzision *herzustellen*. Tätigkeiten dieser Art sind weder reine Bedeutungsanalysen, noch reine Nominaldefinitionen, man bezeichnet sie – im Anschluss an Rudolf Carnap – als *Begriffsexplikationen*.

### 6.2.3 BEGRIFFSEXPLIKATIONEN

Die Begriffsexplikation befasst sich mit Ausdrücken, die zwar in der Umgangssprache oder sogar in der Wissenschaftssprache vorkommen, von denen aber bereits festgestellt worden ist, dass ihre Verwendung mehr oder weniger unklar ist. Der Zweck der Begriffsexplikation besteht dann entsprechend darin, diesen Ausdrücken eine neue und präzise Bedeutung zu geben, um sie für eine philosophische Theorie, also eine strenge Erörterung des Gegenstandsbereiches brauchbar zu machen.

Man geht dabei davon aus, dass durch eine weitere Reflexion auf den tatsächlichen umgangssprachlichen Gebrauch dieser Ausdrücke keine weitere Erkenntnis zu erreichen ist, da entweder die Verdeutlichung verlangenden Sachverhalte zu subtil sind, oder festgestellt wurde, dass der umgangssprachliche Gebrauch ohnehin inkohärent ist. Wie Nominaldefinitionen sind solche Begriffsexplikationen also normativ, da sie eine neue, präzisiertere Ausdrucksverwendung vorschlagen. Wie Nominaldefinitionen können sie daher auch nicht wahr oder falsch sein. Im Gegensatz zu Nominaldefinitionen haben Begriffsexplikationen als „Explicandum“ allerdings einen Ausdruck, der bereits in der Umgangs- oder Wissenschaftssprache vorkommt. Daher sind sie nicht völlig frei in der Stipulation neuer, präzisierter Anwendungsbedingungen, die wir als ‚Explicatum‘ bezeichnen. Würde die „neue“ Bedeutung zu radikal von der umgangs-, bzw. wissenschaftssprachlichen „herkömmlichen“ Bedeutung des Explicandums abweichen, könnte dies (a) zu Missverständnissen führen, und (b) könnte eine Explikation unser *philosophisches* Problem nicht lösen. Letzteres besteht ja nicht darin, dass es *irgendwelche* vagen Ausdrücke in unserer Umgangssprache gibt, sondern darin, dass wir uns für diesen bestimmten Ausdruck interessieren, weil er irgendeine philosophische Relevanz besitzt. Wenn wir also eine Begriffsexplikation solcher Ausdrücke wie ‚Kunst‘, ‚Zahl‘, ‚Wahrscheinlichkeit‘ oder ‚wissenschaftliche Erklärung‘ geben, soll sie natürlich nicht eine völlig von der Umgangssprache abweichende Bedeutung für diese Begriffe stipulieren. Eine Theorie, in der unser dann präzisierter Ausdruck vorkommen soll, soll ja weiterhin über Kunst, oder Zahlen, Wahrscheinlichkeiten

oder wissenschaftliche Erklärungen sein und nicht über irgendetwas beliebiges anderes.

Wie Carnap bereits richtig bemerkte, haben Explikationen die Schwierigkeit, dass die eigentliche Aufgabenstellung unklar ist. Das *Explicandum* ist ja gerade der vage, unklare Begriff. Es entsteht also die Schwierigkeit sich darüber zu verständigen, was man eigentlich machen will, wenn man eine Explikation von Ausdruck *X* angeben möchte. Das Verfahren, das Carnap vorschlägt, besteht darin, zunächst eine Bedeutungsanalyse zu betreiben, soweit das eben möglich ist, und dann – angesichts der Vagheiten, die dabei zu Tage gefördert werden – anhand von Beispielen Verständigung über die Verwendungsweise desjenigen *Explicandums* zu erlangen, dass man sich zur Aufgabe stellen will. Solche Beispielsätze im Fall des Ausdrucks ‚Kunst‘ könnten durch folgende Aufgabenstellung angegeben werden:

- (B1) Ich will ‚Kunst‘ im klassifikatorischen Sinne explizieren, also so, wie in Sätzen wie ‚Das ist aber ein schlechtes Kunstwerk.‘, ‚Das ist ein gutes Kunstwerk.‘, aber nicht wie in Sätzen wie ‚Dieses Geschmiere ist doch keine Kunst.‘, ‚Nur die Arbeiten echter Genies sind Kunst.‘, etc.

Mit (B1) will jemand darauf hinweisen, dass er den Ausdruck ‚Kunst‘ in einem klassifikatorischen Sinn explizieren möchte, so dass das *Explicatum* auch die Redeweise von ‚schlechter Kunst‘ erlaubt, nicht aber in einem lobenden Sinn, der die Redeweise von ‚schlechter Kunst‘ gar nicht mehr zuließe. Solche Beispielsätze sind ebenfalls allesamt Grundlage für *spezielle Adäquatheitsbedingungen* für eine gelungene Explikation. Später lässt sich nämlich testen, ob das *Explicatum* *salva veritate* in diese Beispielsätze für das *Explicandum* ersetzt werden kann. Man kann dies auf einfache Weise so durchführen, dass man Aussagen aus der Umgangssprache aussondert, die den fraglichen Begriff im intendierten Sinne enthalten und dann diese Sätze *mit* ihren Wahrheitsbewertungen zu Adäquatheitsbedingungen macht.

Hat man nun mit Hilfe solcher Adäquatheitsbedingungen eine Verständigung darüber erreicht, was die Aufgabenstellung der Explikation ist, geht die eigentliche Arbeit allerdings erst los. Nun geht es darum, ein geeignetes *Explicatum* zu finden. Auch hierbei ist man nicht völlig frei, sondern hat sich wiederum nach allgemeinen Adäquatheitsbedingungen für gelungene Explikationen zu richten.

Wie aus unseren obigen Bemerkungen bereits hervorgeht, soll die Explikation sich ja möglichst mit der umgangssprachlichen Verwendung decken. Völlige Deckung kann dabei genauso wenig erstrebenswert sein wie die größtmögliche Deckung gegeben die Vagheit des umgangssprachlichen Ausdrucks. Was man ja anstrebt, ist eine Präzisierung, die *fruchtbar* in Theorien weiterverwendet werden kann. Letzteres kann dabei dafür sorgen, dass man zu einem nicht unbeträchtlichen Grad von der umgangssprachlichen Verwendung abweichen wird. So weicht beispielsweise die zoologische Explikation von ‚Fisch‘ von dem umgangssprachlichen Ausdruck ‚Fisch‘ erheblich ab. Der umgangssprachliche Ausdruck, nennen wir ihn ‚Fisch<sub>U</sub>‘ bezeichnet auch Wale und Delfine, während der zoologische Ausdruck ‚Fisch‘, nennen wir ihn ‚Fisch<sub>Z</sub>‘, dies nicht mehr tut. Fisch<sub>U</sub> und Fisch<sub>Z</sub> sind also alles andere als deckungsgleich. Diese Einengung der Bedeutung wurde

in der Zoologie vorgenommen, da man annahm, dass Fisch<sub>z</sub> fruchtbarer zu sein versprach. Man sah, dass „die Tiere, auf die [Fisch<sub>u</sub>] zutrifft, nämlich die Tiere, die im Wasser leben, bei weitem nicht so viele gemeinsame Eigenschaften teilen, wie die Tiere, die poikilotherme Wirbeltiere sind und während ihres gesamten Lebens Kiemen besitzen.“<sup>251</sup> Fisch<sub>z</sub> führt also zu einer größeren Zahl von generalisierbaren Aussagen als eine Explikation, die näher an Fisch<sub>u</sub> angelehnt gewesen wäre.<sup>252</sup>

Schließlich wollen Wissenschaftler aber nicht nur fruchtbare Explikationen, sondern auch *möglichst einfache*, zumindest wenn man die Wahl zwischen verschiedenen, gleich fruchtbaren Explikationen hat. Einfachheit wird dabei einerseits in der Einfachheit der Form der Definition gesehen, zum anderen in der Einfachheit der Form derjenigen Gesetze, die das Explicatum mit anderen wissenschaftlichen Begriffen in Beziehung setzen. Damit hätten wir bereits drei neue Adäquatheitsbedingungen, diesmal für Begriffsexplikationen:

**(AB6) Ähnlichkeit zur Umgangssprache:** Das Explicatum einer Begriffsexplikation soll dem Explicandum möglichst ähnlich sein, d.h., dass in den meisten Fällen, in denen das Explicandum bisher benutzt wurde, nun das Explicatum benutzt werden kann. Ein großer Grad von Ähnlichkeit wird aber nicht gefordert und auch beträchtliche Unterschiede sind erlaubt.

**(AB7) Fruchtbarkeit:** Das Explicatum soll ein fruchtbarer Ausdruck sein, d.h. ein Ausdruck, der zur Formulierung möglichst vieler universaler Aussagen benutzt werden kann.

**(AB8) Einfachheit:** Das Explicatum sollte so einfach wie möglich sein, also so einfach, wie die anderen Adäquatheitsbedingungen dies erlauben.

In diesen Adäquatheitsbedingungen steckt bereits, was John Rawls als „reflektiertes Gleichgewicht“ bezeichnet hat (vgl. Rawls [274]). Begriffsexplikationen sind – wie schon gesagt – im Vergleich mit Bedeutungsanalysen nicht rein deskriptiv, sondern haben normativen Charakter: in ihnen wird präzisiert festgelegt, was mit einem Ausdruck korrekt bezeichnet wird. Diese Normen bauen auf unseren Intuitionen auf, die in (AB6) eingehen, während die Berücksichtigung dieser Intuitionen durch die Erfordernisse einer möglichst präzisen Norm eingeschränkt wird, was (AB7) und (AB8) leisten. Es wird damit ein „reflektiertes Gleichgewicht“ zwischen unseren vorwissenschaftlichen Intuitionen einerseits und unseren wissenschaftlichen Bedürfnissen andererseits geschaffen. Dieses reflektierte Gleichgewicht spielt bei der Bewertung von Begriffsexplikationen eine weitaus größere Rolle als bei der Bewertung von Bedeutungsanalysen, bei denen unsere vorwissenschaftlichen Sprachintuitionen fast alles sind, worauf es ankommt.

---

<sup>251</sup> Carnap [50], 6.

<sup>252</sup> In Cohnitz und Smith [76] diskutieren wir ein Beispiel einer Begriffsexplikation, die diese Adäquatheitsbedingung zu verletzen scheint.

Wie aus unseren Erläuterungen offensichtlich sein sollte, unterliegen nicht alle Typen von Definitionen den gleichen Standards. Will man also eine vorgegebene Definition in der Philosophie bewerten, bzw. kritisieren, muss man sich zunächst Klarheit darüber verschaffen, mit was für einer Art von Definition man es überhaupt zu tun hat. Je nach Definitionstyp gelten verschiedene Adäquatheitsbedingungen, die angeben, was eine angemessene Kritik an einer Definition sein kann, und was nicht.

Wir hatten bereits gesagt, dass alle Definitionen notwendige und hinreichende Bedingungen für das Vorliegen des Definiendums angeben müssen. D.h., dass z.B. eine Definition von Kunst alle Kunstwerke einschließt und alle nicht-Kunstwerke ausschließt. Im Fall von Realdefinitionen und Begriffsexplikationen, die sich ja (ganz oder zum Teil) an unserer intuitiven Sprachverwendung orientieren müssen, können solche Definitionen „zu weit“, oder „zu eng“ ausfallen. Zu weit ist z.B. eine Definition von Kunst, wenn sie auch Dinge umfasst, die nicht zu den Kunstwerken gehören.

(D8)  $x$  ist ein Kunstwerk gdw.  $x$  ein Artefakt ist

ist beispielsweise zu weit, da nach dieser Definition auch Autos und Kugelschreiber und herkömmliche Waschmaschinen Kunstwerke wären, was nun mal nicht stimmt. Zu eng ist eine Definition, wenn sie nicht alle Kunstwerke umfasst.

(D9)  $x$  ist ein Kunstwerk gdw.  $x$  am 19. März 2002 im Louvre in Paris in einem Rahmen hing.

ist in diesem Sinne zu eng, (D9) umfasst zwar (vermutlich) ausschließlich Kunstwerke, aber nur eine kleine Teilmenge derjenigen Gegenstände, die eigentlich Kunstwerke sind.

Begriffsexplikationen können auf ihre Weise ebenfalls zu eng oder zu weit sein, wenn sie nämlich die das Explicandum festlegenden „speziellen Adäquatheitsbedingungen“ nicht erfüllen, also nicht konservativ in Bezug auf diejenigen Sätze sind, die zuvor ausgewählt wurden, um den Begriff einzugrenzen, der durch das Explikat expliziert werden soll.

Inwiefern die Enge oder Weite von Definitionen mit Gedankenexperimenten getestet werden kann, kann man an den Modalitäten ablesen, die durch die jeweiligen Definitionen ausgedrückt werden sollen. Nominaldefinitionen, die bloße Abkürzungen liefern, fallen dabei aus unserem Interessensbereich. Realdefinitionen sind aber für unser Thema interessant. Entweder handelt es sich um Sachanalysen, dann behaupten sie einen naturnotwendigen (bzw. auch einen „metaphysischen“) Zusammenhang. Ein Gedankenexperiment gegen eine solche Sachanalyse müsste dann ein Gegenbeispiel präsentieren, das naturmöglich bzw. metaphysisch möglich ist.

Bedeutungsanalysen hingegen behaupten eine begriffliche Notwendigkeit. Um sie zu widerlegen, würde ein Beispiel ausreichen, das begrifflich möglich ist.

Explikationen schließlich scheinen nur bedingt von Gedankenexperimenten angreifbar zu sein. Schließlich können sie nicht wie Aussagen wahr oder falsch

sein. Nichtsdestotrotz können sie natürlich die Adäquatheitsbedingungen für Explikationen verfehlen und dann inadäquat sein. Inwiefern Gedankenexperimente in diesem Zusammenhang eine Rolle spielen, wird aber erst Gegenstand von Teil 9 sein.

Ausgerüstet mit dieser Zuordnung von Definitionstypen und Modalitäten können wir uns nun in systematischer Weise der spezifischen Kritik an Gedankenexperimenten in der Philosophie zuwenden. Beginnen werden wir mit der Frage, wie die Kritik an Gedankenexperimenten zu bewerten ist, die gegen Realdefinitionen vorgebracht wird.

## 7. Gedankenexperimente gegen Realdefinitionen

Die Einwände gegen philosophische Gedankenexperimente, die wir bisher betrachtet haben, waren sehr allgemeiner Art. Sie bezogen sich auf alle Gedankenexperimente in der Philosophie und nicht auf Gedankenexperimente in bestimmten Teilbereichen oder in Bezug auf bestimmte Fragen. Die Vorwürfe, die wir nun diskutieren werden, sind immer noch allgemein genug, um auf eine große Menge systematischer Teilbereiche der Philosophie ausgeweitet werden zu können. Obwohl diese Vorwürfe in bestimmten Debatten vorgebracht werden, haben sie alle die Gestalt ‚Gedankenexperimente sind unbrauchbar zur Klärung der Frage X, weil X die Eigenschaft Q hat.‘, wobei Q eine Eigenschaft ist, die bei philosophischen Fragen mutmaßlich häufig auftritt. Kontingenterweise sind diese Vorwürfe hauptsächlich in zwei systematischen Teilbereichen erhoben worden, in der Metaphysik personaler Identität und der Erkenntnistheorie. Das hat wohl damit zu tun, dass im ersten Fall Gedankenexperimente sehr häufig vorkommen und in ihnen sehr bizarre Umstände geschildert werden, während im zweiten Fall ein aufstrebendes Naturalisierungsprogramm<sup>253</sup> zur Methodenreflexion Anlass gab.

Die Einwände, die in diesem Kapitel gegen die Methode des Gedankenexperiments erhoben werden, sind jedoch insofern spezifisch, als sie Gedankenexperimente deshalb kritisieren, weil diese gegen Überzeugungstypen gerichtet sind, die sich – aus irgendeinem Grund – zur Kritik durch Gedankenexperimente nicht eignen sollen. Dies erlaubt eine systematische Durchmusterung dieser Kritiken im Anschluss an unsere Überlegungen im vorhergehenden Teil. Im Rest dieser Arbeit wird nun die in Teil 6 skizzierte funktionale Ausdifferenzierung des Gedankenexperiments zur Überzeugungsänderung weiter erläutert und die gegen Gedankenexperimente vorgetragenen Kritiken in Bezug auf ihre Relevanz für diese zu unterscheidenden Funktionen diskutiert.

Beginnen wollen wir in Teil 7 mit der Kritik an Gedankenexperimenten, die sich gegen Realdefinitionen richten, wobei wir zunächst (7.1-7.2) empirische bzw. metaphysische Analysen betrachten und uns dann (in 7.3) der Kritik an Gedankenexperimenten gegen Bedeutungsanalysen zuwenden.

Allen Gedankenexperimenten gegen Realdefinitionen ist gemeinsam, dass in ihnen eine Aussage über eine Möglichkeit bestimmter Art getroffen wird. In Teil 8

---

<sup>253</sup> Mit ‚Naturalismus‘ wird in dieser Arbeit durchgängig die Position bezeichnet, die davon ausgeht, dass es keine genuinen philosophischen Fragen oder Methoden gibt. Jede traditionelle „philosophische“ Frage kann – sofern sie sinnvoll ist – mit den Mitteln der Naturwissenschaften oder der Logik beantwortet werden. Die Position des Naturalismus steht in dieser Arbeit nicht zur Debatte.

werden wir modalepistemologische Kritiken an der Methode des Gedankenexperiments diskutieren, die grundsätzlich in Frage stellen, dass wir gerechtfertigte Aussagen über (nicht-aktuale) Möglichkeiten treffen können.

In Teil 9 werden wir die bis dahin erzielten Ergebnisse zusammentragen und zum Abschluss unserer Untersuchung die Kritik an Gedankenexperimenten gegen Begriffsexplikationen diskutieren. Wie wir im letzten Kapitel gesehen haben, treffen Begriffsexplikationen gar keine Aussagen, sondern machen sprachliche Normierungsvorschläge. Wie sollen Gedankenexperimente zu deren Revision beitragen können?

## 7.1 KRITIK AN GEDANKENEXPERIMENTEN GEGEN SACHANALYSEN

### 7.1.1 GEDANKENEXPERIMENTE SIND UNZUVERLÄSSIG, WENN ES UM EINEN BEGRIFF GEHT, DER KEINE NATÜRLICHE ART BEZEICHNET

Eine der bekanntesten Kritiken an philosophischen Gedankenexperimenten ist von Kathleen Wilkes verfasst worden (vgl. Wilkes [345]).<sup>254</sup> Wilkes Buch *Real People* ist ein Beitrag zur Debatte um diachrone Personenidentität, der den Versuch unternimmt, eine philosophische Position ausschließlich an empirischen Beispielen zu entwickeln. Die Position, die dabei entstanden ist, wie auch die Frage nach der Art ihres Zustandekommens braucht uns hier nicht zu interessieren. Von Interesse ist, dass Wilkes ihrer Abhandlung eine methodologische Kritik an philosophischen Gedankenexperimenten vorausschickt, in der sie zu dem Schluss kommt, dass die Mehrzahl philosophischer Gedankenexperimente aus prinzipiellen Gründen erfolglos ist (weshalb sie für den Rest ihres Buches a priori-Methoden abschwört und sich auf die „Empirie“ verlässt<sup>255</sup>). Dass philosophische Gedankenexperimente mit nur zweifelhaftem Erfolg gekrönt sind, liegt nach Wilkes daran, dass philosophische Gedankenexperimente Gegenstände haben, die in der Regel keine *natürlichen Arten* sind. In den Naturwissenschaften sei dies anders, weshalb Gedankenexperimente dort zuverlässiger seien:

This helps to explain why thought experiments in physics have generally been effective and substantial [...]; why thought experiments in philosophy that pick on natural-kind terms, such as ‘gold’ or ‘water’, seem conclusive [...]; and why other thought experiments in philosophy in general, and those concerning personal identity in particular, are not so [...]. For it is improbable that ‘person’ is in any legiti-

<sup>254</sup> Vgl. zur Kritik an Wilkes auch Martin [200]. Diese Kritik ist in großer Übereinstimmung mit der Auffassung von Gedankenexperimenten, wie sie im letzten Kapitel dieser Arbeit entwickelt wird.

<sup>255</sup> Inwiefern Wilkes Beispiele wirklich „empirisch“ sind, kann man zum Teil anzweifeln. Bei den darauf errichteten Argumenten handelt es sich dann häufig um naturalistische Fehlschlüsse.

mate sense a term that can usefully be constructed as a 'natural kind' [...]. (Wilkes [345], 15)

Wilkes teilt unsere Auffassung, dass Gedankenexperimente zum Test wissenschaftlicher Theorien herangezogen werden (Wilkes [345], 2), wobei sie die Logik des Gedankenexperiments auch in Übereinstimmung mit dem Häggqvist-Schema beschreibt. Demnach testen wir eine philosophische These dadurch, dass wir eine mögliche Welt beschreiben, die in allen relevanten Hinsichten unserer gleicht, in der aber bestimmte andere Dinge vorkommen oder existieren, die in unserer Welt nicht vorkommen und die die zu prüfende These falsifizieren oder unplausibel machen.

Der Hauptpunkt von Wilkes Argumentation bezieht sich darauf, was als „relevant“ zu gelten hat, wenn eine mögliche Welt beschrieben wird, die der unseren in allen „relevanten“ Hinsichten gleichen soll. Würde man nämlich einfach *jede* mögliche Welt als zulässig betrachten, wäre Philosophie von *Science Fiction* nicht mehr zu unterscheiden. Wenn es bei Gedankenexperimenten um mehr als um bloße Hirngespinnste gehen soll, muss man die Zulässigkeit möglicher Welten auf die relevanten möglichen Welten beschränken. Wie soll man dies tun?

An dieser Stelle vergleicht Wilkes Gedankenexperimente mit realen Experimenten. Auch bei diesen müsse man sicherstellen, dass die Experimentalbedingungen das Ergebnis des Experiments nicht verfälschen:

The experimenter – any experimenter, in thought or in actuality – needs to give us the background conditions against which he sets his experiment. If he does not, the results of his experiment will be inconclusive. The reason for this is simple and obvious: experiments, typically, set out to show what difference some factor makes; in order to test this, other relevant conditions must be held constant, and the problematic factor juggled against that constant background. If several factors were all fluctuating, then we would not know which of them (or which combination of them) to hold responsible for the outcome. (Wilkes [345], 7)

Hintergrundwissen über die zu beobachtenden Prozesse und Messapparaturen hilft uns dabei, zulässige von unzulässigen Experimentalbedingungen zu unterscheiden. Wir wissen, gegen welche Einflüsse wir ein Experiment ggf. abschirmen müssen und welche anderen Umstände wir vernachlässigen können, weil sie bekanntermaßen einflusslos sind. Bei einem realen Experiment können wir in der Regel auf eine gut bestätigte Annahme, dass alle relevanten Faktoren bedacht sind, vertrauen. Diese Annahme ist deshalb gut bestätigt, weil die Faktoren, die im Experiment untersucht werden sollen, in Theorien vorkommen, die bereits gut bestätigt sind und die diese Annahme stützen.

Bei Gedankenexperimenten werden die Hintergrundbedingungen von uns stipuliert. Dabei können zwei Probleme auftreten. Erstens kann es sein, dass wir die Hintergrundbedingungen eines Gedankenexperimentes nicht genau genug angeben, weshalb ein Urteil über den geschilderten hypothetischen Fall unterbe-

stimmt wäre.<sup>256</sup> Es könnte ja sein, dass solche Details des Hintergrundes, die nicht ebenfalls angegeben wurden, unser Urteil ändern würden. Einen solchen Einwand haben wir von Dancy in Kapitel 4.2 bereits kennen gelernt. In der Tat verpflichtet sich ein Gedankenexperimentierer darauf, relevante Hintergrunddetails nachliefern zu können, wenn er ein Gedankenexperiment vorschlägt. Dies ist sicherlich kein Einwand gegen Gedankenexperimente *per se*, sondern nur der Rat, beim Gedankenexperimentieren eine Geschichte zu wählen, die entweder sehr detailliert angegeben werden kann, oder so hinreichend nah an der aktuellen Welt ist, dass man annehmen kann, dass die Hintergrundbedingungen so wie in unserer Welt gegeben sind (und die stipulierte Geschichte dann konsistent ist, weil es in der Wirklichkeit keine Widersprüche gibt<sup>257</sup>).

Wilkes ist außerdem nicht entgangen, dass viele erfolgreiche naturwissenschaftliche Gedankenexperimente häufig Umstände schildern, die naturunmöglich sind. So spiele in Stevins Gedankenexperiment (das wir in Kapitel 2 und 3 kennen gelernt haben) absolute Reibungsfreiheit eine Rolle<sup>258</sup>, in Einsteins berühmtem Gedankenexperiment, bei dem er mit Lichtgeschwindigkeit auf einem Lichtstrahl „reitet“, die Möglichkeit, sich mit Lichtgeschwindigkeit fortzubewegen, etc. Man kann von den Hintergrundbedingungen also nicht verlangen, dass sie in einem umfassenden Sinne „möglich“ sind, sondern muss diese Forderung auf einen „relevanten“ Sinn einschränken. Was ist aber relevant? Wilkes verspricht an mehreren Stellen, eine genaue Explikation ihres Relevanzbegriffs zu liefern, unterlässt dies allerdings. Die Beispiele, die sie zur Explikation angibt, sind allesamt Fehlinterpretationen der jeweiligen wissenschaftshistorischen Episode und auch nicht erhellend. So sei Stevins idealisierende Annahme, dass die schiefe Ebene reibungsfrei sei, irrelevant, während Einsteins (implizite) Annahme im Clock-in-the-box-Gedankenexperiment (das wir in Kapitel 2 kennen gelernt haben), dass die Uhr unabhängig von ihrer Bewegung durch das Gravitationsfeld der Erde eine genaue Zeitmessung zuließe, relevant und für das Gedankenexperiment vernichtend sei. Warum das so ist, kann Wilkes natürlich nicht beantworten. Im Gegenteil scheinen ihre Beispiele gerade gegen sie zu sprechen:

Bei Stevin geht es um die Prinzipien der Mechanik, weshalb seine Argumentation, falls sie beinhalten würde, dass auch unter der Annahme der Reibungsfreiheit eine fortwährende Bewegung unmöglich sei, höchst problematisch wäre. Hätte Stevin argumentiert, dass sich die Kette nicht von selbst in Bewegung setzen kann, weil sie sonst ein Perpetuum mobile der zweiten Art wäre, also zur Energiegewinnung genutzt werden könnte, wäre seine Argumentation zwar schlüssig, aber weniger anschaulich. Stevins (angebliche) Hintergrundannahme der Reibungsfreiheit wäre deshalb *gerade* relevant für sein Gedankenexperiment.<sup>259</sup>

<sup>256</sup> Vgl. auch Ward [336] für einen ähnlichen Einwand.

<sup>257</sup> Wer glaubt, dass es in der Wirklichkeit Widersprüche gibt, kann sich dadurch beruhigen, dass dann in der stipulierten Geschichte die Widersprüche an harmloser Stelle auftreten.

<sup>258</sup> Das wird zwar oft behauptet, *de facto* findet sich diese Annahme aber nicht in Stevins Argumentation. Vgl. auch Kühne [175].

<sup>259</sup> Vgl. die Bemerkung der Herausgeber in Stevin [316], 179, FN 1.

Bei Einsteins Gedankenexperiment ist hingegen ungeklärt, welche Verbindung zwischen der allgemeinen Relativitätstheorie und der Quantenmechanik besteht. Dementsprechend unklar ist auch, in welchem Sinne Einsteins Idealisierung von Relevanz für den Ausgang des Gedankenexperimentes ist (was noch mal völlig unabhängig von der Frage ist, ob Bohrs „Rettung“ überhaupt funktioniert, vgl. Kapitel 2). Die beiden Beispiele aus der Physik sind also nicht hilfreich zum Verständnis des Relevanzkriteriums.

Wilkes Beispiele aus der Philosophie helfen auch nicht weiter. So ist sie der Meinung, dass Gedankenexperimente in der Philosophie bezüglich Gold und Wasser erfolgreich sein können, Gedankenexperimente bezüglich Personen aber nicht. Daraus entwickelt sie die weiter unten noch zu diskutierende Auffassung, dass Gedankenexperimente bezüglich natürlicher Arten im Allgemeinen erfolgreicher sein können als Gedankenexperimente über andere Dinge (wie z.B. Personen, die nach Wilkes keine natürliche Art bilden). Nimmt man ihre Beispiele als Beleg für diese These, muss man feststellen, dass Wilkes hier zwei Ebenen durcheinander bringt. Wann genau haben Philosophen eigentlich über Gold und Wasser Gedankenexperimente angestellt? Nachdem die Alchemie nicht mehr wirklich Hochkonjunktur hat, ist die Frage, was Gold und was Wasser ist, eine Frage der Chemie und einer philosophischen Behandlung entrückt. Kripke und Putnam haben das auch nie bestritten. In ihren Gedankenexperimenten geht es nicht um Wasser oder Gold, sondern um Bedeutungstheorien. Es geht um die Frage, ob die Bedeutungen aller unserer Ausdrücke im Fregeschen Sinne durch unsere mentalen Zustände festgelegt sind, oder ob nicht zumindest die Referenz mancher Ausdrücke (Eigennamen, Ausdrücke, die natürliche Arten bezeichnen, etc.) durch extern bestehende Kausalketten gegeben ist. Die Gedankenexperimente zu ‚Wasser‘ und ‚Gold‘ gehen nicht über Wasser und Gold, sondern behandeln ein Problem der Bedeutungstheorie. Die Gedankenexperimente zu ‚Person‘ gehen in der Tat darüber, was Personen sind. Sollte die Unterscheidung von natürlichen und nicht-natürlichen Arten für den Erfolg des Gedankenexperimentierens ausschlaggebend sein, dann sollte die Frage eher lauten, ob ‚Bedeutung‘ eine natürliche Art bezeichnet und sich Putnams und Kripkes Gedankenexperimente *deshalb* von den Gedankenexperimenten zum Personenbegriff unterscheiden. Dies ist noch einmal unabhängig davon, ob man Wilkes anderen Annahmen Glauben schenkt. Beispielsweise der Annahme, dass Gedankenexperimente zum Personenbegriff weniger kontrovers sind als die Gedankenexperimente von Putnam und Kripke (und ob dem – falls dem so ist – zu Recht so ist), und der Annahme, dass ‚Person‘ keine natürliche Art bezeichnet, etc.

Auch in ihren anderen Beispielen scheint Wilkes den eigentlichen Streitpunkt der jeweiligen Debatte nicht zu sehen. So moniert sie an Platons „Ring des Gyges“-Gedankenexperiment, dass unterspezifiziert bliebe, wie man sich die genaue Wirkungsweise des Ringes vorzustellen hätte. Wäre man nur unsichtbar, oder auch unberührbar (und könnte man deshalb einem Wachmann versehentlich in die Arme laufen)? Könnte man in diesem unsichtbaren Zustand Gegenstände wegtragen? Es ist aber *völlig klar*, wie diese Fragen zu beantworten sind, und *völlig*

*unklar*, warum sich hinter den Antworten Probleme für das Gedankenexperiment verbergen sollten.<sup>260</sup>

Entsprechendes gilt für ihre Kritik an Gedankenexperimenten zum Personenbegriff. Zunächst wiederholt sie Fodors Fehler, die Frage, wie wir eine mögliche Welt beschreiben würden, mit der Frage zu verwechseln, wie die Bewohner dieser möglichen Welt ihr Begriffssystem an die Umstände in ihrer Welt angepasst hätten (Wilkes [345], 11). Aber selbst wenn man ihr zugesteht, dass die Umstände, die im Gedankenexperiment geändert werden, keine relevanten Unmöglichkeiten darstellen dürfen, scheint sie nicht verstanden zu haben, was in den jeweiligen Debatten überhaupt von Relevanz ist:

The chief trouble with Wilkes' [...] criticism is that it is far from obvious that the impossibilities she so ably exhibits in many thought experiments about the mind really are relevant to them. Relevance is relative to purpose. The purpose of many philosophical thought experiments appears, at least *prima facie*, to be to investigate "conceptual" theses by studying counterfactual situations. Is it essential to this purpose that those situations be theoretically possible? (Häggqvist [137], 30)

Auch Häggqvist kommt zu dem Ergebnis, dass die „theoretische Unmöglichkeit“ der Hintergrundannahmen bei philosophischen Gedankenexperimenten häufig völlig irrelevant ist. So sind Teletransporter vermutlich theoretisch unmöglich in Wilkes' Sinne. Anzunehmen, dass ein Gedankenexperiment zu diachroner Personenidentität unschlussig ist, wenn Teletransporter in ihm eine prominente Rolle spielen, ist aber bestenfalls unmotiviert (Häggqvist [137], 30).

Sieht man einmal davon ab, dass Wilkes die einzelnen Beispiele eigenwillig interpretiert, fallen noch zwei weitere Irrtümer auf. Zum einen eine Fehlrekonstruktion der Dialektik von Gedankenexperimenten, zum anderen eine Fehldiagnose zum Verhältnis von natürlichen Arten und philosophischen Forschungsmethoden.

Wilkes ist der Meinung, dass die Frage, welche Umstände in einem Gedankenexperiment von Relevanz sind, zum Teil mit Hilfe unserer Theorien über den Gegenstandsbereich beantwortet werden kann. Ist der Gegenstand eines (philosophischen) Gedankenexperiments eine natürliche Art, dann verfügen wir (bei Wilkes

---

<sup>260</sup> Das Gedankenexperiment funktioniert vielleicht zum Teil deswegen, *weil es* in diesen Hinsichten unterspezifiziert ist. Es soll die Intuition abgefragt werden, ob man auch dann noch moralisch handeln würde, wenn unmoralisches Handeln keine negativen Konsequenzen irgendwelcher Art nach sich ziehen würde, (beispielsweise) wenn man unter garantierter Anonymität handeln würde. Die *erwartete* Antwort ist ja, dass man in einem solchen Fall *nicht* moralisch handeln würde. Man würde die Details also zugunsten garantierter Anonymität ausfüllen. Wilkes hat kein Argument, weshalb das nicht gehen sollte, und auch keinen Hinweis, dass die erwartete Reaktion irgendwann umschlagen müsste. Aber selbst wenn es nicht ginge, weil – meinetwegen – garantierte Anonymität mit der Möglichkeit persönlicher Bereicherung konfligiert, ist das *unreflektierte* Urteil, das (naiv) von der Kompatibilität ausgeht und besagt, dass man nur extrinsisch zum moralischen Handeln motiviert ist, trotzdem noch wahr (auch wenn man feststellt, dass man sich aus praktischen Gründen immer in Situationen befindet, in denen es eine solche extrinsische Motivation gibt).

*per definitionem*) auch über (naturwissenschaftliche) Theorien über diesen Gegenstandsbereich, die uns darüber informieren, was von Relevanz ist:

‘Natural kinds’ is an expression which has often come under fire—indicted for vagueness, obscurity, hand-waving. This is a mistake. Whatever the difficulties, for any and all science, of homing in on ‘true’ natural kinds; whatever the difficulties in saying which of them are ‘fundamental’ and which derivative; even if the whole notion of a natural kind is shorthand for the idea of a class of phenomena which more advanced scientific theories will explain in due course in more fundamental terms; despite all that, they serve a useful purpose, even if only to pick out groups of things which science finds useful, profitable, convenient to isolate. They are useful because natural-kind terms provide, in the main, the central explananda and explanata for systematic study: they are the terms for which, and with which, the laws and generalizations of science are framed. Hence ‘water’, ‘mass’, and ‘tiger’ are natural kind terms; ‘fence’, ‘ashtray’, and ‘ornament’ are not. (Wilkes [345], 13-14)

Natürliche Arten sind also diejenigen Dinge, derethalben wir Wissenschaft betreiben und die deshalb auch bereits in vielen gesetzesartigen Verallgemeinerungen Berücksichtigung gefunden haben. Ist der Gegenstand einer wissenschaftlichen Untersuchung *keine* natürliche Art, haben wir entsprechend *keine* wissenschaftlichen Theorien und dementsprechend kein gesichertes Wissen darüber, was für die Hintergrundannahmen im Gedankenexperiment zulässig ist und was nicht.

So we should look rather to the ‘theoretical’, or ‘in principle’ possibility of the *relevant* background conditions – the conditions we need to specify before we can be sure both that the imagined scenario is adequately described, and that the inference from the imagined state to the conclusion can be made. This would be the test of validity for a thought experiment. This we can characterize as a matter of what could or could not happen given our backing scientific knowledge: what our theories allow to be possible or not. So, for example, gold could not, possibly, have had a different atomic number. Again, iron bars cannot, possibly, float on water – nothing which is iron, i.e. which has a specific gravity within the range 7.3-7.8, can float, i.e. have also a specific gravity of less than 1. Once we describe the situation adequately, once we hand ourselves the backing theory of metals, then the (relevant) impossibilities appear – indeed, in the case of the iron bars the theory turns the impossibility into a logical one: nothing can both have and not have a specific gravity of less than 1. Similarly, [...] theory rules out firmly the possibility of carnivorous rabbits on Mars. (Wilkes [345], 18)

Bei natürlichen Arten ist es demnach *erstens* so, dass es metaphysisch determiniert ist, was ihnen in anderen möglichen Welten zustoßen kann und was nicht, und *zweitens* ist es so, dass uns diese metaphysischen Tatsachen auch durch unsere besten Theorien zugänglich sind.<sup>261</sup> Bei nicht-natürlichen Arten kann es sein, dass uns Dinge möglich scheinen, die nicht möglich sind, weil es entweder nicht bestimmt

---

<sup>261</sup> Wilkes Begriff von „theoretischer Möglichkeit“ geht dabei zurück auf Seddons Diskussion „logischer Möglichkeit“ in Seddon [295].

ist, ob es möglich ist oder nicht, oder wir keine Theorien haben, die uns sagen, ob es möglich ist oder nicht.

Lassen wir die Frage der metaphysischen Determiniertheit von Möglichkeiten einmal beiseite. Ohnehin scheint es so, dass wenn etwas nicht in der einen oder anderen Weise determiniert ist, metaphysisch die Dinge zugunsten der Möglichkeit ausgehen. (Angenommen es wäre metaphysisch nicht determiniert, ob ich mich in Düsseldorf aufhalte oder in Neuss, dann ist es sowohl *möglich*, dass ich mich in Düsseldorf aufhalte, als auch möglich, dass ich mich in Neuss aufhalte.)

Nehmen wir also an, es gäbe (beispielsweise darüber, ob Personen einen Teletransport überleben können) etwas zu wissen, aber aufgrund der Tatsache, dass wir keine (wissenschaftlich gut bestätigten) Theorien darüber haben, keine Grundlage zu urteilen. Was uns laut Wilkes fehlt, ist offenbar eine gut bestätigte Theorie über Personen (nicht unbedingt eine über Teletransporter). Die Frage, ob Wasser auch die Molekülstruktur XYZ haben kann, kann mit dem Hinweis beantwortet werden, dass unsere beste Theorie über Wasser besagt, dass es sich um  $H_2O$  handelt, Wasser also *nicht* XYZ sein kann. Die Frage, ob eine Person z.B. einen Teletransport überleben kann, könnte man negativ beantworten, verfügte man über eine Theorie, die besagt, dass Personenidentität Körperidentität voraussetzt, Teletransport aber Körperidentität verletzt (nach Stipulation im Gedankenexperiment), Personen einen Teletransport also nicht überleben können. Leider haben wir keine solche Theorie, die uns über Personen und die für sie relevanten möglichen Umstände aufklären könnte, weshalb sich Gedankenexperimentieren mit Personen verbietet. Ohne eine solche Theorie sind Spekulationen darüber, wie es ist Personen zu teletransportieren eben nur *Science Fiction*.

Eine solche Rekonstruktion stellt die Dialektik von Gedankenexperimenten aber auf den Kopf. Angenommen, wir wollten tatsächlich eine These (T) wie ‚Wasser ist  $H_2O$ ‘ mit einem Gedankenexperiment, in dem eine durchsichtige, trinkbare, geruchs- und geschmacksneutrale Flüssigkeit die Molekülstruktur XYZ hat, testen und fragen uns nun, ob es möglich ist, dass Wasser XYZ ist. Natürlich könnten wir jetzt T heranziehen und feststellen, dass es nicht möglich ist, dass Wasser XYZ ist, wenn T wahr ist. Das wussten wir vorher ja auch schon, sonst hätten wir das Gedankenexperiment nicht entwickelt. T steht ja zur Diskussion, weshalb es eine *petitio principii* wäre, mit der Wahrheit von T zu argumentieren. T sollte also nicht zum Hintergrundwissen gehören, wenn es darum geht, zu beurteilen, ob Wasser XYZ sein kann. Subtrahieren wir T aber von unserem Hintergrundwissen, dann kann man folgende Fälle unterscheiden: (i.) T folgt aus unserem Hintergrundwissen, (ii.) T ist logisch unabhängig von unserem Hintergrundwissen, (iii.)  $\neg T$  folgt aus unserem Hintergrundwissen. Wenn Letzteres der Fall ist, werden wir uns eine mögliche Welt vorstellen können, in der Wasser XYZ ist (oder zumindest irgendeinen Umstand, in dem Wasser nicht  $H_2O$  ist). Daraus folgt aber nicht, dass T falsch ist, denn es kann genauso gut sein, dass T von unserem Hintergrundwissen logisch unabhängig ist (Fall (ii.)). In diesem Fall ist es *trivial*, dass es eine logische Möglichkeit gibt, die mit unserem Hintergrundwissen

vereinbar ist und mit T unvereinbar, besagt daher nichts über die Wahrheit von T.<sup>262</sup> Erst wenn T aus unserem Hintergrundwissen bereits folgt, können wir in der Lage sein, festzustellen, dass es nicht möglich ist, dass Wasser XYZ ist. Nur in diesem Fall hilft uns unser Hintergrundwissen bei der Beurteilung der Möglichkeit also überhaupt weiter, und das ist weder der Fall eines gelungenen Gedankenexperiments (in diesem Fall wissen wir vielmehr nur, dass das Gedankenexperiment eine – im „relevanten“ Sinne – unmögliche Welt beschreibt), noch jeder Fall eines misslungenen Gedankenexperiments, da es ja den Fall offen lässt, dass T von unserem Hintergrundwissen unabhängig und trotzdem wahr ist (vgl. auch das nächste Kapitel, sowie Cohnitz [69], Cohnitz [72]).

Die Logik des Gedankenexperiments ruht nicht hauptsächlich darauf, dass wir explizites Hintergrundwissen aktivieren, um andere explizite Theorien zu testen. Wir testen vielmehr *explizite* Theorien gegen *implizites* Wissen. Das ist die Auffassung, die wir bisher verteidigen konnten, und die seit Ernst Machs erster Beschäftigung mit diesem Thema vorherrschend ist (vgl. Kapitel 2).<sup>263</sup>

Diese Auffassung von Gedankenexperimenten ist schwer vereinbar mit Wilkes Auffassung von natürlichen Arten. Laut Wilkes funktionieren Gedankenexperimente besonders gut, wenn es um Theorien geht, die über natürliche Arten sind. Akzeptiert man – wie Wilkes – Kripkes Analyse metaphysischer Notwendigkeiten als Notwendigkeiten, die a posteriori wahr sein können und zu denen alle Identitätsaussagen (und Subsumtionen) zwischen natürlichen Artausdrücken gehören, kann man diese Auffassung mit unserer Auffassung über Gedankenexperimente nur schwer vereinbaren. Dass Wasser identisch mit H<sub>2</sub>O ist, Wärme identisch mit der Bewegung von Molekülen, oder Katzen Säugetiere sind, sind keine Folgerungen, die wir aus unseren a priori-Überzeugungen gezogen hätten. Nach Kripke handelt es sich um metaphysische Notwendigkeiten *a posteriori*. Diese zum Gegenstand von Lehnstuhl-Experimenten zu machen, scheint besonders problematisch. Wenn ich nicht weiß, dass Wasser H<sub>2</sub>O ist, kann ich mir natürlich vorstellen, dass es XYZ ist. Es scheint mir daher auch (in dieser hypothetischen epistemischen Situation) möglich, dass Wasser eine andere Molekülstruktur hat als H<sub>2</sub>O. Falls Wasser H<sub>2</sub>O ist, ist es – laut Kripke – notwendigerweise H<sub>2</sub>O. In meiner hypothetischen epistemischen Situation haben wir aber gerade festgestellt, dass es XYZ sein kann. Folglich ist Wasser nicht H<sub>2</sub>O. Gedankenexperimente sind – im Gegensatz zu Wilkes Auffassung – nicht auf besondere Weise zuverlässig, wenn es sich um natürliche Arten handelt, sondern offenbar auf besondere Weise unzuverlässig.

---

<sup>262</sup> Das folgt aus der modelltheoretischen Interpretation logischer Unabhängigkeit und der üblichen modalen Interpretation der Modelltheorie. Demnach sind zwei Sätze p und q logisch unabhängig, wenn es eine mögliche Welt gibt, in der p wahr ist und q falsch ist und umgekehrt. Wenn die Konjunktion des Hintergrundwissens K von T logisch unabhängig ist, gibt es eine mögliche Welt, die K erfüllt und T falsifiziert. (Der Fall, dass K inkonsistent ist, wird hier nicht behandelt.)

<sup>263</sup> Vgl. zu dieser Auffassung auch Myers [217], Myers [218], Myers [219].

## 7.1.2 GEDANKENEXPERIMENTE SIND UNZUVERLÄSSIG, WENN ES UM EINEN BEGRIFF GEHT, DER EINE NATÜRLICHE ART BEZEICHNET

Natürlich haben wir noch nicht gezeigt, dass philosophische Gedankenexperimente unproblematisch sind, wenn sie keine natürlichen Arten zum Gegenstand haben, sondern bisher nur herausbekommen, dass Wilkes Gebrauch von Kripkes Auffassung natürlicher Arten nicht besonders einleuchtend ist. Gedankenexperimente bezüglich natürlicher Arten sind – nach allem, was wir bisher über natürliche Arten erfahren haben – eigentlich ein sehr zweifelhaftes Forschungsinstrument. Wir wollen diese Vermutung nun etwas näher betrachten.

In *Sameness and Substance* (1980) verteidigt David Wiggins eine Neo-Lockeanische Auffassung, derzufolge diachrone personale Identität in einem kontinuierlichen Bewusstseinsstrom besteht.<sup>264</sup> Gegen diese Auffassung ist mit Gedankenexperimenten argumentiert worden. Was, wenn sich ein Bewusstseinsstrom teilen könnte? Angenommen, nach einem bestimmten Zeitpunkt  $t_2$ , gäbe es zwei Bewusstseinsströme  $b_1$  und  $b_2$ , für die gilt  $b_1 \neq b_2$ , die aber beide mit dem Bewusstseinsstrom  $b_0$  zu  $t_1$  kontinuierlich sind (vgl. Abbildung 7.1-1). Zu  $t_3$  gibt es dann auch zwei verschiedene Personen,  $P_1$  und  $P_2$ , die jeweils mit einem der beiden Ströme assoziiert sind. Nach Wiggins sollte die Kontinuität zwischen  $b_0$  und  $b_1$  bzw.  $b_2$  hinreichend für personale Identität sein, d.h., dass die Person  $P_0$ , die mit  $b_0$  assoziiert ist, auch mit  $P_1$  und  $P_2$  identisch sein müsste. Da Identität eine transitive Relation ist, müssten aber auch  $P_1$  und  $P_2$  miteinander identisch sein, was offenbar nicht der Fall ist.

Die Analysans-Relation (‘ist psychisch kontinuierlich mit’) hat also andere Eigenschaften als die Analysandum-Relation (‘ist diachron identisch mit’), vorausgesetzt, dass sich Bewusstseinsströme in der beschriebenen Weise teilen können.

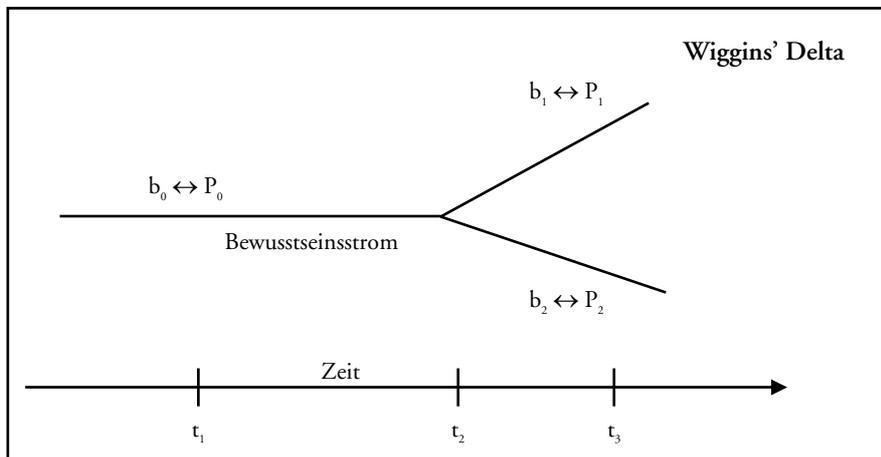


Abbildung 7.1-1

<sup>264</sup> Vgl. hierzu Noonan [236], French [111], Kitcher [164], Beck [17], Wiggins [342].

Können sie das? Wiggins ist der Auffassung, dass, falls es sich bei ‚Person‘ um einen natürlichen-Art-Ausdruck handeln würde, diese Möglichkeit ausgeschlossen werden könnte:

[A]ny would-be determination of a natural kind stands or falls with the existence of lawlike principles that will collect together the actual extension of the kind around an arbitrary good specimen of it; and [...] these lawlike principles will also determine the characteristic development and typical history of members of his extension. [...] If that is supposed, then the first thing we can expect is that the sense of the sortal predicate ‚person‘ will exempt us from counting as genuinely conceivable any narrative in which persons undergo changes that violate the lawlike regularities constituting the actual nomological foundation for the delimitation of the kind we denominate as that of persons. (Wiggins [343], 169-170)

Wiggins argumentiert schließlich dafür, dass ‚Person‘ zwar keine natürliche Art bezeichnet, dass aber jede Person zu einer natürlichen Art gehört, für die jeweils die Möglichkeit einer Verzweigung des Bewusstseinstroms naturgesetzlich ausgeschlossen ist<sup>265</sup>, weshalb oben beschriebener Fall begrifflich unmöglich sei.

Hierbei von „begrifflicher Möglichkeit“ zu sprechen, ist verwirrend, zumindest solange man an der analytisch/synthetisch-Unterscheidung festhalten möchte. Stattdessen trifft man häufig auf die Auffassung, dass es zwischen „begrifflicher“ und „metaphysischer“ Notwendigkeit einen Unterschied gibt, und die von Wiggins diskutierte Modalität die der metaphysischen Notwendigkeit sei. Nach dieser Auffassung kann es sehr wohl sein, dass unser Personen**begriff** es erlaubt, dass Personen Transformationen überleben, die durch gesetzesartige Regularitäten für aktuelle Personen, also die Referenten des Personenbegriffs, eigentlich metaphysisch ausgeschlossen sind. Eine solche *metaphysische* Unmöglichkeit impliziert aber eben keine *begriffliche* Unmöglichkeit, wie wir an der Zwiebel der Möglichkeiten (Abb. 6.1-2) sehen konnten: die Menge der metaphysischen Möglichkeiten könnte eine echte Teilmenge der Menge der begrifflichen Möglichkeiten sein.

Wie auch immer, solange wir uns für *Personen* interessieren, und die Bedingung von *deren* transtemporaler Identität, anstatt für unseren Personen**begriff** und unseren *Begriff* transtemporaler Identität, macht Wiggins Einwand natürlich Sinn. Was Personen zu überleben im Stande sind, kann vielleicht nicht ohne weiteres a priori auf der bloßen Basis unserer sprachlichen Intuitionen herausgefunden werden, wenn ‚Person‘ eine natürliche Art bezeichnet. Wie wir in Kapitel 6.2 gesehen haben besteht der Unterschied zwischen empirischen Analysen und Bedeutungsanalysen ja gerade darin, dass Erstere über die Phänomene reden, nicht über die Ausdrücke, die diese Phänomene bezeichnen.

---

<sup>265</sup> Wiggins [342], Wiggins [343], Wiggins [344] Vgl. auch Beck [17] für eine Kritik an dieser Argumentation Wiggins'. Wilkes' und Nagels Berichte von Komissurotomie-Patienten legen außerdem nahe, dass die Auffassung, dass es naturgesetzlich ausgeschlossen ist, dass ein Bewusstseinsstrom sich teilen kann, empirisch falsch ist, vgl. Wilkes [345], Nagel [220].

## 7.2 DAS APRIORI UND DAS METAPHYSISCHE

Inwiefern wir dazu in der Lage sind, Sachanalysen a priori in Bezug auf ihre Wahrheit oder Adäquatheit zu beurteilen, ist innerhalb der Philosophie – wie wir noch sehen werden – sehr kontrovers. Die Sachanalysen, die dabei im Zentrum der Kontroverse stehen, sind insbesondere solche, die den Status so genannter „metaphysischer Notwendigkeiten“ besitzen. Was metaphysische Notwendigkeiten genau sind, ist umstritten. Es gibt dabei mindestens die folgenden Positionen:

(1) ‚Metaphysische Notwendigkeit‘ bezeichnet eigentlich Naturnotwendigkeit oder nomische Notwendigkeit. Etwas ist metaphysisch notwendig gdw. es aus den Naturgesetzen folgt, bzw. in allen naturmöglichen Welten wahr ist. Etwas ist „metaphysisch möglich“, wenn es mit den Naturgesetzen vereinbar ist, bzw. in mindestens einer möglichen Welt wahr ist, die in der Menge der naturmöglichen Welten ist. (Wenn man außerdem eine analytisch/synthetisch-Unterscheidung vertritt, ist die Menge der naturmöglichen Welten eine echte Teilmenge der möglichen Welten.)

(2) ‚Metaphysische Notwendigkeit‘ bezeichnet wie ‚nomische Notwendigkeit‘ eine Erweiterung der begrifflichen Notwendigkeit um zusätzliche „metaphysische Regularitäten“ (wie etwa „Notwendigkeit der Identität“, „Notwendigkeit der Herkunft“<sup>266</sup>). Etwas ist dann *metaphysisch notwendig*, wenn es in all den möglichen Welten der Fall ist, die Elemente der echten Teilmenge möglicher Welten sind, die durch begriffliche Notwendigkeiten und metaphysische Regularitäten beschränkt ist (vgl. Barwise [11]).

(3) ‚Metaphysische Notwendigkeit‘ bezeichnet wie ‚nomische Notwendigkeit‘ eine Erweiterung der begrifflichen Notwendigkeiten, allerdings nur um Informationen darüber, worauf sich bestimmte Ausdrücke der Sprache in der aktualen Welt beziehen.<sup>267</sup> Die metaphysisch notwendigen Welten sind also (wie in (2) bzw. (1)) eine echte Teilmenge der möglichen Welten.

(4) ‚Metaphysische Notwendigkeit‘ bezeichnet Notwendigkeit im weitesten Sinne. Metaphysische Notwendigkeiten sind in allen möglichen Welten wahr, weil es nur metaphysisch mögliche Welten gibt.<sup>268</sup> Andere Notwendigkeitsbegriffe (wie etwa „nomische Notwendigkeit“) sondern dann jeweils Teilmengen aus der Menge der metaphysisch möglichen Welten aus.

Die Zwiebel der Notwendigkeiten aus Kapitel 6.1 entspricht in *jedem* Fall der *epistemischen* Reihenfolge der verschiedenen Notwendigkeiten in Bezug auf *Sätze*. Wenn wir nur Informationen über die logische Form von Sätzen haben, können wir nur beurteilen, welche Sätze logisch notwendig sind. Wenn wir darüber hinaus

<sup>266</sup> Zur Notwendigkeit der Herkunft („necessity of origin“) und weitere „metaphysische Prinzipien“ dieser Art, vgl. Kripke [172], Rohrbaugh und deRosset [282].

<sup>267</sup> Vgl. Sidelle [297], Hale [139].

<sup>268</sup> Vgl. Stalnaker [313], Kripke [172], Lewis [187], Jackson [158], Chalmers [55].

Informationen über bestimmte Bedeutungspostulate haben, können wir beurteilen, welche Sätze begrifflich notwendig sind. Wenn wir darüber hinaus Wissen davon haben, worauf sich – beispielsweise – die rigide designierenden Ausdrücke unserer Sprache beziehen, haben wir Zugang zu metaphysischen Notwendigkeiten, etc.

Die Zwiebel der Notwendigkeiten entspricht aber nicht unter allen vier Interpretationen von ‚metaphysisch notwendig‘ der *ontologischen* Reihenfolge. Zumindest Auffassung (4) scheint mit einer ontologischen Interpretation der Zwiebel-Idee unvereinbar.

Das liegt daran, dass Auffassung (4) notwendig wahre *Sätze* als Sätze betrachtet, die notwendig wahre *Propositionen* ausdrücken (vgl. Yablo [358], Stalnaker [313]). Die Notwendigkeit von Propositionen ist aber keine Frage von Konventionen, Begriffen oder Bedeutungen. Ob eine *Proposition* notwendig ist, ist eine Frage danach, ob sie in allen möglichen Welten der Fall ist:

It is not just that the metaphysical laws rule out the possibility that gold be something other than an element with atomic number 79 – it is that if we think clearly (in light of empirical facts) about what it would be for there to be gold that was not such an element, we see that there is no such possibility for any laws to rule out. (Stalnaker [313])

Entsprechend beziehen sich auch „bloß logische“ Möglichkeiten auf metaphysische Möglichkeiten. Diese Auffassung erlaubt entsprechend auch keine konventionalistische Reduktion der Modalitäten, die die Notwendigkeit bestimmter *Sätze* auf Bedeutungspostulate o.Ä. alleine zurückführen würde (vgl. Yablo [358], Sidelle [297]).

Da es in dieser Untersuchung darum geht, eine Methodologie des Gedankenexperiments zu entwickeln, die möglichst unabhängig von spezifischen Geltungsansprüchen und Erkenntnisansprüchen ist, müssen wir die Möglichkeit der Existenz metaphysischer Notwendigkeiten zunächst in Erwägung ziehen und die verschiedenen Standpunkte in Bezug auf die Beziehung metaphysischer Notwendigkeiten zu anderen Notwendigkeiten zur Kenntnis nehmen. In diesem Kapitel werden daher im Detail drei verschiedene Auffassungen diskutiert, die diese Beziehung und die Relevanz unserer Intuitionen für die Beurteilung metaphysischer Zusammenhänge unterschiedlich beurteilen, und betrachtet, welche Konsequenzen die jeweilige Auffassung für die Methode des Gedankenexperiments in der Philosophie hat. Diese Ansätze bezeichnen wir als ‚Modaler Rationalismus‘, ‚Putnam Orthodoxie‘ und ‚Moderaten Deskriptivismus‘. Wie wir sehen werden, spielt die ontologische Interpretation von „metaphysischer Notwendigkeit“ in den unterschiedlichen Standpunkten eine geringere Rolle, als die epistemischen Beziehungen zwischen verschiedene Arten von Notwendigkeit. Der „Modale Rationalismus“ und die „Putnam Orthodoxie“ haben erkenntnistheoretisch gesehen die größten Differenzen, wobei sie in der ontologischen Interpretation von „metaphysischer Notwendigkeit“ genau übereinstimmen (nämlich im Sinne von (4)).

## 7.2.1 MODALER RATIONALISMUS

In letzter Zeit hat – insbesondere durch Frank Jackson und David Chalmers – eine Auffassung von der Rolle begrifflicher Analysen wieder an Einfluss gewonnen, die man eigentlich nach Quines ‚Two Dogmas of Empiricism‘ (Quine [269]), spätestens aber nach Kripkes *Naming and Necessity* (Kripke [172]) für erledigt gehalten hat. Ich nenne diese Auffassung (im Anschluss an Chalmers [57]) ‚modalen Rationalismus‘, da sie – wie der erkenntnistheoretische Rationalismus von Descartes – die These vertritt, dass wir substantielle metaphysische Wahrheiten a priori aus der Vernunft erkennen können.<sup>269</sup>

Nach Auffassung des modalen Rationalismus besteht die Aufgabe der Metaphysik zunächst darin, das „location problem“ zu lösen, d.h., die Frage zu beantworten, wie eine wahre Beschreibung der Welt in einem bestimmten Vokabular durch eine Beschreibung in einem angeblich fundamentalen Vokabular adäquat wiedergegeben werden kann. Das metaphysische Problem des Physikalismus – um auf dieses Beispiel zurück zu kommen – ist z.B. das Problem, wie eine wahre Beschreibung eines Ereignisses in (alltags-)psychologischer Ausdrucksweise durch eine Beschreibung adäquat wiedergegeben werden kann, die nur von physikalischem Vokabular Gebrauch macht. Jackson spricht davon, dass zu erklären ist, wie die (angeblich) „fundamentalere“ Beschreibung die weniger fundamentale „wahr macht“. Eine Beschreibung „macht“ dabei eine andere „wahr“, wenn die weniger fundamentale in der fundamentalen „enthalten“ ist. Die weniger fundamentale Beschreibung ist dann „reduktiv erklärbar“ durch die fundamentalere Beschreibung und entsprechende a priori Deduktionen. Die Hauptmethode, Metaphysik zu betreiben, sei demnach Begriffsanalyse: „the very business of addressing when and whether a story told in one vocabulary is made true by one told in some allegedly more fundamental vocabulary“ (Jackson [158], 28). Wir werden dies nun zu erläutern versuchen.

Nach Jackson besteht die Aufgabe der Metaphysik (und darunter fallen für Jackson alle ‚Was ist‘-Fragen, also ein Großteil dessen, was man traditioneller Weise für den Gegenstand analytischer Philosophie hält) darin anzugeben, was es gibt, und wie das, was es gibt, beschaffen ist.<sup>270</sup> Philosophen kommen dieser Aufgabe nicht dadurch nach, dass sie einfach alles, was in einem bestimmten Gegenstandsbereich (oder dem gesamten Universum) vorkommt, auflisten, sondern sie lösen diese Aufgabe, indem sie versuchen, genau die Menge von Basisbestandteilen zu finden, aus der sich alles im jeweiligen Gegenstandsbereich (oder dem gesamten Universum) aufbauen<sup>271</sup> lässt.

By serious metaphysics, I mean metaphysics [...] that acknowledges that we can do better than draw up big lists, that seeks comprehension in terms of a more or less

<sup>269</sup> Dieser Zusammenhang zwischen Vernunftkenntnis und Metaphysik wird (auch unter diesem Label) insbesondere in Chalmers [56] diskutiert.

<sup>270</sup> Wir werden diese Auffassung noch diskutieren.

<sup>271</sup> „Aufbauen“ im Sinne von Carnaps *Aufbau*, also im Sinne einer logischen Konstitution aus einem bestimmten Basisvokabular.

limited number of ingredients, or anyway a smaller list than we started with. How big a list of basic ingredients we need, and even whether there is an ur-set, are matters open to debate; what seems to me obvious is that we can set some limits on what we need [...] and serious metaphysics is the investigation of whether these limits should be set. Thus, by its very nature, serious metaphysics continually faces the location problem. (Jackson [158], 5)

Metaphysische Positionen sind nach dieser Auffassung Supervenienzbehauptungen, also Behauptungen darüber, inwiefern bestimmte Eigenschaften von anderen Eigenschaften notwendig abhängen. So sei die Position des Physikalismus, demzufolge alles „eigentlich“ physikalisch ist, die folgende:

(P<sub>1</sub>) Jedes minimale physikalische Duplikat unserer Welt ist ein Duplikat *simpliciter* unserer Welt.

Was heißt das? Eine *Supervenienzbehauptung* dieser Art quantifiziert zunächst über mögliche Welten, die der aktuellen Welt in bestimmten Hinsichten ähnlich sein sollen. Bei (P<sub>1</sub>) geht es um Welten, die der unseren physikalisch gleichen, in denen also dieselben physikalischen Teilchen<sup>272</sup> mit denselben physikalischen Eigenschaften vorkommen und denselben physikalischen Gesetzen gehorchen. Dass es sich bei diesen möglichen Welten um „minimale“ Duplikate handeln soll, stellt sicher, dass nichts sonst in diesen Welten vorkommt. Ein Physikalismus, der zuließe, dass es neben den physikalischen Teilchen und Gesetzen in allen Duplikaten unserer Welt auch noch eine ordentliche Portion cartesischer *res cogitans* gäbe, wäre von einem Dualismus natürlich nicht zu unterscheiden. Die physikalistische Position zeichnet sich dadurch aus, dass sie behauptet, dass es – neben dem Physikalischen – *sonst nichts* bedarf, um ein Duplikat unserer Welt zu sein. ‚Simpliciter‘ weist darauf hin, dass dieser Physikalismus umfassend gemeint ist. *Alles* ist letztlich physikalisch. Physikalistische Positionen können natürlich auch bescheidener auftreten, wenn sie beispielsweise keine besondere Position hinsichtlich der Frage einnehmen wollen, ob auch abstrakte Objekte wie Zahlen physikalische Objekte sind, oder ob das, was wir für Wahrheiten über solche Objekte halten, keine Wahrheiten sind (oder nicht über irgendwelche Objekte), etc.

Ein bereichsbeschränkter Physikalismus könnte sich beispielsweise auf die Behauptung beschränken, dass alle semantischen Tatsachen physikalische Tatsachen sind, oder dass alle psychischen Tatsachen physikalische Tatsachen sind. Letzteres könnte etwa in folgender These wiedergegeben werden:

(P<sub>2</sub>) Jede Welt, die ein minimales physikalisches Duplikat unserer Welt ist, ist auch ein psychisches Duplikat unserer Welt.

---

<sup>272</sup> Wie dieser Physikalismus im Einzelnen auszubuchstabieren ist, soll nicht unser Problem sein. Wir lassen es hier offen, ob die „Teilchen“ auf einer bestimmten Mikro- oder Makroebene individuiert werden müssen, und ob diese Auffassung voraussetzt, dass die physikalische Welt durch eine konsistente Menge von Gesetzen durch alle Stufen der Körnigkeit der Betrachtung beschrieben werden kann.

Solche Supervenienzthesen sind auch dann wahr, wenn man eine eliminativistische Haltung gegenüber der „supervenierenden“ Ebene hat. Wenn man der Auffassung ist, dass es in unserer Welt schon keine psychischen oder semantischen Tatsachen gibt, ist man natürlich auch der Meinung, dass in einem rein physikalischen Duplikat unserer Welt solche nicht auf einmal auftauchen. Ist man allerdings kein Eliminativist in Bezug auf die supervenierende Ebene, dann verpflichtet die Supervenienzthese – laut Jackson – darauf, dass eine vollständige physikalische Beschreibung unserer Welt, alle Wahrheiten über die psychischen Tatsachen in unserer Welt „enthält“:

Let  $\Phi$  be the story as told in purely physical terms, which is true at the actual world and all the minimal physical duplicates of the actual world, and false elsewhere;  $\Phi$  is a hugely complex, purely physical account of our world. Let  $\Psi$  be any true sentence which is about the psychological nature of our world in the sense that it can only become false by things being different psychologically from the way they actually are: every world at which  $\Psi$  is false differs in some psychological way from our world. Intuitively, the idea is that  $\Psi$  counts as being about the psychological nature of our world because making it false requires supposing a change in the distribution of psychological properties and relations. Now, if  $[P_1]$  is true, every world at which  $\Phi$  is true is a duplicate simpliciter of our world, and so a fortiori a psychological duplicate of our world. But then every world at which  $\Phi$  is true is a world at which  $\Psi$  is true – that is,  $\Phi$  entails  $\Psi$ . (Jackson [158], 25)

Diese Auffassung vom „Enthaltensein“ der supervenierenden Tatsachen in den „fundamentaleren“, impliziert (zunächst) weder, dass die Wahrheit von  $P_1$  oder  $P_2$  a priori eingesehen werden kann, noch, dass die Begriffsanalyse bei der Beurteilung des Wahrheitswertes solcher Thesen von Nutzen sein könnte. Die fragliche Supervenienzbeziehung könnte einfach eine metaphysische Tatsache sein: alle  $\Phi$ -Welten *sind* eben  $\Psi$ -Welten, ohne dass dieser Zusammenhang a priori erkennbar wäre. Diese Option eröffnet sich insbesondere dann, wenn man eine ontologische Auffassung von metaphysischer Notwendigkeit hat, die in etwa der oben diskutierten Auffassung (4) entspricht. Wenn die für „Enthaltensein“ erforderliche Notwendigkeit nicht auf Bedeutungskonventionen (o.Ä.) *reduzierbar* ist, *kann* es durchaus sein, dass solche modalen Zusammenhänge von sprachlichen Zusammenhängen *unabhängig* sind und dann eventuell auch nicht a priori aufspürbar.

In einem verhältnismäßig unproblematischen Sinn verteidigt Jackson zunächst die Rolle der Begriffsanalyse als Voraussetzung für die Beurteilung der Wahrheit einer Supervenienzthese. Bei Supervenienzthesen dieser Art geht es offenbar darum, dass ein Bericht über einen Ereignisverlauf, der in einem Vokabular gegeben wird (beispielsweise in physikalischem Vokabular) einen Bericht „enthält“, der in einem anderen Vokabular gegeben werden kann (beispielsweise in dem der Psychologie oder Alltagspsychologie). Um überhaupt beurteilen zu können, ob eine Enthaltenseins-Beziehung vorliegt – und unabhängig davon, wie man dies festzustellen hat –, wird man zunächst sicherstellen müssen, dass man beide Vokabulare richtig versteht. Wie Ausdrücke in einem Vokabular gemeint sind, findet man am

besten heraus, indem man – als kompetenter Sprecher – imaginäre Fälle sprachlich beurteilt, also Gedankenexperimente<sup>273</sup> anstellt:

With all this behind us, we can state the rationale for conceptual analysis. Serious metaphysics requires us to address when matters described in one vocabulary are made true by matters described in another. But how could we possibly address this question in the absence of a consideration of when it is right to describe matters in the terms of the various vocabularies? And to do that is to reflect on which possible cases fall under which descriptions. And that in turn is to do conceptual analysis. Only that way do we define our subject – or rather, only that way do we define our subject we folk suppose is up for discussion. It is always open to us to stipulate the situations covered by the various descriptive terms, in which case we address subjects of our stipulation rather than the subjects the titles of our books and papers might naturally lead others to expect us to be addressing. (Jackson [158], 41-42)

Jackson argumentiert hier offenbar folgendermaßen: Wenn es bei „seriöser Metaphysik“ darum geht, zu klären, wann ein Bericht über einen Ereignisverlauf, der in einem Vokabular gegeben wird (beispielsweise in physikalischem Vokabular), einen Bericht „enthält“, der in einem anderen Vokabular gegeben ist, dann müssen dafür offenbar zunächst die beiden Vokabulare hinreichend verstanden worden sein. Ein Vokabular klärt man aber dadurch, dass man überlegt, wie man es auf aktuelle und mögliche Fälle anwendet. Dies ist nichts anderes als Begriffsanalyse. Folglich braucht man Begriffsanalyse, wenn man seriöse Metaphysik betreiben will. Ohne Begriffsanalyse kann man zwar auch ähnliche Untersuchungen betreiben (indem man die Bedeutungen der Vokabulare jeweils stipuliert), nur sind diese Untersuchungen dann in Gefahr das ursprüngliche Thema zu verfehlen (zumindest insofern das „ursprüngliche Thema“ in Ausdrücken unserer Umgangssprache formuliert war).

Diese Auffassung ist verhältnismäßig unproblematisch (wie wir noch zeigen werden), behauptet aber auch nur, dass die *Überprüfung* der Wahrheit einer Supervenienzthese eine Klärung der verwendeten Begriffe voraussetzt. Es müsse zunächst einmal geklärt werden, was auf der Ebene der mutmaßlichen Supervenienzbasis und auf der mutmaßlich supervenierenden Ebene jeweils genau behauptet wird, wenn man feststellen will, ob die eine Ebene auf der anderen superveniert.

Die Überprüfung der Frage, ob zwischen diesen beiden Ebenen nun eine Supervenienz-Beziehung vorliegt könnte nach allem, was wir bisher gesagt haben, dann auf andere Weise von statten gehen, ohne Begriffsanalyse und eben a posteriori.

---

<sup>273</sup> Jackson wehrt sich dagegen, philosophische „Gedankenexperimente“ mit wissenschaftlichen gleichzusetzen, da man bei wissenschaftlichen Gedankenexperimenten tatsächlich etwas über die Welt lerne, bei philosophischen Gedankenexperimenten aber nur etwas über unsere Wörter bzw. Begriffe. Falls Jackson darauf hinweisen möchte, dass es in Putnams und Kripkes Gedankenexperimenten nicht um Gold und Wasser geht, hat er mit dieser Auffassung (die wir oben ebenfalls vertreten haben) recht, falls nicht, übersieht er, dass Wörter und ihre Bedeutungen ebenfalls Phänomene dieser Welt sind. Vgl. Jackson [158], 78.

Angenommen es ginge um die Frage, ob Wasser-Tatsachen auf  $H_2O$ -Tatsachen supervenieren. Eine Begriffsklärung des Wasser-Vokabulars und des  $H_2O$ -Vokabulars könnte in einer genauen Angabe resultieren, wie wir jeweils feststellen, ob es sich bei einer Substanz um Wasser bzw.  $H_2O$  handelt. (Dann und nur dann, wenn  $x$  mehr oder weniger durchsichtig ist, den Durst löscht, bei Raumtemperatur flüssig ist, nach nichts schmeckt, ... und bei ca.  $100^\circ\text{C}$  kocht, ist  $x$  Wasser. Dann und nur dann, wenn  $x$  sich in im richtigen Verhältnis in Sauerstoff und Wasserstoff aufspalten lässt, ist  $x$   $H_2O$ .) Damit muss nicht automatisch auch geklärt sein, dass Wasser mit  $H_2O$  identisch ist, Wasser-Tatsachen also in einem logischen Sinn auf  $H_2O$ -Tatsachen supervenieren (wenn – beispielsweise – unsere jeweiligen Testverfahren verschieden sind). Die Feststellung dieser Supervenienzbeziehung könnte vielmehr das Resultat einer empirischen Untersuchung sein, die feststellt, dass diejenige Substanz, die wir mit unserem Testverfahren als „Wasser“ identifizieren, eben diejenige Substanz ist, die wir mit unserem (anderen) Testverfahren als  $H_2O$  identifizieren.

D.h. bei der Enthaltenseins-Beziehung zwischen Supervenienzbasis und der supervenierenden Ebene, muss es sich immer noch nicht um eine analytische Enthaltenseins-Beziehung handeln, oder sonst eine Beziehung, die a priori aufspürbar wäre. Der Brückenschlag zur weitergehenden These des modalen Rationalismus, dass auch die Frage des Bestehens oder Nichtbestehens einer Supervenienz- bzw. Enthaltenseinsbeziehung zwischen den beiden (a priori jeweils geklärten) Ebenen a priori beantwortet werden kann, erfolgt nun über den „Zweidimensionalismus“, eine formalsemantische Theorie, die im modalen Rationalismus erkenntnistheoretisch umgedeutet wird.

#### 7.2.1.1 Zweidimensionalismus<sup>274</sup>

„Zweidimensionalismus“ bzw. „zweidimensionale Semantik“ bezeichnet eine Familie (intensionaler) semantischer Theorien, die zwar einige Grundideen teilen, sich in Bezug auf ihren philosophischen Anspruch (Geht es ihnen nur um semantische Fragen, oder auch um Fragen der Erkenntnistheorie?), wie auch in Bezug auf ihre Erklärungsanspruch als Semantik (Welche semantischen Phänomene sollen rekonstruiert werden?) stark unterscheiden.

Gemeinsam ist den zweidimensionalen Semantiken die Auffassung, dass die Wahrheitswerte von Sätzen sowohl davon abhängen, was die Tatsachen sind, wie auch davon, was die Sätze bedeuten. Diese zweifache Abhängigkeit wird in der zweidimensionalen Semantik dadurch repräsentiert, dass sie Ausdrücken mehr als nur eine Intension zuordnet. Die formale semantische Theorie, die allen zweidimensionalen Semantiken gemeinsam ist, unterscheidet eine Dimension „aktueller Welten“ und „primärer Intensionen“ von einer zweiten Dimension „kontrafaktischer Welten“ und „sekundärer Intensionen“. Daher auch die Bezeichnung „Zweidimensionalismus“. Dieser formale Rahmen wird aber durchaus unterschiedlich interpretiert. Mit Stalnaker [313] und Nimtz [233] kann man drei verschiedene

<sup>274</sup> Die folgende Darstellung orientiert sich an stark an Nimtz [233], Stalnaker und Baldwin [312], Stalnaker [313], Chalmers [61].

Interpretationen unterscheiden: die *Kontext-Interpretation*, die *metasemantische Interpretation* und die *modalrationalistische Interpretation*.<sup>275</sup> Im Folgenden werden wir in 7.2.1.1.1 zunächst den abstrakten Rahmen der zweidimensionalen Semantik rekonstruieren und uns dann in 7.2.1.1.2 die verschiedenen Interpretationen dieses formalen Rahmens ansehen, wobei unser Hauptinteresse dem modalen Rationalismus gelten wird. In 7.2.1.2 wird erläutert, wie der Zweidimensionalismus in der Interpretation durch den modalen Rationalismus die These unterstützt, dass Supervenienzbehauptungen a priori beurteilt werden können. Die Interpretation des Zweidimensionalismus durch den modalen Rationalismus ist höchst umstritten. In 7.2.1.3 werden wir einige der Argumente betrachten, die gegen diese Interpretation gerichtet sind.

#### 7.2.1.1.1 Der abstrakte Rahmen des Zweidimensionalismus

Intensionale Semantiken lassen sich durch fünf Ideen charakterisieren (vgl. zum Folgenden Nitz [233]):

(1) *Bedeutung ist Repräsentation*. Die Bedeutung eines Satzes besteht darin, wie der Satz repräsentiert, wie die Dinge sind.

(2) *Der repräsentationale Gehalt (eines Aussagesatzes) besteht in Wahrheitsbedingungen*. Wie ein Satz repräsentiert, wie die Dinge sind, ist durch seine Wahrheitsbedingungen festgelegt. Wie ein Satz S die Dinge repräsentiert, kann von den Situationen abgelesen werden, in denen S wahr ist.

(3) *Wahrheitsbedingungen sind Wahrheitswertverteilungen über mögliche Welten*. Wahrheitsbedingungen können als Zuweisung von Wahrheitswerten zu möglichen Welten verstanden werden. Die Wahrheitsbedingungen von S können demnach so modelliert werden, dass S der Wert ‚wahr‘ in allen möglichen Welten zugewiesen wird, die mit S übereinstimmen, und der Wert ‚falsch‘ in allen Welten, die mit S nicht übereinstimmen. Wir hatten schon gesagt, dass die Menge der möglichen Welten alle Weisen enthält wie unsere Welt (vollständig spezifiziert) sein kann.

(4) *Extensionen sind kompositional*. Der Wahrheitswert eines Satzes mit bestimmter formaler Struktur ist festgelegt durch die Referenz seiner Teilausdrücke. Die Objekte, die durch singuläre Terme bezeichnet werden, die Mengen, die durch Prädikate bezeichnet werden, und die Wahrheitswerte, die durch Sätze bezeichnet werden, sind die *Extensionen* der singulären Terme, Prädikate und Sätze. Die Extension eines Satzes in einer Welt *w* ist daher bestimmt durch die Extension seiner Teilausdrücke in dieser Welt *w*. ‚Der Kanzler der Bundesrepublik Deutschland im Jahre 2004 hat gefärbte Haare.‘ ist demnach wahr in *w*, weil ‚der Kanzler der Bundesrepublik Deutschland im Jahr 2004‘ in *w* ein Objekt bezeichnet, das in der Menge der Objekte ist, die ‚hat gefärbte Haare‘ in *w* bezeichnet.

(5) *Intensionen sind ebenfalls kompositional*. Die Wahrheitsbedingungen eines Satzes sind durch die Extensionen bestimmt, die seine Teilausdrücke in den möglichen Welten haben. Eine Zuordnung von Extensionen zu möglichen Welten, also

---

<sup>275</sup> Stalnaker bezeichnet die drei Interpretationen als ‚Kaplan Paradigm‘, ‚Metasemantic Paradigm‘ und ‚Generalized Kaplan Paradigm‘.

eine Funktion  $f: W \rightarrow E$  von möglichen Welten in die Menge der Extensionen, ist eine *Intension*. So ist die Intension des Ausdrucks ‚der Kanzler der Bundesrepublik Deutschland im Jahr 2004‘ eine Funktion, die der aktuellen Welt Gerhard Schröder als Wert zuweist, und einer möglichen Welt, in der Edmund Stoiber Bundeskanzler ist, Edmund Stoiber.

Folglich sind die Intensionen von Sätzen bestimmt durch die Intensionen ihrer Teilausdrücke. ‚Der Kanzler der Bundesrepublik Deutschland im Jahr 2004 hat gefärbte Haare.‘ hat die Wahrheitsbedingungen, die der Satz hat, weil ‚Der Kanzler der Bundesrepublik Deutschland im Jahr 2004‘ in verschiedenen möglichen Welten ein Individuum bezeichnet, und ‚hat gefärbte Haare‘ eine (variiierende) Menge von Objekten. Aus diesen fünf Ideen folgt ebenfalls, dass Bedeutungen Intensionen sind, und dass Bedeutung eng mit Modalität (Möglichkeit und Notwendigkeit) verbunden ist.

#### 7.2.1.1.1.1 Kripkeanismus

Saul Kripke (vgl. Kripke [172]) ging von diesen beiden letzten Folgerungen aus, um daraus weitergehende philosophische Konsequenzen zu ziehen. Wenn wir einen Eigennamen (wie ‚Gerhard Schröder‘ oder ‚Hesperus‘) oder einen Ausdruck für eine natürliche Art (wie etwa ‚Gold‘) einführen, dann tun wir dies, um damit ein bestimmtes Objekt oder eine bestimmte Substanz zu bezeichnen. Laut Kripke sind es nun diese Objekte bzw. Substanzen, die die Intensionen der Ausdrücke ‚Gerhard Schröder‘, ‚Hesperus‘ oder ‚Gold‘ festlegen, nicht aber die Weisen, wie wir Gerhard Schröder, Hesperus oder Gold identifizieren.

‚Hesperus‘ bezeichnet in jeder möglichen Welt dasjenige Objekt, das in unserer Welt der hellste Stern am Abendhimmel ist, unabhängig davon, ob es in jeder dieser (kontrafaktischen) möglichen Welten auch diese Rolle innehat. Genauso bezeichnet ‚Gold‘ in jeder möglichen Welt dasjenige Metall, aus dem in unserer Welt die Eheringe gemacht sind, unabhängig davon, ob diese Substanz in anderen möglichen Welten auch dazu verwendet wird.<sup>276</sup> Kripke zieht aus diesen Überlegungen folgende Thesen:

(a) *Einige Ausdrücke bezeichnen starr.* Eigennamen (wie ‚Hesperus‘ oder ‚Gerhard Schröder‘) und Ausdrücke für natürliche Arten (wie ‚Gold‘ oder ‚Wasser‘) bezeichnen ein und dasselbe Objekt bzw. ein und dieselbe Substanz in jeder möglichen Welt.

(b) *Viele Identitätsaussagen sind notwendig wahr (wenn sie wahr sind).* Da ‚Hesperus‘ und ‚Venus‘ in allen möglichen Welten dasselbe Objekte bezeichnen, das sie in dieser Welt bezeichnen (weil beides Eigennamen sind), und in dieser Welt der hellste Stern am Abendhimmel die Venus ist, ist ‚Hesperus = Venus‘ notwendig wahr, also wahr in allen möglichen Welten.

---

<sup>276</sup> In diesen Beispielen wird davon ausgegangen, dass wir ‚Hesperus‘ eingeführt haben, um damit dasjenige Objekt am Abendhimmel zu bezeichnen, das am stärksten leuchtet, und ‚Gold‘ eingeführt haben, um diejenige Substanz zu bezeichnen, aus der in unserer Umgebung die Eheringe hergestellt werden.

(c) *Einige notwendige Wahrheiten sind a posteriori.* Da dasjenige chemische Element, aus dem in dieser Welt die meisten Eheringe gemacht sind, das Element mit Atomzahl 79 ist, ist ‚Gold ist das Element mit Atomzahl 79‘ notwendig wahr. Nichtsdestotrotz mussten wir empirische Untersuchungen anstellen, um herauszufinden, welches die Atomzahl desjenigen Elements ist, aus dem hier die meisten Eheringe gemacht sind. ‚Gold ist das Element mit Atomzahl 79‘ ist also eine notwendige Wahrheit *a posteriori*.

Kripke schließt daraus, dass notwendige Wahrheit und a priori Zugang nicht zusammenfallen (Metaphysik und Erkenntnistheorie werden getrennt), und dass das Wissen, dass ein kompetenter Sprecher verwendet, um die Extension eines Ausdrucks zu identifizieren, nicht den Bezug und die Wahrheitsbedingungen seiner Äußerungen festlegen kann (Semantik und Erkenntnistheorie werden getrennt, vgl. Nimtz [233]).

#### 7.2.1.1.1.2 Die Grundideen des Zweidimensionalismus

Die traditionelle intensionale Semantik (Carnap [49], Carnap [53], Føllesdal [110]) weist jedem Satz *eine* Intension zu. Diese Intension erfasst wie der Wahrheitswert eines Satzes von den jeweiligen Tatsachen abhängt und wie er mit ihnen variiert. Die *Standardintension* von ‚Gerhard Schröder hat gefärbte Haare.‘ bewertet diesen Satz mit ‚wahr‘ in eben den Welten, die damit übereinstimmen, was dieser Satz besagt. In Welten, in denen das Objekt, das mit ‚Gerhard Schröder‘ bezeichnet wird, in der Menge der Objekte ist, die mit ‚hat gefärbte Haare‘ bezeichnet wird, ist der Satz wahr, sonst falsch. Wenn der Satz (gemäß dem eben Gesagten) in  $w_1$  und  $w_2$  wahr ist und in  $w_3$  falsch, kann die Intension (für diese kleine Teilmenge der möglichen Welten) in folgender *eindimensionalen* Matrix dargestellt werden:

$w_1$	$w_2$	$w_3$
w	w	f

Abbildung 7.2-1  
Eindimensionale Standardintension von  
‚Gerhard Schröder hat gefärbte Haare.‘

Zweidimensionale Semantiken machen nun darauf aufmerksam, dass der Wahrheitswert eines Satzes auch davon abhängt (und damit variieren kann) *was* der Satz jeweils besagt. Dass der Satz ‚Gerhard Schröder hat gefärbte Haare‘ in unserer Welt den Wahrheitswert hat, den er hat, liegt daran, dass er die Proposition ausdrückt, *dass Gerhard Schröder gefärbte Haare hat*. Zweidimensionale Semantiken stimmen darin überein, dass eine formale Semantik dieser zweifachen Abhängigkeit des Wahrheitswertes von Tatsachen und Bedeutung Rechnung tragen muss und dass beide Abhängigkeiten durch eine erweiterte Mögliche-Welten-Semantik

modelliert werden können. Hierzu muss nur eine Unterscheidung zwischen „kontrafaktischen“ und „aktualen“ Welten eingeführt werden, sowie eine Unterscheidung dreier Arten von Intensionen (vgl. Nimtz [233]).

Die zweifache Abhängigkeit des Wahrheitswertes von Tatsachen und Bedeutung wird besonders deutlich an Sätzen mit indexikalischen Ausdrücken wie ‚ich‘ oder ‚jetzt‘. Ob ‚Ich habe keine gefärbten Haare.‘ in einer Welt wahr ist, hängt davon ab, was in dieser Welt der Fall ist, wie auch davon, wer diesen Satz in dieser Welt überhaupt äußert. Wenn Gerhard Schröder diesen Satz äußert, dann ist dieser Satz wahr in einer möglichen Welt  $w$ , wenn Gerhard Schröder in  $w$  keine gefärbten Haare hat. Wenn jemand anders diesen Satz äußert, hat der Satz entsprechend andere Wahrheitsbedingungen.

Ob ein Satz in einer *kontrafaktischen* Welt  $w$  wahr ist, hängt also von den Tatsachen ab. Einerseits von den Tatsachen in der *kontrafaktischen* Welt  $w$ , auf die sich der Satz bezieht, andererseits von den Tatsachen in der *aktualen* Welt, die festlegen, was er bedeutet (beispielsweise Tatsachen darüber wer einen Satz äußert). Die kontrafaktischen und aktualen Welten, die hier unterschieden werden, sind dabei keine unterschiedlichen Entitäten (vgl. Jackson [158]). Was hier unterschieden wird, sind vielmehr zwei verschiedene Rollen, die dieselben möglichen Welten einnehmen können (wenn wir für die als aktual betrachteten Welten einen Sprecher, einen Ort und eine Zeit spezifizieren).

Die Unterscheidung zwischen kontrafaktischen und aktualen (bzw. „als aktual betrachteten“) Welten erlaubt es drei verschiedenen Arten von Intensionen auseinander zu halten.<sup>277</sup>

Die *primäre Intension* eines Ausdrucks weist einem Ausdruck für jede als aktual betrachtete Welt eine Extension zu.<sup>278</sup> Es handelt sich also um eine Funktion  $f: W_A \rightarrow E$  von als aktual betrachteten Welten zu Extensionen.

Die *sekundäre Intension* eines Ausdrucks weist ihm eine Extension in jeder als kontrafaktisch betrachteten Welt zu.<sup>279</sup> Es handelt sich hierbei also um eine Funktion  $f: W_C \rightarrow E$  von als kontrafaktisch betrachteten Welten zu Extensionen.

Die *zweidimensionale Intension* weist dem Ausdruck für jede als aktual betrachtete Welt eine sekundäre Intension zu. Diese Funktion,  $f: W_A \rightarrow (W_C \rightarrow E)$ , von aktualen Welten zu sekundären Intensionen stellt den Zusammenhang zwischen den beiden obigen Intensionsbegriffen her (vgl. Nimtz [233]).

Mit diesen Intensionen kann die Abhängigkeit des Wahrheitswertes von Tatsachen und Bedeutung nun modelliert werden. So könnte man dem Satz ‚Ich habe gefärbte Haare.‘ zunächst eine primäre Intension zuweisen, die dem Satz jeweils in Abhängigkeit davon eine Extension zuweist, wer diesen Satz äußert. Dies würde erfassen, wie die Wahrheitsbedingungen des Satzes davon abhängen, von wem der Satz geäußert wird.

<sup>277</sup> Vgl. Chalmers [55], Cohnitz [71].

<sup>278</sup> Jackson spricht hier von der „A-Intension“, Stalnaker von der „diagonalen Intension“ Jackson [158], Stalnaker [313].

<sup>279</sup> Die sekundäre Intension heißt bei Jackson ‚C-Intension‘, vgl. Jackson [158].

Die sekundäre Intension würde dem Satz dann jeweils in Abhängigkeit davon eine Extension zuweisen, ob derjenige, der den Satz geäußert hat, in den kontrafaktischen Umständen gefärbte Haare hat oder nicht. Dies würde erfassen, wie die Wahrheitsbedingungen des Satzes von den jeweiligen Tatsachen abhängen.

Die zweidimensionale Semantik verbindet nun diese beiden Aspekte, indem sie erfasst, welche sekundäre Intension durch die Äußerung des Satzes in einer als aktual betrachteten Welt ausgedrückt wird. Die daraus resultierende formale Struktur (Abbildung 7.2-2) ist ebenfalls allen zweidimensionalen Semantiken gemeinsam. Eine solche zweidimensionale Matrix stellt die Intensionen eines Satzes (hier für eine kleine Teilmenge der möglichen Welten) dar. Jeder Zeile entspricht eine sekundäre Intension, die Diagonalen von oben links nach unten rechts entspricht eine primäre Intension, die gesamte Matrix stellt eine einzelne zweidimensionale Intension dar (vgl. Nimtz [233]).

		Dimension 2 Kontrafaktische Welten $\Rightarrow$		
		$w_1$	$w_2$	$w_3$
Dimension 1 Aktuale Welten $\Downarrow$	$w_1^*$	w	w	f
	$w_2^*$	f	w	w
	$w_3^*$	w	w	f

Abbildung 7.2-2

Soweit zunächst die abstrakte Darstellung der Grundideen des Zweidimensionalismus. Kommen wir nun zu den verschiedenen Interpretationen dieses formalsemantischen Analyseinstruments.

#### 7.2.1.1.2 Verschiedene Interpretationen des abstrakten Rahmens

Die verschiedenen Interpretationen des Zweidimensionalismus stimmen ebenfalls alle darin, dass Dimension 2 (Abbildung 7.2-2) und die primäre Intension erfassen, wie die Extension eines Ausdrucks davon abhängt, was er bedeutet. Allerdings sind die verschiedenen Interpretationen uneinig darüber, was die Natur dieser Abhängigkeitsbeziehung ist. Die unterschiedlichen Interpretationen geben jeweils unterschiedliche Antworten auf die folgenden beiden Frage: (a) ‚Was sind die aktuellen Welten?‘ und (b) ‚Wofür braucht man den Apparat von aktuellen Welten und primären Intensionen?‘

### 7.2.1.1.2.1 Die Kontext-Interpretation des Zweidimensionalismus<sup>280</sup>

Ausgehend von der Unterscheidung zwischen sprachlichen *tokens*, Ausdrucksvorkommnissen in konkreten Äußerungssituationen, und sprachlichen *types*, Ausdrücken außerhalb von konkreten Äußerungssituationen, beobachtet David Kaplan (vgl. Kaplan [163]), dass eine semantische Asymmetrie zwischen indexikalischen *tokens* und indexikalischen *types* besteht. Während indexikalische *tokens* zwar eine Referenz aber keine deskriptive Bedeutung haben, haben indexikalische *types* zwar eine deskriptive Bedeutung, aber keine Referenz. ‚Ich‘ in Gerhard Schröders Äußerung ‚Ich habe keine gefärbten Haare.‘ trägt zur Bedeutung des Gesamtsatzes nur die Referenz auf ein bestimmtes Individuum bei: der Satz drückt eine Proposition über Gerhard Schröder aus. Außerhalb eines konkreten Äußerungskontextes bezieht sich ‚Ich‘ andererseits auf nichts, *hat* aber eine deskriptive Bedeutung, die jeder Sprecher kennen muss: ‚Ich‘ bezieht sich auf den Sprecher der Äußerung im jeweiligen Äußerungskontext. Der Satz-*type* ‚Ich habe gefärbte Haare.‘ drückt also zunächst keine (Standard-)Proposition aus. Wenn es die Bedeutung des Satzes aber versteht, versteht jedes Mitglied der Sprachgemeinschaft auch, welche (Standard-)Proposition ein Satz-*token* dieses *types* ausdrückt, wenn ein solches *token* in einem Kontext geäußert wird (vgl. Nimtz [233]).

Kaplan unterscheidet daher zwischen zwei Arten von Bedeutung, „Gehalt“ („content“) und „Charakter“ („character“). Der Gehalt eines Ausdrucks ist das, worauf sich der Ausdruck bezieht, der Charakter eine konventionell bestimmte Regel, die angibt, welchen Gehalt ein *token* eines Ausdrucks-*types* ausdrückt, wenn es in einem spezifizierten Kontext geäußert wird. Innerhalb des zweidimensionalen Rahmens werden Gehalte nun als sekundäre Intensionen modelliert: Die sekundäre Intension eines Satz-*tokens* spezifiziert seine Wahrheitsbedingungen und repräsentiert die Proposition, die der Satz als bestimmte (konkrete) Äußerung ausdrückt. Charaktere sind hingegen zweidimensionale Intensionen. Die zweidimensionale Intension eines Satz-*types* spezifiziert eine sekundäre Intension für jede als aktual betrachtete mögliche Welt. Die zweidimensionale Intension repräsentiert also, wie die jeweils ausgedrückte Proposition mit dem Kontext variiert, in dem ein *token* des *types* geäußert wird.

Bei Kaplan sind aktuelle Welten also mögliche Äußerungskontexte. Aktuelle Welten und primäre Intensionen brauchen wir in der formalen Semantik, wenn wir die Kontextabhängigkeit der Sprache modellieren wollen (vgl. Nimtz [233]).

### 7.2.1.1.2.2 Die metasemantische Interpretation des Zweidimensionalismus

Stalnakers metasemantische Interpretation des Zweidimensionalismus kann als Versuch verstanden werden, das *Informativitätsproblem* notwendiger Wahrheiten in den Griff zu bekommen (vgl. Stalnaker [311], Nimtz [233], Haas-Spohn [134]).<sup>281</sup> Das Informativitätsproblem entsteht dadurch, dass die folgenden drei Aussagen *prima facie* miteinander unvereinbar scheinen:

<sup>280</sup> Die Darstellung orientiert sich an Nimtz [233]. Vgl. auch Haas-Spohn [134].

<sup>281</sup> Vgl. zu diesem Problem insbesondere Bremer und Cohnitz [34], Cohnitz [69], Haas-Spohn [134].

- (i) Da notwendig wahr, schließt die Proposition, die durch ‚Hesperus = Phosphorus‘ ausgedrückt wird, keine Möglichkeit aus.
- (ii) Ein Satz kann nur dann zur Kommunikation von kontingenter Information verwendet werden, wenn die Proposition, die er übermittelt, Möglichkeiten ausschließt.
- (iii) ‚Hesperus = Phosphorus‘ kann zur Kommunikation von kontingenter Information über die Welt verwendet werden.

Stalnaker glaubt, dass der Widerspruch zwischen (i)-(iii) nur scheinbar besteht. Um ihn aufzulösen, unterscheidet er zwischen der Proposition, die bei einem informativen Gebrauch von ‚Hesperus = Phosphorus‘ vermittelt wird, und der Proposition, die bei diesem Gebrauch tatsächlich ausgedrückt wird. Die ausgedrückte Proposition ist notwendig wahr und wird durch die semantischen Regeln der Sprache festgelegt. Die übermittelte Proposition wird hingegen aus den pragmatischen kommunikativen Absichten des Sprechers erschlossen und ist kontingent.

Pragmatische Interpretationsmaximen können Hörer dazu veranlassen, in Situationen, in denen ein Sprecher einen Satz äußert, der nach den semantischen Regeln notwendig wahr ist, den fraglichen Satz umzuinterpretieren. Hört man in einer Äußerungssituation beispielsweise den Satz ‚Hesperus = Phosphorus‘, der offenbar in der kommunikativen Absicht geäußert wurde, eine kontingente Information über die Welt zu vermitteln, und stellt man fest, dass ‚Hesperus = Phosphorus‘ nach den semantischen Regeln der Sprache eigentlich eine notwendige Wahrheit ausdrückt, könnte man folgende Überlegung anstellen:

- (i) ‚Hesperus‘ wurde als Name für den hellsten Stern am Abendhimmel eingeführt, ‚Phosphorus‘ als Name für den hellsten Stern am Morgenhimmel.
- (ii) Welche Objekte diese Bezeichnungsweise herausgriff, hing von den astronomischen Tatsachen in unserer Welt ab. Diese astronomischen Tatsachen waren kontingent. Falls sie anders gewesen wären, hätten ‚Hesperus‘ und ‚Phosphorus‘ zwei verschiedene Objekte bezeichnet.
- (iii) Der Sprecher möchte übermitteln, dass unsere Welt eine Welt ist, in der die astronomischen Tatsachen nicht auf diese Weise anders sind. Er möchte übermitteln, dass unsere Welt der Proposition entspricht, *dass der hellste Stern am Morgenhimmel mit dem hellsten Stern am Abendhimmel identisch ist.* (Vgl. Nimtz [233])

Die Standardbedeutungen von Sätzen werden in Stalnakers Interpretation des zweidimensionalen Apparats durch sekundäre Intensionen repräsentiert. Die Propositionen, die in Reinterpretationen zugewiesen werden, werden als primäre Intensionen repräsentiert.

In Stalnakers Interpretation sind die aktual betrachteten möglichen Welten demnach mögliche alternative Umwelten, in denen wir unsere Ausdrücke einge-

führt haben könnten. Die als aktual betrachteten möglichen Welten und die primären Intensionen repräsentieren bestimmte metasemantische Tatsachen wie die semantische Bedeutung unserer Ausdrücke von den Gegebenheiten in der aktuellen Welt abhängt. Dieses metasemantische Wissen ziehen wir manchmal pragmatisch heran, um ansonsten unverständlich Äußerungen zu interpretieren.

#### 7.2.1.1.2.3 Die modal-rationalistische Interpretation des Zweidimensionalismus

Sowohl Kaplans als auch Stalnakers Interpretation des Zweidimensionalismus sind bloße Erweiterungen von Kripkes Auffassungen. Beide Ansätze sind völlig damit vereinbar, dass Metaphysik und Erkenntnistheorie getrennt sind, wie auch damit, dass Erkenntnistheorie und Semantik getrennt sind (vgl. Chalmers [61], Nimtz [233]).

In Kaplans Interpretation, bei der es um die Repräsentation innersprachlicher Regeln zum Umgang mit Kontextabhängigkeit geht, kommen die „metaphysischen Notwendigkeiten“, also die Notwendigkeiten, die nach Kripke a posteriori aber notwendig wahr sind, *gar nicht* als Sonderproblem vor. Da Namen und Artausdrücke keine klarerweise indexikalischen Ausdrücke sind, haben metaphysische Notwendigkeiten eine in allen möglichen Welten wahre primäre Intension und eine in allen möglichen Welten wahre sekundäre Intension (vgl. Stalnaker [313]). Außerdem verlangt die Kontext-Interpretation des Zweidimensionalismus, dass sich in jeder als aktual betrachteten Welt auch die zu analysierende Äußerung befindet. Dies macht diese Interpretation für epistemologische Zwecke aber unbrauchbar: Warum sollte es a priori sein, dass in jeder möglichen Welt jemand etwas sagt?<sup>282</sup>

Auch in der metasemantischen Interpretation bleibt die Trennung zwischen Metaphysik und Semantik, wie zwischen Erkenntnistheorie und Semantik bestehen. Die metasemantische Analyse erfasst Beziehungen zwischen Tatsachen der Ausdruckseinführung und der Ausdrucksbedeutung, diese Tatsachen müssen aber nicht allen kompetenten Sprechern zugänglich sein und sind darüber hinaus auch nicht Bestandteil der sprachlichen Bedeutung dieser Ausdrücke. Der metasemantische Apparat *kann* zur Analyse der kognitiven Signifikanz von Äußerungen eingesetzt werden, aber nur dann, wenn eine Interpretation der Äußerung im Sinne der tatsächlichen sprachlichen Bedeutung an Interpretationsmaximen scheitert.

In der modal-rationalistischen Interpretation des Zweidimensionalismus werden die Zusammenhänge zwischen Semantik und Metaphysik bzw. zwischen Erkenntnistheorie und Semantik nun wieder hergestellt.

Der modale Rationalismus geht dabei von stark kontroversen Annahmen aus. Erstens nimmt er an, dass es so etwas geben kann, wie eine kanonische, neutrale Beschreibung der Welt, die es (prinzipiell) erlaubt, alle wahren Sätze a priori zu erschließen, wie auch die Referenz aller Ausdrücke a priori zu kennen. So kann von einer Beschreibung des Verhaltens, des Aussehens und des Aufbaus chemischer Substanzen, die das Wort ‚Gold‘ nicht enthält, a priori erschlossen werden, dass ‚Gold ist die chemische Substanz mit Atomzahl 79‘ wahr ist (vgl. Chalmers [60],

<sup>282</sup> Vgl. insbesondere die Argumentation in Chalmers [61].

Chalmers [57], Chalmers [59], Chalmers [61], Chalmers und Jackson [63], Jackson [158]).

Zweitens nimmt der modale Rationalismus an, dass es so etwas wie epistemisch mögliche Welten gibt: vollständige kanonische Beschreibungen, die die Welt so darstellen, wie sie *a priori* sein könnte. Für eine solche vollständige „epistemische Möglichkeit“ folgt dann natürlich (durch die erste Annahme) auch, dass ein Sprecher *a priori* weiß, welche seiner Sätze in dieser epistemischen Möglichkeit wahr wären und worauf sich seine Ausdrücke beziehen würden.

Nach der Interpretation des modalen Rationalismus verbinden Sprecher mit ihren Ausdrücken und Sätzen Funktionen von solchen epistemischen Möglichkeiten in die Menge der Extensionen. Diese „epistemische Intension“, die für jeden Sprecher mit jedem seiner Ausdrücke verbunden ist, ist von fundamentaler Bedeutung für die Signifikanz des jeweiligen Ausdrucks (vgl. Nimtz [233]).

*Erstens* bestimmt sie die kognitive Signifikanz. Die epistemische Intension ist das, was ein Ausdruck für einen Sprecher bedeutet. Wenn ‚Wasser‘ in jeder epistemischen Möglichkeit zum Espresso gereicht wird, dann gehört dies zur Bedeutung dieses Ausdrucks für diesen Sprecher.

*Zweitens* bestimmt die epistemische Intension die Extension in der aktuellen Welt, da die aktuelle Welt einfach eine aktualisierte epistemische Möglichkeit ist.

*Drittens* bestimmt die epistemische Intension die kontrafaktische Intension aller Ausdrücke, deren kontrafaktische Intension von der aktuellen Welt abhängt. Angenommen die epistemische Intension für ‚Gold‘ bestimmt, dass Gold dasjenige Metall ist, aus dem in unserer Umgebung die Eheringe gemacht sind. Da die Eheringe in unserer Umgebung aus dem chemischen Element mit der Atomzahl 79 gemacht sind und ‚Gold‘ starr das bezeichnet, was es in der aktuellen Welt bezeichnet, bezeichnet, ist ‚Gold hat die Atomzahl 79.‘ wahr in allen möglichen kontrafaktischen Welten.

Der modale Rationalismus identifiziert primäre Intensionen mit diesen epistemischen Intensionen. Diese Intensionen sind dadurch festgelegt, wie wir unsere Ausdrücke verstehen. Die sekundäre Intension repräsentiert die Extension eines Ausdrucks in kontrafaktischen Alternativwelten, und die zweidimensionale Intension gibt an, wie die sekundäre Intension von der primären Intension abhängt. Als aktual betrachtete Welten sind in dieser Interpretation also epistemische Möglichkeiten, und primäre Intensionen sind erforderlich, um die epistemische Abhängigkeit der Bedeutung zu analysieren.

Den modalen Rationalisten zufolge haben wir dadurch, dass wir den Begriff ‚Wasser‘ erworben haben, bereits substanzielles *a priori* Wissen erworben: Wir können nur auf der Grundlage unseres Bedeutungswissens eine zweidimensionale Matrix wie in Abbildung 7.2-3 vollständig rekonstruieren, da wir die epistemisch interpretierten primären Intensionen unserer Ausdrücke kennen:

What we can know independently of knowing what the actual world is like can properly be called *a priori*. The sense in which conceptual analysis involves the *a priori* is that it concerns A-extensions at worlds, and so A-intensions [primäre Intensionen], and accordingly concerns something that does, or does not, obtain in-

dependently of how things actually are. When we do conceptual analysis of K-hood, we address the question of what it takes to be a K in the sense of when it is right, and when it is wrong, to describe some situation in terms of 'K', and so we make explicit what our subject is when we discuss Ks. The part of this enterprise that addresses the question of what things are K at a world, under the supposition that that world is the actual world, is the a priori part of conceptual analysis, because the answer depends not at all on which world is in fact the actual world [...]. (Jackson [158], 51)

Was wir nicht a priori wissen können, ist in welcher Zeile, also in welchem Referenzfixierungskontext wir uns befinden.

Auswertungskontext

		Auswertungskontext			
		<i>aktuale Welt</i>	$w_1$	$w_2$	$w_3$
Referenzfixierungskontext	<i>aktuale Welt</i>	H <sub>2</sub> O	H <sub>2</sub> O	H <sub>2</sub> O	H <sub>2</sub> O
	$w_1$	XYZ	XYZ	XYZ	XYZ
	$w_2$	H <sub>2</sub> O	H <sub>2</sub> O	H <sub>2</sub> O	H <sub>2</sub> O
	$w_3$	QXY	QXY	QXY	QXY

Abbildung 7.2-3

Zweidimensionale Intension von ‚Wasser‘  
(In  $w_1$  erfüllt XYZ die „Rolle“ von Wasser,  
in  $w_3$  QXY)

### 7.2.1.2 Der Brückenschlag

Zwischen der a priori zugänglichen primären Intension eines Ausdrucks und seiner gegebenenfalls nur a posteriori zugänglichen sekundären Intension besteht nun – laut modalem Rationalismus – ein systematischer Zusammenhang, der mit folgender These wiedergegeben werden kann:

- (2D) Ein Satz S ist a posteriori notwendig wahr, gdw. er eine kontingente primäre Intension und eine notwendig wahre sekundäre Intension hat.

Der Satz

- (S) Wasser ist H<sub>2</sub>O.

wäre ein Beispiel für einen solchen Satz: Gegeben unsere obigen Annahmen ist er zwar notwendig wahr, bzw. besitzt eine sekundäre Intension, die allen kontrafaktischen Welten den Wahrheitswert ‚wahr‘ zuweist, sein Wahrheitswert ändert sich aber, wenn mit dem Auswertungskontext auch der Referenzfixierungskontext variiert:

	<i>aktuale Welt</i>	$w_1$	$w_2$	$w_3$
<i>aktuale Welt</i>	WAHR	WAHR	WAHR	WAHR
$w_1$	FALSCH	FALSCH	FALSCH	FALSCH
$w_2$	WAHR	WAHR	WAHR	WAHR
$w_3$	FALSCH	FALSCH	FALSCH	FALSCH

sekundäre Intension von ‚Wasser ist H<sub>2</sub>O‘

primäre Intension von ‚Wasser ist H<sub>2</sub>O‘

Abbildung 7.2-4

Betrachten wir zum Vergleich den Satz

(S\*) Der Urmeter in Paris ist 1m lang.,

der – laut<sup>283</sup> Kripke (vgl. Kripke [172]) – zwar a priori wahr aber nichtsdestotrotz kontingent ist, und bewerten ihn in der aktuellen Welt, in einer Welt  $w_1$ , in der der Pariser Urmeter 98cm lang ist,  $w_2$ , in der er ebenfalls 100cm lang ist und  $w_3$ , in der er 102cm lang ist:

<sup>283</sup> Wir folgen in dieser Darstellung den Standardbeispielen unkritisch. Vgl. aber Needham [222] und Loomis [189].

	<i>aktuelle Welt</i>	$w_1$	$w_2$	$w_3$
<i>aktuelle Welt</i>	WAHR	FALSCH	WAHR	FALSCH
$w_1$	FALSCH	WAHR	FALSCH	FALSCH
$w_2$	WAHR	FALSCH	WAHR	FALSCH
$w_3$	FALSCH	FALSCH	FALSCH	WAHR

sekundäre Intension von  
'Der Urmeter in Paris ist  
1m lang'

primäre Intension von  
'Der Urmeter in Paris ist  
1m lang'

Abbildung 7.2-5

Dementsprechend gilt im Zweidimensionalismus auch Folgendes:

(2D\*) Ein Satz S ist a priori wahr gdw. er eine notwendig wahre primäre Intension hat.

„Substanziell“ ist unser a priori Wissen nun insofern, als es uns direkt wie indirekt über metaphysische Möglichkeiten informiert. Folgen wir zunächst dem indirekten Weg (aus Chalmers [55], Byrne [47]).

Betrachten wir hierzu eine metaphysische These T, die zwar eine metaphysische Notwendigkeit ausdrücken soll, aber zugleich a posteriori herausgefunden wurde (etwa eine Konsequenz aus  $P_1$ ). Sei  $\Phi$  ein physikalischer Satz, der genau in den Welten wahr ist, die minimale physikalische Duplikate der aktuellen Welt sind,  $\Psi$  ein Satz, der eine wahre Proposition über Mentales ausdrückt, aber nicht a priori aus den physikalischen Tatsachen folgt, dann ist

(T)  $\Phi \supset \Psi$ .

a posteriori und notwendig wahr, wenn  $P_1$  wahr ist (es wird angenommen, dass für jeden wahren Satz  $\alpha$ ,  $\Phi \supset \alpha$  notwendig wahr ist, da  $\Phi$  die physikalischen Wahrheiten der Welt vollständig determiniert und alles über dem Physikalischen

superveniert). Sei nun für jeden Satz  $\alpha$ ,  $\alpha^p$  ein Satz, dessen sekundäre Intension gleich der primären Intension von  $\alpha$  ist.<sup>284</sup>

Hier stellt sich aber nun ein Dilemma für den Physikalisten. Die primäre Intension von  $\Phi$  ist entweder gleich seiner sekundären Intension, oder nicht. Angenommen, sie ist gleich, dann ist die primäre Intension von  $\Phi \supset \Psi$  gleich der sekundären Intension von  $\Phi^p \supset \Psi^p$ , gleich der sekundären Intension von  $\Phi \supset \Psi^p$ . Die primäre Intension von  $\Phi \supset \Psi$  sollte nach physikalistischer Auffassung aber kontingent sein, da nach (2D) eine a posteriori Notwendigkeit eine kontingente primäre Intension haben muss. Wenn  $\Phi \supset \Psi$  eine kontingente primäre Intension hat, und diese gleich der sekundären Intension von  $\Phi \supset \Psi^p$  ist, ist der Physikalismus aber falsch, da  $\Psi^p$  wahr ist (in der aktualen Welt fallen primäre und sekundäre Intension zusammen) und ja für jeden wahren Satz  $\alpha$  gilt, dass  $\Phi \supset \alpha$  notwendig wahr ist, da  $\Phi$  die physikalischen Wahrheiten der Welt vollständig determiniert und alles über dem Physischen superveniert,  $\Phi \supset \Psi^p$ , also keine kontingente sekundäre Intension haben darf, wenn der Physikalismus wahr sein soll.

Nehmen wir also an, dass die primäre Intension von  $\Phi$  sich von seiner sekundären Intension unterscheidet. In diesem Fall ist der Physikalismus zwar nicht falsch, aber als metaphysische Aussage leer, da die Welt durch  $\Phi$  sozusagen metaphysisch unterbestimmt bleibt. Zu behaupten, alles supervenierte auf dem Physischen, würde dann nicht mehr das ausdrücken, was eigentlich intendiert war.

Eine solche indirekte Argumentation gegen den Physikalismus beruht etwa auf dem Zombie-Gedankenexperiment, wie wir es in 4.1 kennen gelernt haben. Wenn eine physikalische Duplikat-Welt der aktualen Welt vorstellbar ist, in der mindestens einer von uns ein Zombie ist, dann ist diese Welt entweder metaphysisch oder epistemisch möglich.<sup>285</sup> Nach obiger Analyse wäre beides für den Physikalismus vernichtend.<sup>286</sup>

Der Brückenschlag von epistemischer zu metaphysischer Notwendigkeit ergibt sich aus dem Zweidimensionalismus aber auch direkt.<sup>287</sup> Angenommen es ginge darum, die Supervenienzbeziehung zwischen  $H_2O$ -Tatsachen und Wassertatsachen zu etablieren. Laut Jackson besteht zwischen den beiden folgenden Sätzen die oben erläuterte Enthaltenseins-Beziehung:

(S<sub>1</sub>)  $H_2O$  bedeckt den größten Teil der Erdoberfläche.

(S<sub>2</sub>) Wasser bedeckt den größten Teil der Erdoberfläche.

<sup>284</sup> Wir nehmen einmal dem Argument zuliebe an, dass dies ohne weiteres geht, etwa indem man alle eventuellen rigiden Designatoren durch Kennzeichnungen ersetzt.

<sup>285</sup> Der Zusammenhang von Vorstellbarkeit zu Möglichkeit wird uns im nächsten Kapitel beschäftigen.

<sup>286</sup> Für eine Diskussion dieser Auffassung vgl. Cohnitz [71], Perry [255], Byrne [47]. An dieser Stelle würde uns eine inhaltliche Auseinandersetzung mit dem Zombie-Argument zu weit in die Philosophie des Geistes führen. Es geht hier nur um die Frage, wie der modale Rationalismus von solchen Argumenten Sinn macht.

<sup>287</sup> Wohlgemerkt spielt in obigem Argument der besondere Charakter des Phänomenalen gar keine Rolle. Insofern unterscheiden sich Chalmers' und Jacksons Argument durchaus von denen, die Kripke [172] gegen den Physikalismus vorbringt. Vgl. aber Yablo [360].

Tatsächlich „enthält“  $S_1$  die Tatsache, die durch  $S_2$  ausgedrückt wird, wenn alle Wasser-Welten  $H_2O$ -Welten sind. In diesem Fall ist in jeder Welt, in der  $S_1$  wahr ist, auch  $S_2$  wahr. Allerdings – so wird man einwenden wollen – ist das Konditional » $S_1 \supset S_2$ « (wenn auch notwendig, so doch) nur *a posteriori* wahr.

Der Zweidimensionalismus führt die Aposteriorizität solcher Konditionale auf das Fehlen von Kontextwissen zurück. Wie wir oben gesagt haben, können wir laut Zweidimensionalismus die Intensionen-Matrix für jeden Ausdruck der Sprache, den wir verstehen, *a priori* rekonstruieren. Bei metaphysischen Notwendigkeiten können wir den Wahrheitswert aber unter Umständen erst dann beurteilen, wenn wir über zusätzliches Wissen darüber verfügen, in welchem Referenzfixierungskontext wir uns befinden. Würde man solche Kontextinformationen zu  $S_1$  hinzufügen, könnte man sehr wohl *a priori* wissen, dass  $S_2$  in  $S_1$  enthalten ist:

- ( $S_1$ )  $H_2O$  bedeckt den größten Teil der Erdoberfläche.
- ( $S_{1a}$ )  $H_2O$  ist die wässrige Substanz, die wir kennen.
- ( $S_2$ ) Wasser bedeckt den größten Teil der Erdoberfläche.

Auch dies scheint vielleicht noch unproblematisch zu sein. Obzwar man von  $S_1$  nicht *a priori* zu  $S_2$  übergehen kann, kann man von  $S_1$  und  $S_{1a}$  sehr wohl *a priori* auf  $S_2$  schließen (Jackson [158], 82).

Dass der Physikalismus auf die These verpflichtet ist, dass alle wahren Sätze über Mentales aus einer vollständigen, wahren physikalischen Beschreibung der Welt *a priori* deduzierbar sein sollen, begründet Jackson damit, dass eine „seriöse“ metaphysische Position auch die relevanten Informationen über den Kontext „lokalisieren“ muss:

The crucial point here is the way that the contextual information, the relevant information about how things actually are, by virtue of telling us in principle the propositions expressed by the various sentences (or, equivalently, the C-propositions [sekundäre Propositionen] associated with them) enables us to move *a priori* from the  $H_2O$  way things are to the water way they are. But if physicalism is true, all the information needed to yield the propositions being expressed about what the actual world is like in various physical sentences can be given in physical terms, for the actual context is givable in physical terms according to physicalism. Therefore, physicalism is committed to the in principle *a priori* deducibility of the psychological from the physical. (Jackson [158], 83)

Damit kann „seriöse“ Metaphysik grundsätzlich *a priori* betrieben werden, solange es ihr „nur“ darum geht, das „location-problem“ zu lösen.<sup>288</sup>

---

<sup>288</sup> Jackson nennt dies die „moderate“ Auffassung von Metaphysik. „Moderat“ ist diese Auffassung insofern, als sie nur Aussagen relativ zu unserem Verständnis und Gebrauch des relevanten Vokabulars macht. „Nicht-moderat“ wäre eine Metaphysik, wenn sie von sich bean-

### 7.2.1.3 Kritik am modalen Rationalismus

Der Zweidimensionalismus leistet also Folgendes: (a) Er klärt in gewissem Sinne das Verhältnis von logischen<sup>289</sup> und metaphysischen Möglichkeiten. Es gibt nur eine Menge möglicher Welten: die metaphysischen Möglichkeiten, die sowohl die Rolle des Auswertungs- wie auch die Rolle des Referenzfixierungskontexts spielen können. Wird ein Referenzfixierungskontext betrachtet, der von der aktuellen Welt abweicht, wird gegebenenfalls eine bloß logische Möglichkeit repräsentiert (Jackson 2000). Solche „logischen“ Möglichkeiten, die wir durch Vorstellbarkeitstests aufspüren können, können dabei niemals vollständige Täuschungen sein. Wir können keine Möglichkeiten *halluzinieren*, schlimmstenfalls handelt es sich um eine bloße *Illusion*: Hinter jeder logischen Möglichkeit verbirgt sich eine – unter Umständen falsch beschriebene – metaphysische Möglichkeit. In manchen Fällen können wir aber a priori ausschließen, dass es sich um eine Illusion handeln könnte. Wenn das relevante Kontextwissen, das zur Referenzfixierung nötig ist, uns vollständig gegeben ist (wir also ausschließen können, dass weitere Informationen über den Referenzfixierungskontext unser modales Urteil abändern würden), kann es sich auch nicht um eine modale Illusion handeln.<sup>290</sup>

(b) Kripke hatte zunächst Epistemologie und Metaphysik getrennt, indem er scheinbar aufwies, dass es Notwendigkeiten gibt, die nicht a priori sind (vgl. Nitz [233], Chalmers [61]). Im Zweidimensionalismus werden diese Bereiche wieder miteinander verbunden. Es wird auch geklärt, in welchem Sinn logische (also „epistemische“) Notwendigkeiten metaphysisch kontingent sein können bzw. metaphysische Notwendigkeiten logisch (also „epistemisch“) kontingent. Die Antwort des Modalen Rationalisten ist, dass, wenn wir die von den Sätzen ausgedrückten Propositionen betrachten und zwischen einer primären Proposition und einer sekundären Proposition genauso unterscheiden wie zwischen primärer und sekundärer Intension, Aposteriorizität und Kontingenz, wie auch Apriorizität und Notwendigkeit zusammenfallen, je nachdem welche Proposition betrachtet wird.

Trotz dieser vorteilhaften Aspekte ist der modale Rationalismus gegenwärtig stark umstritten. Dabei sind vor allem zwei Problembereiche hervorgetreten:

(i.) Es ist unklar, wie die primären Intensionen abgegrenzt werden. (ii.) Es ist unklar, ob (2D) und (2D\*) wahre Verallgemeinerungen darstellen.

#### 7.2.1.3.1 Die Abgrenzung der primären Intensionen

Entweder ist der Zweidimensionalismus nichts anderes als ein Neo-Deskriptivismus, der sehr wahrscheinlich falsch ist, oder der Zweidimensionalismus wird leer, weil er als primäre Intension praktisch alles zulässt.

---

spruchte Aussagen treffen zu können, die unabhängig von unserem Vokabular gültig sind.

Vgl. Chalmers und Jackson [63].

<sup>289</sup> ‚Logisch‘ wird an dieser Stelle synonym mit ‚begrifflich‘ und ‚epistemisch‘ verwendet. Es handelt sich also um Möglichkeiten, die das fragliche Individuum nicht schon aufgrund idealer a priori Reflexion auf sein Bedeutungswissen ausschließen kann.

<sup>290</sup> Insofern erfüllt der modale Rationalismus, was Hintikka, Fodor und Mayer gefordert haben, nämlich eine Theorie zu liefern, die erklärt, warum wir bestimmtes Wissen a priori haben sollten.

In der Variante des gehaltvollen (aber vermutlich falschen) Neo-Deskriptivismus können die epistemischen Intensionen, die ein Sprecher mit einem Ausdruck verbindet, zum Teil in der Form von Beschreibungen aufgefasst werden. Der Teil der Bedeutung von ‚Wasser‘, der „im Kopf“ ist (und damit die primäre Intension von Wasser festlegt), wäre so etwas wie

(DW) das flüssige, durchsichtige, geschmacks- und geruchslose Zeug, das sich in Flüssen und Meeren befindet, als Regen vom Himmel fällt und in südlichen Ländern zum Espresso gereicht wird

Je nachdem, was in einer als Referenzfixierungskontext betrachteten Welt unter DW fällt (sei es  $H_2O$ , wie in der aktuellen Welt, oder XYZ), wäre dann in der primären Intension von ‚Wasser‘. Eine solche Auffassung hat mit mindestens drei Gegenargumenten zu kämpfen (vgl. zum Folgenden Nimtz [231], Nimtz [232]): (1.) dem Unwissenheitsargument, (2.) dem Subjektivitätsargument und (3.) dem Überzeugungsrevisionsargument.

#### 7.2.1.3.1.1 Das Unwissenheitsargument

Für eine sehr große Zahl von Begriffen kann ein gewöhnlicher Sprecher vermutlich keine solche Beschreibung wie in DW angeben, weil er über die fragliche Substanz oder Eigenschaft schlicht zu wenig weiß (z.B. ‚Elektron‘ oder ‚Tapir‘). Dennoch würde man ihm in diesen Fällen häufig nicht absprechen wollen, dass er trotzdem über einen solchen Begriff verfügt.

Sprecher, die sehr bizarre Auffassungen über eine Substanz oder eine Eigenschaft haben, könnten sich mit ihren Begriffen nicht auf diese beziehen. Entsprechend könnten wir uns auch nicht auf dieselben Eigenschaften oder Substanzen wie sie beziehen, wenn wir sie korrigieren wollten. Das ist ziemlich unplausibel (vgl. Nimtz [231]).

#### 7.2.1.3.1.2 Das Subjektivitätsargument

Als Reaktion auf obigen Einwand wird von manchen modalen Rationalisten (vgl. Chalmers [59], Chalmers und Jackson [63]) zugestanden, dass die primären Intensionen von Ausdrücken von Sprecher zu Sprecher variieren können. Dann ist aber fraglich, ob primäre Intensionen noch ihre Erklärungsfunktion erfüllen können. Primäre Intensionen sollten ja insbesondere erklären, welche Wahrheiten einer Sprache a priori sind, und damit, was zu den analytischen Wahrheiten einer Sprache gehört. Wenn die primären Intensionen von Sprecher zu Sprecher variieren, dann sind für verschiedene Sprecher einer Sprache auch verschiedene Sätze a priori bzw. analytisch. Will man einen übergreifenden Begriff von Apriorizität simpliciter für eine Sprache angeben, steht man vor folgenden, gleichermaßen implausiblen Alternativen:

Weite Apriorizität simpliciter

Ein Satz ist *a priori simpliciter* in  $L$  gdw. er für mindestens einen Sprecher von  $L$  zu mindestens einem Zeitpunkt a priori ist.

Enge Apriorizität simpliciter

Ein Satz ist *a priori simpliciter in L* gdw. er für alle Sprecher von *L* zu allen Zeitpunkten a priori ist. (Vgl. Chalmers und Jackson [63].)

Apriorizität simpliciter

Ein Satz ist *a priori simpliciter in L* gdw. er für einen Sprecher von *L* auf der Grundlage idealer rationaler Reflektion a priori ist.

Alle drei Apriorizitätsbegriffe sind kritisiert worden (vgl. Nimtz [231], Nimtz [232]). Ersterer scheint viel zu viele Sätze einer Sprache als analytisch auszuzeichnen, der zweite zu wenige. Der dritte Vorschlag schließlich ist zumindest klärungsbedürftig: Angenommen ich weiß fast nichts über Tapire, außer, dass sie relativ exotische Säugetiere mit vier Beinen und einer komischen Schnauze sind, wie soll ideale Reflexion das verbessern?

#### 7.2.1.3.1.3 Das Überzeugungsrevisionsargument

Der modale Rationalismus versucht, einen Unterschied zwischen unserem begrifflichen Wissen und unserem Faktenwissen zu explizieren. DW repräsentiert beispielsweise unser begriffliches Wissen über ‚Wasser‘; dass Wasser eine bipolare Molekülstruktur hat, ist ein Teil unseres Faktenwissens über Wasser. Ersteres bestimmt die Referenz von ‚Wasser‘, Letzteres gehört zu den Dingen, die unsere Experten über Wasser herausgefunden haben. Diese Darstellung ist vermutlich inadäquat.

Zum *Einen* ist auch das von Experten herausgefundene Faktenwissen nicht unbeteiligt an der Genese der primären Intension, die ein Sprecher mit ‚Wasser‘ verbindet: Wenn ich weiß, dass Wasser eine bipolare Molekülstruktur hat, werde ich vermutlich Schwierigkeiten haben, mir eine Welt vorzustellen, in der das wässrige Zeugs eine solche Struktur nicht hat. Philosophen, die einen moderaten semantischen Holismus vertreten, würden zumindest behaupten, dass die in die Alltagstheorie einfließenden wissenschaftlichen Erkenntnisse über die mit den jeweiligen Ausdrücken bezeichneten Dinge und Arten die primäre Intension eines Ausdrucks mit der Zeit verändern.<sup>291</sup>

Zum *Anderen* ist es durchaus vorstellbar, dass unsere Experten herausfinden, dass Teile von DW revidiert werden müssen: Das, was vom Himmel fällt, ist streng genommen kein Wasser, sondern verwandelt sich erst durch einen komplizierten Vorgang der Induktion<sup>292</sup> in Wasser, wenn es mit dem Wasser der Meere oder Flüsse in Berührung kommt.

#### 7.2.1.3.2 Der Status von (2D) und (2D\*)

Es ist unklar, ob (2D) bzw. (2D\*) wahre Verallgemeinerungen darstellen. Möglicherweise gibt es notwendige Wahrheiten, die aufgrund der in ihnen vorkommen-

<sup>291</sup> Vgl. hierzu die Position des „moderaten Deskriptivismus“, die weiter unten beschrieben wird.

<sup>292</sup> ‚Induktion‘ im Sinne einer magnetischen Induktion, oder etwas Ähnlichem. Es geht hier nicht darum, ein chemisch plausibles Beispiel zu finden.

den Ausdrücke eine notwendige primäre Intension haben, aber nicht a priori wissbar sind. So etwas wäre mutmaßlich der Fall, wenn es Theoreme der elementaren Zahlentheorie gäbe, die wir prinzipiell nicht wissen können.<sup>293</sup> Solche Sätze hätten eine notwendige primäre wie sekundäre Intension aufgrund der in ihnen vorkommenden Ausdrücke, wären aber falls wahr, nicht a priori wahr. Ein ähnliches Beispiel könnte vorgebracht werden, wenn es „starke metaphysische Notwendigkeiten“ gäbe, also beispielsweise in einer konsistenten Spielart des Theismus Gott zwar notwendig existiert, aber für uns epistemisch unerreichbar wäre (und Atheismus immer eine logische Möglichkeit bliebe) (vgl. Chalmers [55], Yablo [360]). Wenn (2D) oder (2D\*) aber nur für manche Sätze wahr sind, für andere aber nicht, dann verliert der Zweidimensionalismus einiges an Attraktivität, insbesondere für das Gebiet der modalen Epistemologie. Schließlich könnte es ja sein, dass gerade das Gebiet, das zur Debatte steht, die entsprechenden Anomalien aufweist und deswegen die Vorstellbarkeit eines relevanten Sachverhalts in diesem Gebiet keinen Rückschluss auf Metaphysisches erlaubt.<sup>294</sup> Da es sich beim modalen Rationalismus um eine Theorie handelt, die aktuell ausgearbeitet wird, macht es wenig Sinn, sie hier einer umständlicheren Analyse zu unterziehen. Es sollte aber klar geworden sein, dass der modale Rationalismus – wenn er sich verteidigen lässt – eine Fundierung des Gedankenexperimentierens darstellt, die unbeschadet von der Frage ist, ob es sich beim Gegenstand der Philosophie um natürliche Arten handelt. Betrachten wir zum Vergleich die Auffassung, die Gedankenexperimentieren mit natürlichen Arten wesentlich problematischer sieht: die Putnam Orthodoxie.

---

<sup>293</sup> Wie Gregory Chaitin in mehreren Schriften behauptet. Vgl. <http://www.umcs.maine.edu/~chaitin/>. Eine inhaltliche Auseinandersetzung mit diesem Thema muss an dieser Stelle aus Platzgründen leider ausbleiben. Vgl. aber Bremer und Cohnitz [34], Raatikainen [271], Van Lambalgen [334], Raatikainen [270], Fallis [101].

<sup>294</sup> Für weitere Kritik am modalen Rationalismus, vgl. Yablo [361] und insbesondere Soames [302].

## 7.2.2 DIE PUTNAM ORTHODOXIE

Eine im Wesentlichen auf Putnams Arbeiten (Putnam [263], Putnam [265], Putnam [266]) zurückgehende Auffassung von der externen Bestimmung der Bedeutung eines Ausdrucks leugnet schlechterdings, dass es eine systematische Beziehung zwischen primären und sekundären Intensionen gibt, wie der modale Rationalismus sie voraussetzt. ‚Wasser‘ bezeichnet nicht deswegen  $H_2O$ , weil  $H_2O$  die „Rolle des wässrigen Stoffs in unserer Welt“ spielt, also  $H_2O$  eben dasjenige Zeug ist, was wir mit (DW) eindeutig bezeichnen, sondern schlicht und ergreifend, weil es  $H_2O$  war, was wir irgendwann einmal ‚Wasser‘ genannt haben.

Als dieser Ausdruck in die Sprache eingeführt wurde, um sich mit ihm auf eine bestimmte Substanz zu beziehen, wurde eine „kausale Referenzkette“ gestartet, an deren anderem Ende wir nun stehen. Die Vorstellung dabei ist, dass unsere Ausdrücke aus einer Tradition stammen, sich mit ihnen auf bestimmte Dinge zu beziehen. Diese Tradition hat irgendwann in ferner oder naher Vergangenheit dadurch begonnen, dass jemand den Ausdruck in die Sprache eingeführt hat, um sich auf etwas Bestimmtes zu beziehen. Wenn wir einen Ausdruck in einer Tradition verwenden, die kausal auf diesen Akt der Ausdruckseinführung (der „Taufe“) zurückgeführt werden kann, dann referiert unser Ausdruck immer noch auf dasselbe, worauf der Ausdruck bei der Taufe referierte. Welche Auffassungen über den Referenten (beispielsweise) des Ausdrucks ‚Wasser‘ die Mitglieder der verschiedenen Sprachgemeinschaftsgenerationen in der Zwischenzeit gehabt haben mögen, spielt für die Existenz einer solchen Referenzkette überhaupt keine Rolle.

Für Putnam scheint nichts dagegen zu sprechen, dass der erste Verwender von ‚Wasser‘ der Meinung war, es handle sich dabei um ein großes Lebewesen.<sup>295</sup> Es ist zwar so, dass wir ein Stereotyp haben müssen, das mit dem Rest der Sprachgemeinschaft übereinstimmt, damit uns die Sprachgemeinschaft zugesteht, dass wir einen Begriff erworben haben (vielleicht im Fall von ‚Wasser‘ so etwas wie (DW)). Dieses Stereotyp kann aber an der Natur von Wasser auch völlig vorbeigehen. Es muss sich dabei nicht um eine wahre Meinung, noch um eine in allen Einzelheiten universell geteilte handeln:

---

<sup>295</sup> Putnam drückt sich aber unklar aus. So legen seine Formulierungen in Putnam [265] nahe, dass wir einen Ausdruck *als* Ausdruck für eine Substanz bestimmter Art *intendieren* müssen, was ja bedeuten würde, dass wir zumindest so viel über den Referenten bereits wissen bzw. vermuten. So schreibt Putnam: „We picture the term “water” as becoming connected at some point in its history with the idea that substances possess a subvisible structure [...]. We picture “water” as acquiring a rigid use: as being used to denote whatever is substance-identical with (most of) the paradigms in our actual environment” (Putnam [265], 61). Diese Auffassung steht allerdings in Spannung mit den Bemerkungen in Putnam [263] zu Bleistiften, die sich als Organismen entpuppen können. Organismen und Artefakte haben aber sicherlich sehr unterschiedliche Identitätsbedingungen. Wir kommen hierauf in Kapitel 8.2.9 ausführlicher zu sprechen.

Nach dieser Auffassung wird von jemandem, der weiß, was „Tiger“ bedeutet (oder der, wie wir uns entschlossen haben zu sagen, das Wort „Tiger“ erworben hat), *verlangt* zu wissen, daß *stereotypische* Tiger gestreift sind. Genauer: Es gibt *ein* Tiger-Stereotyp (er mag noch andere haben), das die Sprachgemeinschaft als solches voraussetzt; es wird von ihm verlangt, daß er dieses Stereotyp hat und daß er im Prinzip weiß, daß diese Kenntnis obligatorisch ist. Und dieses Stereotyp muß das Merkmal des Gestreiftseins umfassen, wenn ihm der Erwerb des Wortes „Tiger“ bescheinigt werden soll.

Daß ein Merkmal (z.B. das Gestreiftsein) in dem mit einem Wort *X* verknüpften Stereotyp enthalten ist, heißt nicht, daß es eine analytische Wahrheit wäre, daß alle *Xs* oder überhaupt irgendwelche *Xs* dieses Merkmal aufweisen. Dreibeinige oder albinotische Tiger sind keine logisch unmöglichen Entitäten. Die Entdeckung, daß unser Stereotyp sich auf nicht normale oder nichtrepräsentative Elemente einer natürlichen Art berufen hat, ist nicht die Entdeckung eines logischen Widerspruchs. (Putnam [263], 68)

Einen Begriff erworben zu haben, bedeutet also, ein relevantes Stereotyp erworben zu haben. Dieses braucht weder aus wahren Verallgemeinerungen über den Referenten zu bestehen, noch aus Verallgemeinerungen, die sich als notwendige und hinreichende Bedingungen im Sinne einer Definition angeben lassen. Darüber hinaus braucht das Stereotyp auch nicht eindeutig sein: viele Sprecher des Deutschen können Buchen und Ulmen nicht unterscheiden, verfügen aber trotzdem über beide Begriffe. Dies gilt nicht nur für die Dinge, mit denen sich Naturwissenschaftler beruflich auseinandersetzen, sondern auch für solche, die eher in den Aufgabenbereich des Philosophen zu fallen scheinen, wie beispielsweise Personen. Auch für den Personenbegriff vermutet Putnam, dass er keine monokriteriale Verwendungsweise besitzt, d.h., dass er nicht synonym mit einer in hinreichenden und notwendigen Bedingungen ausdrückbaren Beschreibung gebraucht wird:

Es gibt da einen Irrenwitz, wo einer gerade aus der Irrenanstalt entlassen werden soll. Er wurde von den Ärzten auf Herz und Nieren geprüft und hat dabei einen geistig völlig gesunden Eindruck gemacht. So wollen sie ihn entlassen, und wie er geht, erkundigt sich einer der Ärzte noch beiläufig: „Was wollen Sie denn draußen werden?“ – „Ein Teekessel.“ Der Witz wäre unverständlich, wenn es buchstäblich unvorstellbar wäre, daß eine Person ein Teekessel sein könnte. (Putnam [263], 61)

Wie hieraus deutlich wird, läßt der Personenbegriff offenbar alles als Person zu, das Stereotyp wird also kaum als eine Menge von informativen notwendigen Bedingungen repräsentiert sein. Was folgt nun aus dieser Auffassung für das Gedankenexperiment? Wir wollen zunächst zwei Fragen unterscheiden, die man unabhängig voneinander behandeln kann. Die erste Frage ist, ob – falls unser Bedeutungswissen bei rigide verwendeten Ausdrücken in keinem systematischen Zusammenhang mit der „externen“ Bedeutung dieser Ausdrücke steht – sich aus Gedankenexperimenten etwas lernen läßt, wenn diese Gedankenexperimente rigide verwendete Ausdrücke zum Gegenstand haben (bzw. die Dinge zum Gegenstand haben, die durch rigide verwendete Ausdrücke rigide bezeichnet werden). Die

zweite Frage ist, ob – falls unser Bedeutungswissen häufig nicht als Menge von hinreichenden und notwendigen Bedingungen repräsentiert werden kann, sondern eher dem Modell eines Stereotypen entspricht – sich aus Gedankenexperimenten etwas lernen lässt, wenn diese Gedankenexperimente Ausdrücke zum Gegenstand haben, die nach dem Modell eines Stereotyp mental repräsentiert sind (bzw. die Dinge zum Gegenstand haben, die durch solche Ausdrücke bezeichnet werden).

In diesem Abschnitt wollen wir uns der ersten Frage widmen und die zweite Frage bis zum nächsten Kapitel zurückstellen, da sie von der Frage unabhängig ist, ob philosophische Probleme (manchmal) natürliche Arten betreffen und sich Gedankenexperimente deshalb (manchmal) verbieten. Insbesondere scheinen Spekulationen über die Repräsentationsstruktur unserer Begriffe die Möglichkeit von Bedeutungsanalysen als solche in Frage zu stellen und die Frage nach der Möglichkeit von Gedankenexperimenten gegen Sachanalysen nur indirekt zu betreffen.

Wie wir zu Anfang schon gesagt haben, ist die Philosophie des Geistes nicht das einzige Gebiet der systematischen Philosophie, in dem sich methodologische Dispute um die Zulässigkeit von Gedankenexperimenten entzündet haben. Mindest in der Erkenntnistheorie gibt es ebenfalls eine rege Diskussion darüber, ob philosophische Gedankenexperimente zuverlässige Verfahren darstellen, um philosophische Thesen zu diskutieren. Auch in der Erkenntnistheorie ist dabei die Auffassung geäußert worden, dass sie dies nicht können, sofern der Gegenstand der Erkenntnistheorie eine natürliche Art bildet:

In short, I see the investigation of knowledge and philosophical investigation generally, on the model of investigation of natural kinds.

This point is quite important, for what it means is that a good deal of the work involved in defining the subject matter under investigation is actually done by the world itself rather than the investigator. (Kornblith [167], 11)

Der Gegenstand eines Mineralogen – beispielsweise – ist die natürliche Gesteinsart, zu der die meisten der eingesammelten Proben gehören, ohne dass der sammelnde Wissenschaftler in irgendeiner Weise in der Lage sein muss, die „wesentlichen“ Eigenschaften dieser Mineralart angeben zu können. Was diese Eigenschaften sind, und welche der gesammelten Proben zu der Art gehören, hängt nicht von unserem Mineralogen und seinem Begriff von dieser Art ab. Es hängt von den Eigenschaften dieser Art ab, mutmaßlich von ihrer chemischen bzw. physikalischen Mikrostruktur, die der Mineraloge nicht kennen muss, von deren Existenz er vielleicht gar nichts weiß. Dementsprechend wird selbst der Gegenstandsbereich einer Wissenschaft nicht hauptsächlich durch die Überzeugungen der Wissenschaftler über ihren Gegenstandsbereich definiert, sondern durch die tatsächlichen Eigenschaften der Dinge in der Natur:

Subject matter is defined by way of connections with real kinds in the world, and what we regard as central or defining features does not determine the reference of our terms. (Kornblith [167], 12)

Genauso verhält es sich nach dieser Auffassung auch in der Erkenntnistheorie. Wissen ist ein natürliches Phänomen und sollte dementsprechend auch mit Methoden untersucht werden, mit denen man sonst auch natürliche Phänomene untersucht. Bei anderen natürlichen Arten sind wir nicht der Meinung, dass eine Reflexion auf unsere Intuitionen enorme Erkenntniszuwächse verspricht, stattdessen stellen wir Untersuchungen draußen in der Welt an.

Sofern wir starke Intuitionen in Bezug auf solche natürlichen Phänomene haben, handelt es sich um fehlbare a posteriori Überzeugungen, die wir durch die Alltagspsychologie mitbekommen haben. In einem vortheorietischen Stadium mag es sinnvoll sein, diese Überzeugungen, die vielleicht nur implizit in institutionalisierten Verhaltensweisen (inkl. Redeweisen) enthalten sind, durch Gedankenexperimente ans Tageslicht zu befördern. Etwa so, wie Wissen, das wir über physikalische Vorgänge haben, sinnvoll und berechtigt in naturwissenschaftlichen Gedankenexperimenten heraus*philosophiert* werden kann (man denke an Stevin und Galilei), können auch philosophische Gedankenexperimente unsere vortheorietischen Überzeugungen offen legen. Diese Überzeugungen sind dann so zuverlässig wie der Mechanismus, durch den sie zustande gekommen sind.

Dass wir bei solchen Intuitionen häufig einer Meinung sind, muss dabei kein Indikator für die Richtigkeit unserer gemeinsamen impliziten Hintergrundtheorie sein, sondern mag nur ein Indikator dafür sein, dass diese Hintergrundtheorie eben von vielen geteilt wird. Dennoch – solange keine bessere Theorie verfügbar ist, kann es in so einem vortheorietischen Stadium sinnvoll sein, Gedankenexperimente anzustellen und Intuitionen abzufragen. Ab einem bestimmten Zeitpunkt fällt diese Rechtfertigung für die „Innenschau“ aber weg:

At the same time, however, I do not think that this can be the whole story here, and this is where the difference between the practice of naturalists and that of anti-naturalists comes into play. If my account is correct, then what we ought to be doing is not just consulting the beliefs we already have, but more directly examining the external phenomena; only then would appeals to intuition be given what, on my view, is their proper weight. Thus, appeal to intuition early on in philosophical investigation should give way to more straightforwardly empirical investigations of external phenomena. (Kornblith [167], 15)

Sobald ein gewisses vortheorietisches Stadium verlassen ist, spielen imaginäre Fälle und unsere intuitiven Reaktionen auf diese keine bedeutende Rolle mehr – oder so sollte es zumindest sein. Falls unser Gegenstand eine natürliche Art ist, sind unsere Intuitionen in doppelter Hinsicht unzuverlässige Informanten:

[T]he appeal to imaginary cases and what we are inclined to say about them is both overly narrow and overly broad in its focus. It is overly narrow because serious empirical investigation of a phenomenon will often reveal possibilities which we would not, and sometimes could not, have imagined before. It is overly broad because many imaginable cases are not genuine possibilities and need not be accounted for by our theories. We might be able to imagine a rock with a certain combination of color, hardness, malleability, and so on, and such a rock, were it to exist, might be difficult or impossible to fit into our current taxonomy. But this raises no problem

at all for our taxonomic principles if the imagined combination of properties is nomologically impossible. On the naturalistic view, the same may be said for testing our empirical views against merely imaginable cases. (Kornblith [167], 16)

Folgt man der Auffassung, dass (a) Philosophie nicht in der Untersuchung unserer Begriffe von bestimmten Phänomenen besteht, sondern ausschließlich die Phänomene selbst zum Gegenstand hat, (b) die Phänomene (mutmaßlich) natürliche Arten bilden, und es der Fall ist, dass (c) die vortheoretische Phase der wissenschaftlichen Beschäftigung mit diesen Phänomenen verlassen ist und unsere Hintergrundüberzeugungen keine fruchtbare Informationsquelle mehr sind, dann scheint man Gedankenexperimente in der Tat nur dadurch rechtfertigen zu können, dass man die dahinter stehende kausale Referenztheorie angreift.<sup>296</sup>

Nun sind die Annahmen (a) und (c) aber nicht so offensichtlich, wie Kornblith und die anderen Vertreter der Putnam Orthodoxie uns Glauben machen wollen, weshalb sich noch eine dritte Position angeben lässt, die die Rolle von Gedankenexperimenten in einer Philosophie, deren Hauptfragen natürliche Arten betreffen, anders definiert.

### 7.2.3 MODERATER DESKRIPTIVISMUS

Ob Wissen eine natürliche Art bildet, kann hier nicht diskutiert werden<sup>297</sup>, genauso wenig wie die Frage, ob Personen eine natürliche Art sind, oder ob es sich bei Bewusstsein um eine natürliche Art handelt. Die Frage von Kapitel 7.2 ist ja, was, *wenn* diese Dinge natürliche Arten sind? Sollten Gedankenexperimente dann aus der Philosophie ausgeschlossen werden?

Selbst die radikale Position, die wir im vorhergehenden Abschnitt kennen gelernt haben, würde diese Konsequenz nicht behaupten. Nach der Putnam Orthodoxie können Gedankenexperimente durchaus fruchtbar sein, ebenso, wie sie dies in den Naturwissenschaften auch sein können, wobei die Frage, wie und warum sie das sein können, so zu beantworten wäre, wie wir dies in Kapitel 3 getan haben. Aber vielleicht kann man noch etwas mehr sagen.

Laut Kornblith „definiert“ der Wissenschaftler, der natürliche Arten untersucht, seinen „Gegenstand“ nur partiell. Damit ist gemeint, dass ein Wissenschaftler, der vor einem großen Meer steht, mit dem Finger auf die Wasseroberfläche deutet und verkündet ‚Ich möchte wissen, was das da genau ist.‘, ein Forschungsproblem formuliert hat, das letztlich *nur* durch die Aufdeckung der Eigenschaften von H<sub>2</sub>O

<sup>296</sup> Eine Alternativposition könnte beispielsweise der im vorherigen Abschnitt erläuterte modale Rationalismus darstellen, oder vielleicht eine Variante der Transzendentalphilosophie (vgl. Misselhorn [211]). Eine solche Diskussion kann und soll hier nicht geleistet werden. Wie schon gesagt, gehört die Diskussion um den modalen Rationalismus zu einem Gebiet, das gegenwärtig fast unüberschaubar ist, und die Auseinandersetzung mit der Transzendentalphilosophie würde eine Grundlagendiskussion erfordern, die ebenfalls den hier gesteckten Rahmen sprengen würde.

<sup>297</sup> Vgl. Goldman [126], Goldman und Pust [127], Kornblith [167].

gelöst werden kann, unabhängig davon, was unser Wissenschaftler in dem Moment über H<sub>2</sub>O oder Molekülstrukturen im Allgemeinen weiß. Dass dem so ist, liegt daran, dass ‚das da‘ eben auf Wasser verweist und Wasser eben H<sub>2</sub>O ist. Der Wissenschaftler hätte sich natürlich auch für andere Dinge interessieren können, weil er sich aber für Wasser interessiert, ist nun durch die Art des Beschaffenseins des Untersuchungsgegenstandes vorgegeben, wie das Problem am sinnvollsten anzugehen und letztlich zu lösen ist.<sup>298</sup>

Es mag sein, dass dieses Bild für bestimmte Wissenschaftszweige tatsächlich zutrifft, und dass man mit einer kausalen Referenztheorie deshalb auch gut erklären kann, wie es möglich ist, dass Wissenschaftler sich vor 200 Jahren schon mit denselben Dingen wie heutige Wissenschaftler beschäftigt haben, obwohl sie radikal andere Vorstellungen davon hatten, was diese Dinge sind.

‚Defining the subject‘ bedeutet aber nicht nur, dass man den Gegenstand identifiziert, mit dem man sich wissenschaftlich auseinandersetzen will, sondern auch, dass man mit bestimmten Fragen an diesen Gegenstand herantritt, die man beantwortet haben möchte. Welche Fragen man beantwortet haben möchte, ist nicht (nur) durch die Welt determiniert, sondern tatsächlich zunächst etwas, was der Wissenschaftler für sich festlegen kann. Typischerweise sind philosophische Probleme nicht von der ‚Was ist das da?‘-Art, sondern beziehen sich auf Sachverhalte, Zusammenhänge oder Dinge, die in einem relativ ungeklärten kausalen Verhältnis zu uns stehen, so dass ohne Weiteres eben *nicht* offensichtlich ist, welche Untersuchungsmethoden herangezogen werden müssen, um das fragliche Problem zu lösen. Es ist gerade die Frage, ob die philosophische Frage danach, was Bewusstsein ist, durch die Neurowissenschaften beantwortet wird, bzw. ob die Fragen der Erkenntnistheorie durch die Psychologie beantwortet werden.

Während es in Bezug auf die ‚Was ist das da?‘-Frage vor dem großen Ozean halbwegs klar ist, dass ein erheblicher Teil der Antwort aus der Chemie kommen wird, ist es eben *nicht* ausgemacht, dass die Psychologie unsere Fragen in der Erkenntnistheorie, die Biologie unsere Fragen in der Debatte um personale Identität, oder die Neurowissenschaften unsere Fragen in Bezug auf das Leib-Seele-Problem beantwortet. Hier spielen Gedankenexperimente offenbar eine besonders prominente Rolle. Obwohl es bei diesen „philosophischen Problemen“ um mutmaßliche natürliche Arten „geht“, untersucht die Philosophie dennoch zunächst unsere Begriffe. Selbst für ein naturalistisches Verständnis vom „Gegenstand“ der Philosophie kommt es darauf an, zu klären, ob das, was wir mit ‚Bewusstsein‘ meinen, das ist, was von den Neurowissenschaften erklärt wird, um dann beurteilen zu können, ob das Leib-Seele-Problem in den Neurowissenschaften am Besten aufgehoben ist.

Eine solche Auffassung wertet das Zombie-Gedankenexperiment dementsprechend nicht als den schlagenden Nachweis, dass eine bestimmte metaphysische Auffassung (Physikalismus) falsch ist, sondern als Hinweis darauf, dass es eine Diskrepanz gibt zwischen dem, was wir als Antwort auf unsere Frage nach dem We-

---

<sup>298</sup> Mutmaßlich durch chemische Untersuchungen und die Analyse von Molekülstrukturen.

sen des Bewusstseins erwarten, und dem, was uns die Neurowissenschaften als Antwort liefern.<sup>299</sup>

Betrachten wir hierzu die Anforderungen an wissenschaftliche Erklärungen, die durch das Hempel-Oppenheim Schema, bzw. seine Nachfolgekonzeptionen formuliert werden.<sup>300</sup> Natürlich kann es pragmatische Umstände geben, unter denen wir auch solche wissenschaftliche Erklärungen akzeptieren, bei denen es weder einen semantischen noch einen ontologischen Zusammenhang zwischen Explanans und Explanandum gibt. Selbst solche pragmatisch zufrieden stellenden Erklärungen sind in ihrem *explanativen* Wert aber immer nur so gut, wie sie uns über den so genannten „idealen Erklärungstext“ informieren. Bei diesem handelt es sich um

an inter-connected series of law-based accounts of all the nodes and links in the causal network culminating in the explanandum, complete with a fully detailed description of the causal mechanisms involved and theoretical derivations of all the covering laws involved. This full-blown causal account would extend, via various relations of reduction and supervenience, to all levels of analysis, i.e., the ideal text would be closed under relations of causal dependence, reduction, and supervenience. (Railton [272], 247)

Die wichtigste Eigenschaft dieser Erklärungstexte ist aber ihre ausnahmslos *deduktive* Struktur. Entweder wird das Explanandum deduktiv aus dem Explanans abgeleitet, oder zumindest wird eine Aussage über die (objektive, irreduzible) Wahrscheinlichkeit des Explanandums aus dem Explanans abgeleitet (beispielsweise (und vermutlich ausschließlich) in der Quantenmechanik). In beiden Fällen muss das Explanandum-Vokabular in einer *semantischen* Enthaltenseins-Beziehung zum Explanans-Vokabular stehen, damit der geforderte deduktive Zusammenhang bestehen kann. Auch wenn die tatsächliche Angabe eines solchen Erklärungstextes nie gefordert ist, ist er eine regulative Idee, an der Begriffe wie ‚Erklärungsstärke‘ etc. erläutert werden können. Eine „vollständige“ wissenschaftliche Erklärung bestünde in der Angabe eines solchen Textes. Wäre ein solcher Text prinzipiell in Bezug auf bestimmte erlaubte Explanans-Vokabulare und bestimmte Explananda ausgeschlossen, wäre eine *vollständige* wissenschaftliche Erklärung in diesem Sinne nicht möglich. In einem solchen Fall wollen wir von einer *Erklärungslücke* sprechen. Betrachten wir unter dieser Perspektive nochmals Jacksons Gedankenexperiment mit der Neurophysiologin Mary:

[...] Jackson's story about Mary does show something important about Mary's epistemic situation; in particular, her ability to explain qualia in physical terms. For if

---

<sup>299</sup> Ein Beispiel für eine solche Auffassung ist Nagel [221].

<sup>300</sup> Um an dieser Stelle keine Debatte über die richtige Explikation ‚wissenschaftlicher Erklärung‘ anzubrechen, gehe ich von der erweiterten „idealen Erklärungstext“-Konzeption nach Railton [272] aus, die ich in Cohnitz [68] erläutere und verteidige. Joseph Levine hat gezeigt, dass sich die folgenden Überlegungen auch auf eine ontische (Salmon) wie auch eine vereinheitlichungstheoretische (Kitcher, Friedman) Konzeption übertragen lassen. Vgl. Levine [184].

Mary could really explain the character of sensory experiences by reference to the underlying physical process, then it seems that she shouldn't learn anything new when she finally experienced red for herself. She should have expected it to be like that. The fact that it seems so clear that she would learn what it's like to experience red is testimony to the explanatory gap that separates physical theory and conscious experience. (Levine [184], 77)

Es besteht (falls das Mary-Gedankenexperiment und das Zombie-Gedankenexperiment irgendetwas zeigen) eben eine Lücke zwischen dem, was wir als Erklärung erwarten, und dem, was wir als Erklärung von einer bestimmten Wissenschaft bekommen. *Ob* solche Lücken bestehen, hängt in entscheidender Weise davon ab, wie wir *unsere* „philosophischen“ Fragen verstehen, die wir in *unseren* Begriffen formulieren. Es hängt davon ab, wie das Explanandum formuliert ist.

Dabei ist eine solche Position keineswegs dazu verpflichtet anzunehmen, dass unsere Begriffe unwandelbar sind und scharf von unserem (empirischen und damit synthetischen) Hintergrundwissen getrennt werden können.<sup>301</sup> Es kann durchaus sein, dass sich Erklärungslücken im weiteren Verlauf einer Wissenschaftsgeschichte schließen.<sup>302</sup>

Es kann aber auch sein, dass ein bestimmtes Naturalisierungsprogramm Aspekte außer acht lässt, die – auf dem Hintergrund von Begriffsanalysen – relativ klar benannt werden können und vielleicht durch ein anderes, erweitertes Programm behandelt werden können, dass Erklärungslücken also nicht bloße Artefakte unserer Sprachverwendung sind, sondern einen ganz realen Kern haben.

So wird der naturalisierten Erkenntnistheorie in der Nachfolge von Quine von manchen Autoren vorgeworfen, dass wir von einer Erkenntnistheorie eine normative Komponente erwarten (sie soll Auskunft darüber geben, was man tun *soll*, um zu Überzeugungen zu gelangen), Quines Programm eine solche Komponente aber nicht enthält (die Psychologie klärt uns nur darüber auf, wie wir *de facto* zu Überzeugungen gelangen). Neuere Naturalisierungsprogramme versuchen diesen normativen Aspekt einzuholen, indem sie beispielsweise die Frage danach, was man tun *soll*, um zu Überzeugungen zu gelangen, durch die ameliorative Psychologie beantworten<sup>303</sup> (wobei es jetzt nicht darum geht, ob dieses neue Programm problemfrei ist, oder auch nur dem erforderlichen normativen Aspekt gerecht wird – das muss an anderer Stelle erörtert werden). In diesem Fall konnte die Begriffsanalyse davon *überzeugen*, dass ein Naturalisierungsprogramm zu kurz griff und die eigentliche Fragestellung (in diesem Fall die Fragestellung der Erkenntnistheorie) nicht abdeckte.

<sup>301</sup> Vgl. zu dieser Frage McGinn [206], Papineau [249], Levine [184], Chalmers [55].

<sup>302</sup> Diese Hoffnung hegt beispielsweise Pauen [251] in Bezug auf die Physikalismus-Debatte in der Philosophie des Geistes.

<sup>303</sup> Vgl. zu dieser Problematik Bishop und Trout [26], alle Beiträge in Kornblith [166], Kornblith [167], Stich [317].

### 7.3 KRITIK AN GEDANKENEXPERIMENTEN GEGEN BEDEUTUNGSANALYSEN

Die in Kapitel 7.2 dargestellte (durchaus naturalistische) Auffassung ist mit der Möglichkeit, dass es sich bei den Gegenständen der Philosophie um natürliche Arten handelt, und der Auffassung, dass philosophische Probleme Sachprobleme und keine Sprachprobleme sind, absolut vereinbar. Ihrzufolge tragen Gedankenexperimente substantiell zur Lösung philosophischer Probleme bei, weil sie uns darüber Auskunft geben, wie *wir* die Begriffe verstehen, in denen *unsere* philosophischen Fragestellungen formuliert sind. Eine Aufgabe, die offenbar nicht trivial ist – wie die Philosophiegeschichte zeigt – und die auch intuitiv sinnvoll erscheint. Antworten sind eben so gut, wie sie sich auf die gestellten Fragen beziehen und diese klären. Was aber die gestellte Frage jeweils ist, hängt in entscheidender Weise von unseren Interessen, unseren Erwartungen und unserem Vorverständnis ab.<sup>304</sup> Mit dieser Fragestellung versuchen wir nun nicht mehr Sachanalysen mit Hilfe von Gedankenexperimenten zu kritisieren, sondern befinden uns offensichtlich im Bereich der Bedeutungsanalyse.

Dennoch kann man diese Position aus naturalistischer Perspektive auf mindestens zwei Weisen in Frage stellen: (1.) Was heißt hier ‚uns‘ und ‚wir‘? Die soeben geschilderte Position ist doch letztlich nur die Philosophie wohlhabender Weißer (vgl. Weinberg, Nichols und Stich [338]). (2.) Was heißt hier ‚Begriffe‘? Geht diese Position nicht von einer sehr naiven Vorstellung davon aus, welche internen Repräsentationen mit Bedeutungen verbunden sind (vgl. Gendler [119], Gendler [121])?

#### 7.3.1 DIE RELATIVITÄT VON INTUITIONEN

Um die Karten gleich auf den Tisch zu legen: falls unsere Philosophie „die Philosophie wohlhabender Weißer“ ist, ist gar nicht klar, was daran so schlimm sein sollte. Obwohl das auf den ersten Blick (und vermutlich auch auf den zweiten Blick) nicht wirklich nach *political correctness* aussieht, ist diese Position, wie wir sehen werden, völlig untadelig. Aber gehen wir das Problem in behutsamen Schritten an.

In *From Metaphysics to Ethics* beschreibt Frank Jackson (wie wir schon erläutert haben) zunächst<sup>305</sup> eine vergleichsweise unproblematische Position in Bezug auf die Rolle von Bedeutungsanalysen in der Philosophie, die in etwa dem zu entspre-

---

<sup>304</sup> ... unter der Annahme, dass *wir* diejenigen sind, die die Frage gestellt haben.

<sup>305</sup> Das ist die Position, die Jackson zunächst als „moderate Metaphysik“ anpreist, die sich aber nachher als modaler Rationalismus entpuppt. Im Folgenden soll es nur um die im vorhergehenden Abschnitt charakterisierte Position gehen, nicht um die Enthaltenseins-These.

chen scheint, was wir im letzten Abschnitt als moderaten Deskriptivismus bezeichnet haben. Auch nach dieser Position geht es zunächst nur darum, die Begriffe der Alltagstheorie zu rekonstruieren, um diese gegebenenfalls mit Begriffen anderer Theorien zu vergleichen.

I am sometimes asked – in a tone that suggests that the question is a major objection – why, if conceptual analysis is concerned to elucidate what governs our classificatory practice, don't I advocate doing serious opinion polls on people's responses to various cases? My answer is that I do – when it is necessary. Everyone who presents the Gettier case to a class of students is doing their own bit of fieldwork, and we all know the answer they get in the vast majority of cases. But it is also true that often we know that our own case is typical and so can generalize from it to others. It was surely not a surprise to Gettier that so many people agreed about his cases. (Jackson [158], 36-37)

Obwohl Jackson hier eine Auffassung wiedergibt, die durchaus von vielen geteilt wird, nämlich, dass die Gettier-Fälle nahezu universell überzeugend seien, ist von empirischer Seite eingewandt worden, dass sie so *universell* überzeugend nicht sind, und dass auch andere „Lieblingskinder“ der gedankenexperimentellen Begriffsanalyse in Bezug auf ihre Universalität überschätzt werden.

Eine der in diesem Zusammenhang bekanntesten Untersuchungen ist die von Weinberg, Nichols und Stich (Weinberg, Nichols und Stich [338]), sowie die Nachfolgeuntersuchung von Machery, Mallen, Nichols und Stich (Machery, Mallen, Nichols und Stich [196]). Da beide Untersuchungen noch recht jungen Datums sind, sollen sie in einiger Ausführlichkeit dargestellt werden.

Psychologische Untersuchungen, die ein Psychologenteam unter der Leitung von Richard Nisbett durchgeführt hat, deuten darauf hin, dass Ostasiaten (Chinesen, Koreaner und Japaner) eher „holistisch“ denken, während Westler (Amerikaner europäischer Abstammung) eher „analytisch“ denken.<sup>306</sup>

Untersuchungen, die von einem Team unter der Leitung von Jonathan Haidt<sup>307</sup> in Bezug auf moralische Intuitionen durchgeführt wurden, deuten darauf hin, dass diese Intuitionen signifikant mit der sozioökonomischen Stellung der Befragten variieren. Diese beiden Befunde nehmen Weinberg et al. zum Ausgangspunkt, ähnliche Untersuchungen in Bezug auf epistemische Intuitionen anzustellen. Könnte es nicht sein, dass unsere intuitiven Reaktionen in Bezug auf erkenntnistheoretische Gedankenexperimente ebenfalls variieren, je nachdem, aus welchem Kulturkreis wir stammen bzw. welchen sozioökonomischen Status wir haben? Die Untersuchung von Weinberg et al. bezog sich auf folgende Hypothesen:

(H1) Epistemische Intuitionen variieren zwischen Kulturen.

---

<sup>306</sup> Vgl. dazu Nisbett, Peng, Choi und Norenzayan [235] und Nisbett [234]. Nisbett vermutet „tiefe“ geistesgeschichtliche Ursachen. Die These, dass kognitive Strukturen kulturabhängig sind, ist nicht neu und wurde in den 80er Jahren auf der Grundlage linguistischer Untersuchungen bereits diskutiert. Vgl. Bloom [28].

<sup>307</sup> Vgl. Haidt, Koller und Dias [138].

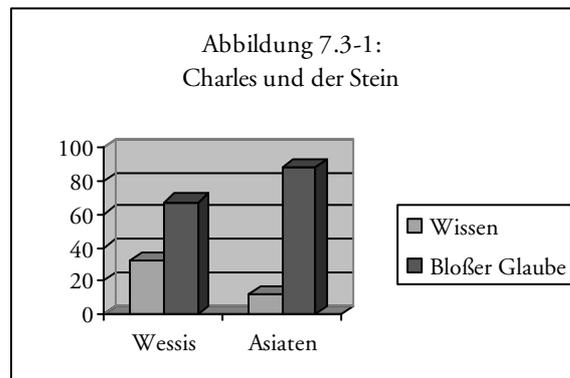
(H2) Epistemische Intuitionen variieren zwischen sozioökonomischen Gruppen.

In einer ersten Versuchsreihe wurden die epistemischen Intuitionen von Ostasiaten mit denen von Westlern verglichen.<sup>308</sup> Zu diesem Zweck präsentierte man den Versuchspersonen kurze Gedankenexperimente<sup>309</sup>, wie etwa das folgende, das einem Gedankenexperiment aus der erkenntnistheoretischen Internalismus/Externalismus-Debatte entspricht:

One day Charles is suddenly knocked out by a falling rock, and his brain becomes re-wired so that he is always absolutely right whenever he estimates the temperature where he is. Charles is completely unaware that his brain has been altered in this way. A few weeks later, this brain-rewiring leads him to believe that it is 71 degrees in his room. Apart from his estimation, he has no other reason to think that it is 71 degrees. In fact, it is at that time 71 degrees in his room. Does Charles really know that it was 71 degrees in the room, or does he only believe it? (Weinberg, Nichols und Stich [338], 439)

Wie man in Abbildung 7.3-1 sehen kann, waren die Befragten beider Kulturkreise mehrheitlich der Meinung, dass es sich bei Charles' Überzeugung nicht um Wissen handelt, die Ostasiaten („Asiaten“) waren allerdings in signifikant größerem Ausmaß dieser Meinung als die Amerikaner europäischer Abstammung („Wessis“).

Weinberg et al. untersuchten daraufhin, ob sich durch eine Veränderung der im



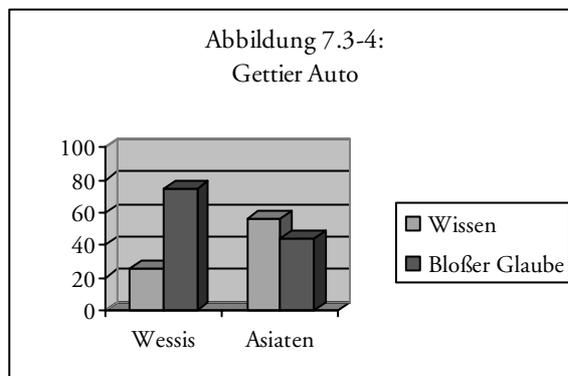
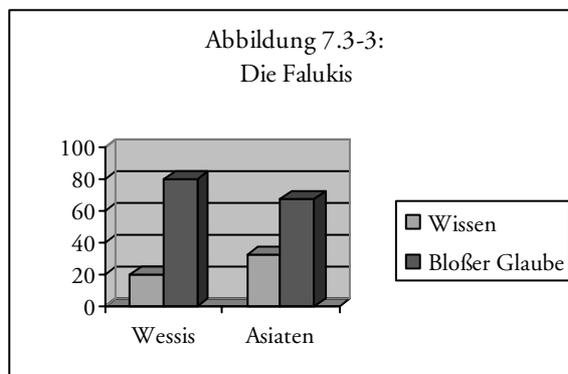
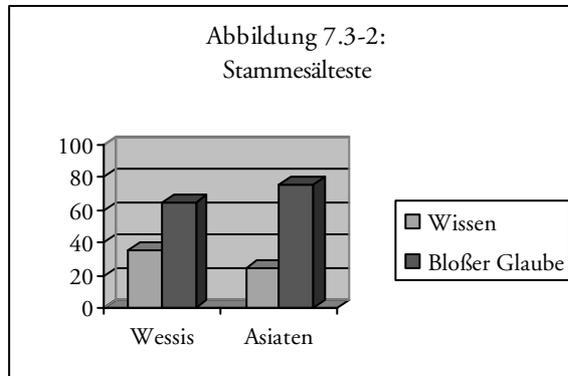
Gedankenexperiment geschilderten Geschichte dieses Verhältnis beeinflussen lässt. Zunächst änderten sie die Geschichte insofern ab, als der Protagonist seine neue Fähigkeit nicht dadurch erhält, dass er durch einen Stein getroffen wird, sondern dadurch, dass – ohne, dass er es weiß – die Dorfältesten

beschlossen haben, dass sein Gehirn von Wissenschaftlern so umgestaltet wird, dass er zuverlässig die genaue Temperatur einschätzen kann. Wieder waren sich Wessis und Asiaten mehrheitlich einig, dass es sich auch hierbei nicht um Wissen handelt, allerdings verschwand bei dieser Geschichte der signifikante Unterschied zwischen den beiden Gruppen, wie in Abbildung 7.3-2 zu erkennen ist.

Durch eine weitere Geschichte, in der nicht ein einzelnes Individuum mit der Fähigkeit beglückt wird, die Temperatur zuverlässig einschätzen zu können, son-

<sup>308</sup> Die Untersuchung wurde an Studierenden (undergraduates) der Universität Rutgers durchgeführt.

<sup>309</sup> Im Sinne von  $T_3$  aus Kapitel 3.1.



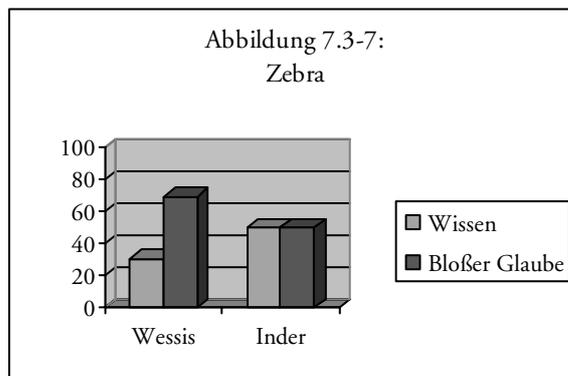
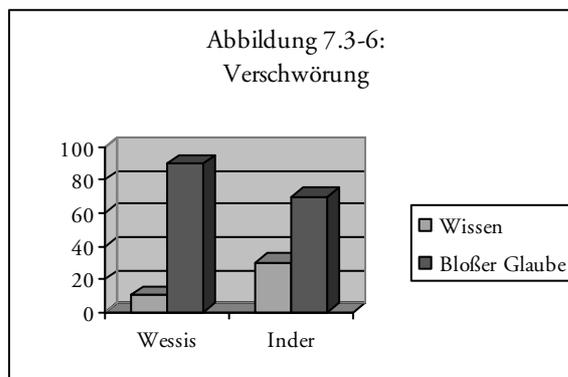
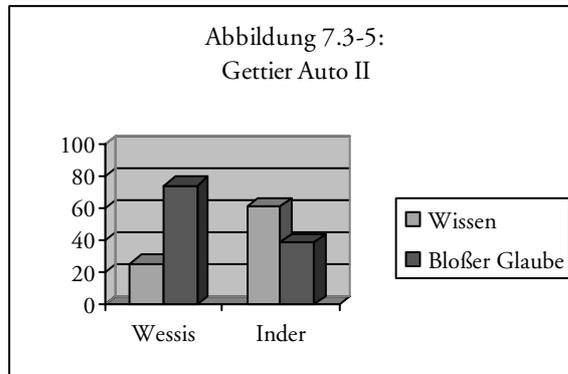
pistemische Gesamtsituation aber Situationen *ähnelt*, bei denen das Subjekt Wissen hat. Von solchen Überlegungen ausgehend, prognostizierten Weinberg et al., dass Asiaten Gettier-Fälle anders beurteilen als Wessis. Wie Abbildung 7.3-4 zeigt, ist die Beurteilung des Gedankenexperiments sogar mehrheitlich umgekehrt.

In dem zu beurteilenden Gedankenexperiment ging es darum, dass Bob glaubt, dass Jill einen amerikanischen Wagen fährt, *weil* er glaubt, dass Jill einen Buick fährt. Jill hat aber den Buick gegen einen Pontiac getauscht. Ist Bobs (wahre) Ü-

dem eine gesamte Inselpopulation, die „Faluki“, durch einen radioaktiven Meteor in diese Lage versetzt wird, konnte das ursprüngliche Verhältnis zwischen Wessis und Asiaten sogar umgedreht werden (allerdings nicht statistisch signifikant), wie Abbildung 7.3-3 zeigt.

Diese Resultate stehen im Einklang mit den Hypothesen von Nisbett et al., dass Asiaten ein stärkeres Gewicht auf die Harmonie einer Gemeinschaft legen als die eher individualistisch denkenden Wessis.

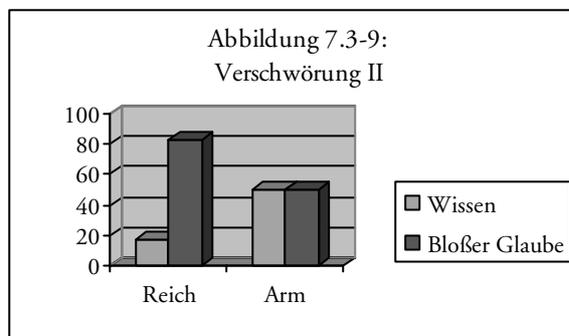
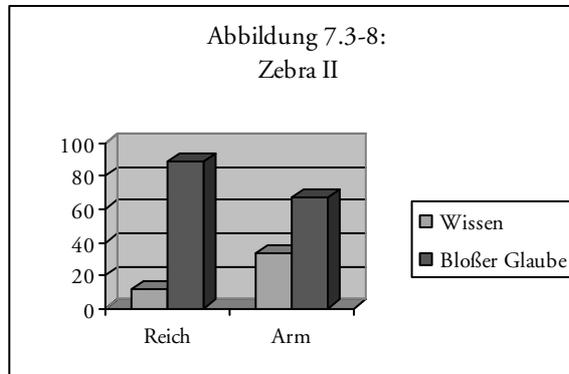
Nisbett et al. hatten außerdem vermutet, dass Asiaten stärker auf Ähnlichkeiten achten als Wessis und dass Wessis stärker kausale Verbindungen betonen als Asiaten. Bei den klassischen Gettier-Fällen, die wir zu Anfang dieses Kapitels besprochen haben, ist es häufig so, dass eine wahre Überzeugung durch eine falsche Überzeugung *kausal* hervorgerufen wurde, die e-



berzeugung, dass Jill einen amerikanischen Wagen fährt, Wissen? Motiviert durch diese Befunde, allerdings ohne irgendeine weitergehende theoretische Fundierung, verglichen Weinberg et al. die gemessenen Wessi-Reaktionen auf dieses Gedankenexperiment mit den Reaktionen, die Personen zeigen, die vom Indischen Subkontinent abstammen. Auch die „Inder“ zeigten mehrheitlich die umgekehrte Bewertung des Falls (Abbildung 7.3-5).

Inder zeigten darüber hinaus signifikant abweichende Beurteilungen anderer Fälle, bei denen zwischen Asiaten und Wessis kein signifikanter Unterschied festgestellt werden konnte. So beurteilten die Inder das Verschwörungsgedankenexperiment (7.3-6), wie auch das Zebra-Gedankenexperiment (7.3-7) signifikant weniger eindeutig wie die befragten Wessis.

Im Verschwörungsgedankenexperiment geht es um „Jim“, der auf der Grundlage veröffentlichter Untersuchungsergebnisse glaubt, dass die Einnahme von Nikotintabletten bei gleichzeitiger Rauchabstinenz die Wahrscheinlichkeit von Krebs nicht erhöht. Jim ist sich allerdings nicht im Klaren darüber, dass es sein könnte, dass die Untersuchungsergebnisse, auf die er seine Überzeugung stützt, von der Tabakindustrie lanciert wurden und eigentlich nicht stimmen. Die Tabakindustrie hat diese Untersuchungsergebnisse nicht lanciert. Handelt es sich bei Jims Überzeugung um Wissen?



Bei dem Zebra-Gedankenexperiment geht es um einen ähnlich gelagerten Fall, bei dem der junge Vater „Mike“ mit seinem Nachwuchs den Zoo besucht und seinem Sohn ein Zebra zeigt. Was Mike nicht weiß – was aber den Älteren in Mikes Gesellschaft bekannt ist: man kann Maultiere so als Zebras verkleiden, dass man den Unterschied kaum feststellen kann. Auch Mike kann den Unterschied nicht feststellen. Mike zeigt seinem Sohn tatsächlich gerade ein Zebra. *Weiss* Mike, dass es sich um eins handelt?

Insoweit scheinen die empirischen Ergebnisse tatsächlich (H1) zu bestätigen. Die Verteilung der Intuitionen in Bezug auf erkenntnistheoretische Gedankenexperimente variiert zwischen Kulturen.

Weinberg et al. sind bei diesem Ergebnis nicht stehen geblieben, sondern haben nach den Befunden von Haidt et al. auch vermutet, dass epistemische Intuitionen, so wie moralische Intuitionen, vom sozioökonomischen Status der Befragten abhängen können. Von hohem sozioökonomischen Status („Reich“) war eine Versuchsperson dann, wenn sie angab mindestens für ein Jahr ein College besucht zu haben, wer nie ein College besucht hatte, wurde als von niedrigem sozioökonomischen Status betrachtet („Arm“).

Eins der getesteten Gedankenexperimente war eine Variante des soeben geschilderten Zebra-Falls, wobei aber in der Geschichte nicht darauf hingewiesen wurde, dass die Älteren in Mikes Gesellschaft (im Gegensatz zu Mike) von den Täuschungsmöglichkeiten wissen, sondern nur gesagt wurde, dass der fragliche Betrachter von seiner Beobachtungsposition aus ein verkleidetes Maultier von einem Zebra nicht habe unterscheiden können. Dieses Gedankenexperiment führte zu signifikanten Abweichungen in der anteilmäßigen Beurteilung zwischen den beiden Gruppen (Abbildung 7.3-8).

Ein ähnlich eindeutiges Ergebnis erzielte der Versuch mit der Krebs-Verschwörungsgeschichte, die auch zu unterschiedlichen Beurteilungsproportionen zwischen Wessis und Indern geführt hatte (Abbildung 7.3-9).<sup>310</sup> Diese Ergebnisse stützen damit auch (H2), da sie darauf hindeuten, dass die Beurteilung von Gedankenexperimenten vom sozioökonomischen Status der Befragten abhängen kann.

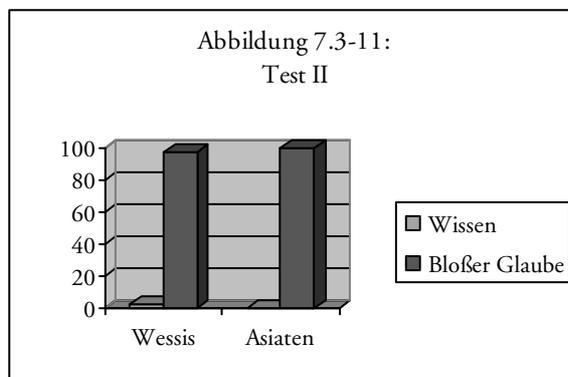
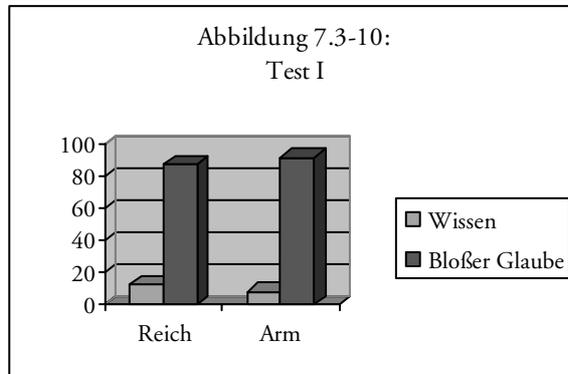
*Was folgt daraus?*

Eine nahe liegende Replik auf diese Ergebnisse besteht in dem Hinweis, dass das Wort ‚Wissen‘ im Alltag unterschiedliche Verwendungen hat. Zumindest im Englischen kann ‚know‘ manchmal auch nur starke subjektive Gewissheit ausdrücken. In diesem Sinne kann man auch etwas ‚wissen‘, wenn die fragliche Überzeugung falsch ist. Dabei handelt es sich nach allgemeinem Verständnis nicht um einen Hinweis darauf, dass Wahrheit keine notwendige Bedingung für Wissen ist, sondern einfach um eine weitere Bedeutung des Wortes ‚to know‘. Um ein solches Missverständnis auszuschließen, also um festzustellen, ob die Befragten ‚Wissen‘ im erkenntnistheoretisch relevanten Sinne verstehen, wurde außerdem folgende Geschichte getestet:

Dave likes to play a game with flipping a coin. He sometimes gets a “special feeling” that the next flip will come out heads. When he gets this “special feeling,” he is right about half the time, and wrong about half the time. Just before the next flip, Dave gets that “special feeling,” and the feeling leads him to believe that the coin will land heads. He flips the coin, and it does land heads. Did Dave really know that the coin was going to land heads, or did he only believe it? (Weinberg, Nichols und Stich [338], 450)

---

<sup>310</sup> Die folgende Abbildung orientiert sich an dem Histogramm in Weinberg, Nichols und Stich [338], S. 448 und nicht, wie die vorhergehenden Abbildungen, an den Untersuchungsergebnissen, die sich bei Weinberg, Nichols und Stich [338] im Anhang finden. Laut den Zahlen im Anhang würden die Armen den Verschwörungsfall genau umgekehrt beurteilen (ca. 85% für ‚Wissen‘). Vermutlich handelt es sich bei den Zahlenwerten im Anhang aber um einen Tippfehler. Da die Gesamtzahl der Armen beim Zebra-Fall bei 24 liegt, wird sie vermutlich auch beim Verschwörungsfall bei 24 und nicht bei 14 liegen; das Ergebnis war also 12:12, was 50% für ‚Wissen‘ entspricht und mit dem Histogramm auf S. 448 sowie mit der Tatsache übereinstimmt, dass Weinberg et al. diese ansonsten überraschend radikale Abweichung nicht entsprechend kommentieren.



Sowohl im Vergleich zwischen Wessis und Asiaten, wie auch im Vergleich zwischen Reichen und Armen waren sich die Gruppen einig, wie ‚Wissen‘ verstanden wird (vgl. Abbildung 7.3-10, 7.3-11).

Außerdem zeigen die Untersuchungen von Machery et al. (Machery, Mallon, Nichols und Stich [196]), dass diese kulturell bedingten Abweichungen kein Resultat eines möglicherweise falsch verstandenen Wissensbegriffs sind, sondern sich auch bei *semantischen*<sup>311</sup> Intuitionen bemerkbar machen.

Wir hatten bereits gesagt, dass neben den Ge-

tier-Fällen die Gedankenexperimente zur kausalen Referenztheorie zu den Lieblingskindern der Philosophie gehören. Wie wir vor einigen Abschnitten erläutert haben, war – zum Beispiel – Kathy Wilkes ja der Meinung, dass diese Gedankenexperimente wesentlich weniger kontrovers sind als die meisten anderen Gedankenexperimente in der Philosophie, und nahm dieses Datum zum Anlass, eine Theorie zu entwickeln, die den Kripke-Putnam Gedankenexperimenten eine epistemische Sonderrolle zuordnete. Dass diese Theorie nicht viel taugt, haben wir gezeigt. Dennoch scheint Wilkes (und nicht nur sie) davon auszugehen, dass die Gedankenexperimente zur kausalen Referenztheorie zu universell geteilten Resultaten geführt haben, im Gegensatz zu manchen anderen Gedankenexperimenten der Philosophie.

Aber auch diese Vermutung, dass die Gedankenexperimente zur kausalen Referenztheorie zu universell geteilten Intuitionen führen, wird durch neuere empirische Untersuchungen stark in Zweifel gezogen. Machery et al. haben Versuchspersonen dabei wieder Gedankenexperimente vorgelegt und deren intuitive Reaktionen getestet. Im Zentrum standen dabei Gedankenexperimente, die von Kripke gegen eine deskriptivistische Theorie von Eigennamen vorgebracht wurden.

<sup>311</sup> ‚Semantisch‘ im Sinne von ‚auf den Bedeutungsbegriff bezogen‘.

Nach der deskriptivistischen Theorie der Eigennamen (wie sie etwa von Frege vertreten wurde) verbindet ein Sprecher mit einem Eigennamen eine Beschreibung. Das Individuum, das der Sprecher bei der Verwendung eines Eigennamens bezeichnet, ist das Individuum, das die Beschreibung am Besten erfüllt. Wenn kein Individuum die Beschreibung erfüllt, die der Sprecher mit dem fraglichen Eigennamen verbindet, bezeichnet seine Verwendungsweise des Eigennamens überhaupt niemanden.

Nach der mit dieser Auffassung konkurrierenden kausalen Referenztheorie ist die Referenz eines Eigennamens nicht durch die Beschreibung festgelegt, die ein Sprecher mit einem Ausdruck verbindet, sondern *extern* durch eine kausale Kette bestimmt, in der der Sprecher zur ursprünglichen Einführung des fraglichen Eigennamens steht. Der Eigenname bezeichnet dann dasjenige Individuum, das am Anfang der Kausalkette mit diesem Namen „getauft“ wurde, sofern es ein solches Individuum gibt. Dies ist dann völlig unabhängig davon, welche Beschreibung der Sprecher mit dem Namen verbindet und unabhängig davon, was der Sprecher über das Bestehen oder Nichtbestehen einer solchen Kausalkette glaubt.

Bei diesem Experiment wurden 31 von Europäern abstammende amerikanische Studierende (undergraduates) in Rutgers befragt („Wessis“) und 40 chinesische Studierende (undergraduates) an der englischsprachigen University of Hong Kong („Asiaten“). Es wurden 4 Gedankenexperimente vorgegeben, zu denen jeweils eine Frage im Multiple Choice-Verfahren zu beantworten war. Eine der vorgegebenen Antworten entsprach dabei der Interpretation einer kausalen Referenztheorie, die andere Antwort der Interpretation durch eine deskriptivistische Theorie. Von den 4 getesteten Gedankenexperimenten schildern zwei ein Gedankenexperiment, das Kripkes Gödel-Gedankenexperiment nachempfunden ist.<sup>312</sup> In diesem Gedankenexperiment verwendet ein Sprecher einen Eigennamen in der Absicht, sich auf eine Person zu beziehen, von der er glaubt, dass sie eine bestimmte historische Leistung vollbracht hat. Tatsächlich hat die historische Person dieses Namens diese Leistung aber gar nicht vollbracht, sondern eine andere Person. Bezieht sich der Eigenname trotzdem auf die historische Person dieses Namens? Die deskriptivistische Theorie müsste dies zurückweisen.

---

<sup>312</sup> „Nehmen Sie an, daß Gödel gar nicht wirklich der Urheber dieses Theorems [der Unvollständigkeit der Arithmetik] war. In Wirklichkeit hat die betreffende Arbeit ein Mann namens „Schmidt“ getan, dessen Leiche vor vielen Jahren unter mysteriösen Umständen in Wien gefunden wurde. Sein Freund Gödel kam irgendwie in den Besitz des Manuskripts, und seither wurde das Manuskript Gödel zugeschrieben. Nach der in Frage stehenden Auffassung [d.h. der deskriptivistischen Theorie] will also unser gewöhnlicher Mensch, wenn er den Namen „Gödel“ verwendet, in Wirklichkeit auf Schmidt referieren, weil Schmidt die Person ist, die als einzige die Beschreibung „der Mann, der die Unvollständigkeit der Arithmetik entdeckte“ erfüllt. [...] Da also so der Mann, der die Unvollständigkeit der Arithmetik entdeckte in Wirklichkeit Schmidt ist, referieren wir, wenn wir über „Gödel“ reden, in Wirklichkeit immer auf Schmidt. Es scheint mir jedoch, daß das nicht so ist. Es ist einfach nicht so.“ (Kripke [172], 99).

Die beiden anderen Gedankenexperimente sind Kripkes Jona-Gedankenexperiment nachempfunden.<sup>313</sup> Dabei geht es um die Frage, ob ein Eigenname auch dann eine Person bezeichnet, wenn alles, was man über diese Person zu wissen glaubt, niemals von irgendjemandem erfüllt wurde (weil es beispielsweise nur Legende ist). Nach der deskriptivistischen Theorie bezeichnet ein solcher Eigenname dann keine reale Person.

Wählte eine Versuchsperson eine Antwort, die mit der kausalen Referenztheorie übereinstimmt, wurde dies mit 1 bewertet, sonst mit 0. Die einzelnen Resultate wurden einfach aufsummiert, so daß im statistischen Mittel einer Testgruppe ein Wert zwischen 0 und 2 pro Gedankenexperimenttyp resultiert. Die Ergebnisse waren die Folgenden:

	Punktzahl
<b>Gödelfälle</b>	
Wessis	1,13
Asiaten	0,63
<b>Jonafälle</b>	
Wessis	1,23
Asiaten	1,32

Abbildung 7.3-12

Nach Standardsignifikanztest ( $t$ -Test) weichen Wessis von Asiaten bei den Gödelfällen signifikant ab ( $p < 0,05$ ).

Die Ergebnisse bezüglich semantischer Intuitionen deuten also zunächst in dieselbe Richtung, wie die Ergebnisse zu epistemischen Intuitionen, weshalb der Einwand, dass es sich um Eigentümlichkeiten des Wissensbegriffs handelt, die die kulturelle Diskrepanz erklären können, vorerst wenig überzeugen kann.

Diese empirischen Untersuchungen stecken noch weitgehend in den Kinderschuhen, und es macht an dieser Stelle vermutlich wenig Sinn, an dem genauen Experimentaufbau, der Signifikanz der Ergebnisse und der Formulierung der Gedankenexperimente herumzudeuten. Es ist absolut vorstellbar, dass mittelfristig empirische Belege vorliegen, die mit ähnlichen, überzeugenderen Resultaten auf-

<sup>313</sup> „Angenommen es sagt jemand, daß kein Prophet je von einem großen Fisch oder Wal verschlungen wurde. Folgt daraus, daß Jona nicht existiert hat? Es scheint immer noch die Frage zu sein, ob die biblische Darstellung einer Person ist, die nicht existiert hat, oder ob sie eine legendäre Darstellung ist, die sich um eine wirkliche Person rankt. Im letzteren Fall ist es durchaus naheliegend zu sagen, daß es Jona zwar gegeben hat, daß aber niemand die Dinge vollbracht hat, die gewöhnlich mit ihm in Zusammenhang gebracht werden.“ (Kripke [172], 80).

warten können. Also, was für Konsequenzen sollte man aus solchen Ergebnissen ziehen?

Weinberg et al., wie auch Machery et al. sind sich ziemlich sicher, dass ihr Ergebnis tieferschürfende Konsequenzen für die Begriffsanalyse haben sollte, es ist aber alles andere als klar, weshalb dem so sein sollte. Was sicherlich stimmt, falls sich die Ergebnisse von Weinberg et al. und Machery et al. bestätigen lassen, ist, dass Äußerungen wie etwa die von Jackson, nach der ein Philosoph in aller Regel davon ausgehen kann, dass seine intuitiven Reaktionen universell geteilt werden, furchtbar naiv sind. Ob es für Jacksons Methodologie allerdings irgendwelche Folgen hat, dass unsere Intuitionen (und damit nach Jackson auch unsere Begriffe) von Kultur zu Kultur und zwischen Reich und Arm variieren, vermag ich nicht zu erkennen. Wie oben bereits erwähnt, gibt es durchaus Varianten des modalen Rationalismus, bei denen davon ausgegangen wird, dass die primären Intensionen, die wir mit einem Ausdruck verbinden, von Sprecher zu Sprecher variieren, wenn nicht sogar von Äußerung zu Äußerung (vgl. Chalmers [59]). (Wie das im modalen Rationalismus gelingt, kann im Detail in den Arbeiten von David Chalmers nachvollzogen werden.<sup>314</sup>) Aber ohne so weit gehen zu wollen, warum sollte ein Philosoph – sofern er an dem Alltagsverständnis irgendeines Begriffes interessiert ist – nicht genau an demjenigen Verständnis interessiert sein, das ihn und seine engsten Kollegen dazu treiben, sich mit philosophischen Problemen auseinanderzusetzen? Es mag sein, dass eine Rekonstruktion der Alltagstheorie von ‚Bedeutung‘ (sofern das überhaupt jemand anstrebt) keinen universellen Anspruch haben kann, also nicht mit dem Anspruch auftreten kann, einen universell geteilten Bedeutungsbegriff zu explizieren, wenn es den nicht gibt. Darin besteht aber doch noch kein gravierendes Problem. Machery et al. schreiben zu einer solchen Verteidigung Folgendes:

We find it *wildly* implausible that the semantic intuitions of the narrow cross-section of humanity who are Western academic philosophers are a more reliable indicator of the correct theory of reference [...] than the differing semantic intuitions of other cultural or linguistic groups. Indeed, given the intense training and selection that undergraduate and graduate students in philosophy have to go through, there is good reason to suspect that the alleged *reflective* intuitions may be *reinforced* intuitions. In the absence of a principled argument about why philosophers' intuitions are superior, this project smacks of narcissism in the extreme. (Machery, Mallon, Nichols und Stich [196], 9)

Dieser Narzissmus-Vorwurf ist aber zurückzuweisen. Wie Nelson Goodman schon treffend bemerkt hat, ist das Wittgensteinsche Bild vom Philosophen, der der Fliege hilft, den Ausweg aus dem Glas zu finden, in verschiedener Hinsicht korrekturbedürftig.<sup>315</sup> Eine solche Korrektur besteht darin, festzustellen, dass philosophische Probleme in aller Regel auftreten, weil *der Philosoph* sie als problematisch empfindet.

---

<sup>314</sup> ... insbesondere in Chalmers [60].

<sup>315</sup> Cohnitz und Rossberg [75], Kapitel 3.

Die Theorien, die Philosophen typischerweise mit Gedankenexperimenten testen, sind keine Theorien, die aufgestellt worden sind, um Indern mit geringem sozioökonomischen Status einen Problemlösungsvorschlag zu unterbreiten für irgendeins der (mutmaßlich interessanten, drängenden und wichtigen) Probleme, die ein sozioökonomisch schlecht gestellter Nachfahre eines Einwohners des indischen Subkontinents haben mag. Es geht nicht darum, sich narzisstisch über den Wissensbegriff oder Bedeutungsbegriff dieser Leute hinwegzusetzen, sondern es geht darum, dass deren Wissensbegriff oder Bedeutungsbegriff, wenn er denn vom Wissens- und Bedeutungsbegriff der westlichen Philosophenwelt verschieden ist, eben gerade nicht den Anlass zu demjenigen philosophischen Problem gegeben hat, das mit der Hilfe der Begriffsanalyse partiell einer Lösung zugeführt werden soll.

Nach dem weiter oben skizzierten Modell des moderaten Deskriptivismus testen Gedankenexperimente Rekonstruktionen unserer Begriffe, weil wir feststellen wollen, was wir unter einem bestimmten Begriff verstehen, um beispielsweise beurteilen zu können, ob eine Theorie einer Einzelwissenschaft (oder vielleicht, wie bei der Explikation des Begriffs der logischen Folgerung, auch eine formale Theorie) ein philosophisches Problem, das mit diesem Begriff in engem Zusammenhang steht, lösen kann oder nicht. Die Frage ‚Was ist, wie funktioniert und wozu gibt es Bewusstsein?‘ ist eine Frage, die *wir* gestellt haben. Was eine *für uns* befriedigende Antwort auf diese Frage ist, hängt davon ab, was *wir* mit ‚Bewusstsein‘ meinen. Was andere mit Bewusstsein meinen, spielt dabei eine vergleichsweise geringe Rolle.

Wenn es so ist, dass die Institution „akademische westliche Philosophie“ durch subtile Mechanismen, die mutmaßlich auf unsichtbare Hand-Prozesse zurückgeführt werden können, dafür sorgt, dass unter denjenigen, die gemeinsam an der Lösung dieser Frage arbeiten, durch Verstärkung und einseitige Ernährung mit Beispielen ein gemeinsames Verständnis dafür entwickelt wird, was man mit ‚Bewusstsein‘ meint, so dass diese Gruppe von Sprechern Intuitionen hat, die auf lange Sicht konvergieren, dann scheint das von unserem Standpunkt aus *begrüßenswert*. Auf diese Weise würden diese subtilen Mechanismen sicherstellen, dass man (diachron und synchron) an demselben Problem arbeitet (vgl. die Ergebnisse in Nichols, Stich und Weinberg [230], bei diesen Untersuchungen wurde (angeblich) herausgefunden, dass Versuchspersonen, die schon mehr Philosophiekurse besucht hatten, auch in deutlicherer Anzahl die „philosophischen“ Intuitionen teilten<sup>316</sup>).

Dass Antworten und Erklärungen so gut sind, wie sie sich auf das Verständnis und die Interessen beziehen, die hinter der ursprünglichen Frage gestanden haben, mindert auch nicht den Objektivitätsanspruch dieser Erklärungen.<sup>317</sup> Solange eine Begriffsanalyse von sich nicht behauptet, einen universal geteilten Begriff zu explizieren, sollte sie mit diesen empirischen Befunden keine Probleme haben.

---

<sup>316</sup> Diese Ergebnisse werden hier nicht diskutiert, weil die genauen Zahlen unbekannt sind.

<sup>317</sup> Vgl. Cohnitz [68].

Das gilt selbst für normative Befunde. So schreiben Nichols et al. in Bezug auf skeptische Argumente Folgendes:

And, of course, those skeptical arguments give us no reason at all to think that what High SES white Western philosophers call 'knowledge' is any better, or more important, or more desirable, or more useful than what these other folks call 'knowledge' [...]. Without some reason to think that what white, Western, High SES philosophers call 'knowledge' is any more valuable, desirable, or useful than any of the commodities that other groups call 'knowledge', it is hard to see why we should care if we can't have it. (Nichols, Stich und Weinberg [230])

Dagegen ist zu erwidern, dass gerade *weil* der Wissensbegriff anderer Gruppen von dem unserer Gruppe abweicht, die Frage völlig irrelevant ist, ob dieses ‚Wissen‘ in irgendeiner Hinsicht brauchbar oder wertvoll ist. Das hieße Äpfel mit Birnen zu vergleichen.

Etwas anders liegt die Sache bei ethischen Fragestellungen und moralischen Intuitionen. Wenn moralische Intuitionen abgefragt werden, um zu Aussagen über die *allgemeine Akzeptabilität* einer Sozialnorm (o.Ä.) Aussagen zu treffen, ist Vorsicht geboten. Der Schluss ‚X ist für mich intuitiv akzeptabel.‘ ‚X ist für alle intuitiv akzeptabel.‘ wird durch die empirischen Befunde von Haidt et al. stark in Zweifel gezogen. Was einem sozioökonomisch besser Gestellten als moralisch untadelig erscheint, mag für einen sozioökonomisch schlechter Gestellten völlig inakzeptabel sein.

Haidt hatte Versuchspersonen aus den Vereinigten Staaten und aus Brasilien kurze Geschichten wie die folgende vorgelegt:

A man goes to the supermarket once a week and buys a dead chicken. But before cooking the chicken, he has sexual intercourse with it. Then he cooks it and eats it. (Haidt, Koller und Dias [138])

Das Ergebnis war, dass Versuchspersonen mit niedrigem sozioökonomischem Status der Auffassung waren, dass die Person in dem Beispiel etwas moralisch Verwerfliches tut, während Versuchspersonen mit hohem sozioökonomischem Status diese Auffassung nicht teilten.

Wie wir in 4.2 schon erläutert haben, sind Prognosen darüber, wie eine bestimmte Sozialnorm oder Regelung von anderen empfunden wird, nur bedingt auf der Grundlage der eigenen Intuitionen zuverlässig. Geht es also um die Abschätzung des Grades der umfassenden Interessenberücksichtigung einer normativen (moralischen) Regelung, sind die empirischen Befunde zu berücksichtigen, und – wo immer möglich – sollten empirische Untersuchungen an die Stelle empirischer Spekulationen auf der Grundlage eigener Intuitionen treten.

In 4.3 haben wir aber auch schon erläutert, dass „Gedankenexperimente“ zur Einschätzung des Grades, zu dem eine Norm in umfassender Weise die Interessen der Beteiligten berücksichtigt, eine Sonderform des Gedankenexperiments in der Ethik darstellen. Die Kernrolle von Gedankenexperimenten in der Ethik bestand darin zu testen inwiefern sich unsere moralischen Intuitionen mit Regelungsvor-

schlagen (bzw. moralischen Prinzipien) in einem reflektierten Gleichgewicht befinden. Warum sollten dazu die Intuitionen weißer Reicher wichtiger sein? Mutmaßlich, weil diese Intuitionen im relevanten Sinne *reflektiert* sind. Im engeren Sinne geht es ja nicht um weiße Reiche, sondern um Philosophen, die an einer bestimmten ethischen Debatte teilnehmen und möglicherweise moralische Intuitionen haben, die andere Mitglieder der Gesellschaft nicht haben. Das ist aber *an sich* völlig unproblematisch. Es geht darum, moralische Prinzipien zu finden, die eben *normativ* sind. Wenn jeder aus seinem Bauchgefühl heraus schon wüsste, was in einer gegebenen Situation das moralisch Richtige ist, wären Bioethik-Kommissionen sinnlose Veranstaltungen. Wer die Ethik ernst genug nimmt, um sich um ihre methodologische Fundierung zu sorgen, sollte sie auch als wichtig genug ansehen, um ihr eine besondere Kompetenz zuzugestehen.<sup>318</sup>

Wie gesagt, die Dinge liegen anders, wenn man Intuitionen zu anderen Zwecken heranzieht. Dass ein moralisches Prinzip unsere reflektierten Gleichgewichtstests besteht, bedeutet noch nicht, dass dieses Prinzip auf allgemeine Akzeptanz treffen wird. Aus diesem Grund nimmt die Ethik Akzeptabilität als zusätzliches Kriterium zur Bewertung von moralischen Prinzipien, wenn es um konkrete Sozialnormen geht.

Zusammenfassend ist zu sagen, dass der Konzeption der *Bedeutungsanalyse* keine methodologisch bedeutsamen Schwierigkeiten aus der Tatsache erwachsen, dass Intuitionen zwischen Kulturkreisen und sozioökonomischen Gruppen variieren. Wie wir im nächsten Kapitel noch erläutern werden, sollten die empirischen Ergebnisse dazu anregen, Intuitionsbehauptungen genau zu prüfen. Es scheint aber nicht allgemein sinnvoll zu sein, Gedankenexperimente durch groß angelegte empirische Studien zu ersetzen. Begriffsanalysen – wie wir sie verstehen – zielen nicht primär darauf ab, einen universal geteilten Begriff herauszuschälen.

Aber selbst wenn man sich auf diese Position zurückzieht, ist die Begriffsanalyse noch nicht im sicheren Hafen angelangt. Wie – beispielsweise – Jackson sich an manchen Stellen ausdrückt, sollen Begriffsanalysen unsere Alltagstheorie ans Tageslicht befördern. Es geht uns um die rationale Rekonstruktion von Begriffen in so etwas wie hinreichende und notwendige Bedingungen. Warum sollte man annehmen, dass es überhaupt eine Alltagstheorie gibt, oder dass sie sich mit einer Menge notwendiger und hinreichender Bedingungen rekonstruieren lässt? Gibt es hier nicht noch weitere Gründe, die gegen die prinzipielle Möglichkeit bzw. Durchführbarkeit von Bedeutungsanalysen sprechen?

---

<sup>318</sup> Das schließt nicht aus, dass es Positionen in der Ethik gibt, die sich zur Aufgabe machen, moralische Alltagsintuitionen in einer ethischen Theorie aufzufangen. Diese Positionen sollten dabei den eigenen Intuitionen dementsprechend weniger Vertrauen entgegenbringen, als sie dies mutmaßlich bisher getan haben.

## 7.3.2 DIE REPRÄSENTATIONSSTRUKTUR UNSERER BEGRIFFE

Spätestens seit Ludwig Wittgensteins zweitem Hauptwerk, *Philosophische Untersuchungen* (Wittgenstein [352] [PU], insbesondere §65-77) spekulieren Philosophen darüber, dass einige, wenn nicht alle Begriffe, die wir verwenden, auf der Grundlage von Ähnlichkeitsurteilen verwendet werden, die sich nicht durch eine Angabe notwendiger und hinreichender Bedingungen rekonstruieren lassen. Wenn diese Spekulation zutrifft, scheint die Konzeption der Bedeutungsanalyse in Schwierigkeiten. Die Bedeutungsanalyse versucht ja gerade hinreichende und notwendige Bedingungen als Bedeutung anzugeben.

Nach der herkömmlichen Vorstellung besitzen wir einen Begriff (wobei offen bleibt, ob wir den Begriff erworben haben oder einige Begriffe angeboren sind) und können deswegen ein Wort, das diesen Begriff ausdrückt, auf bestimmte Dinge applizieren. Auf diese Weise klassifizieren wir Gegenstände oder Ereignisse mit Wörtern. In der Regel – selbst wenn wir den Begriff erworben haben – bedeutet das nicht, dass wir ohne zu zögern sofort sagen könnten, nach welchen Kriterien wir im Allgemeinen ein bestimmtes Wort applizieren. Begriffsanalyse scheint die Aufgabe zu haben, die Kriterien, nach denen wir bestimmte Wörter applizieren, offen zu legen und zu systematisieren. Manchmal verwenden wir zur Applikation eines Wortes Kriterien, von denen wir glauben, dass sie mit den eigentlichen Anwendungsbedingungen eines Begriffs korreliert sind. Deswegen denken sich Philosophen imaginäre Fälle aus, bei denen diese Korrelationen nicht auftreten (Fodor hatte das erläutert), um herauszufinden, welche Anwendungsbedingungen tatsächlich für ein bestimmtes Wort gelten. Der Hintergrund dieses Verfahrens ist die Annahme, dass unsere Klassifikationspraktiken nicht zufällig und veränderlich sind, sondern einem Muster folgen. Bestimmte Dinge klassifizieren wir als X, andere nicht – die Frage ist, was die Dinge, die wir als X klassifizieren, gemeinsam haben und von den Dingen unterscheidet, die wir nicht so klassifizieren.

Was, wenn es gar nichts gibt, das allen Dingen, die wir als X klassifizieren gemeinsam ist (also, abgesehen davon, dass wir sie als X klassifizieren)? Wittgensteins alternative Vorstellung war in etwa die, dass wir manche Wörter nur aufgrund bloßer Ähnlichkeiten zwischen Dingen verwenden, wobei diese Ähnlichkeitsurteile in unterschiedlichsten Hinsichten getroffen werden. Wittgensteins prominentes Beispiel ist die Verwendung des Ausdrucks ‚Spiel‘:

66. Betrachte z.B. einmal die Vorgänge, die wir „Spiele“ nennen. Ich meine Brettspiele, Kartenspiele, Ballspiele, Kampfspiele, usw. Was ist all diesen gemeinsam? [W]enn du sie anschaust wirst du zwar nicht etwas sehen, was allen gemeinsam wäre, aber du wirst Ähnlichkeiten, Verwandtschaften, sehen, und zwar eine ganze Reihe. [...] Schau z.B. die Brettspiele an, mit ihren mannigfachen Verwandtschaften. Nun geh zu den Kartenspielen über: hier findest du viele Entsprechungen mit jener ersten Klasse, aber viele gemeinsame Züge verschwinden, andere treten auf.

Wenn wir nun zu den Ballspielen übergehen, so bleibt manches Gemeinsame erhalten, aber vieles geht verloren. (Wittgenstein [352] PU, §66)

Manche Dinge bezeichnen wir also als ‚Spiel‘, weil sie sich in ihrem Unterhaltungsaspekt sehr ähneln, manche, weil es in ihnen in ähnlicher Weise um Konkurrenz geht, etc.

Dabei muss es keine Menge notwendiger Bedingungen geben, die all diese Dinge gemeinsam haben. Man könnte sich vorstellen, dass wir mal diese mal jene Hinsicht zum Anlass genommen haben, das Wort ‚Spiel‘ auf diese Aktivitäten auszuweiten, ohne dass alle zwei Aktivitäten, die wir als ‚Spiel‘ bezeichnen, in derselben Hinsicht (oder in überhaupt einer Hinsicht) gleich sein müssen. Diese Auffassung hat in verschiedenen systematischen Teilbereichen der Philosophie unterschiedlich viel Anklang gefunden, wobei bisweilen der Schluss gezogen worden ist, dass eine systematische Theorie des fraglichen Gegenstandsbereichs wegen der Familienähnlichkeitsstruktur seiner Begriffe nicht möglich sei.<sup>319</sup>

Auch von Stich und Weinberg wird dieser Einwand vorgebracht. Sie beziehen sich dabei insbesondere auf Frank Jacksons Position, dass die Begriffsanalyse dazu diene, so etwas wie eine Alltagstheorie ans Licht zu befördern. Sie versuchen nachzuweisen, dass die Annahme, es gäbe so etwas wie eine Alltagstheorie, die in unserem Bedeutungswissen abgelegt ist, mit modernen Theorien darüber wie unser Bedeutungswissen abgelegt ist, inkompatibel ist:

For, while some researchers who study commonsense concepts and the ways in which ordinary folk classify things into categories would agree that commonsense (or “folk”) theories guide our classificatory intuitions involving *some* terms or concepts, many researchers hold that our classificatory intuitions about *many* concepts are guided by cognitive structures that are very different from folk theories. (Stich und Weinberg [320], 2)

Stich und Weinberg führen als Beispiel für die Auffassung der „many researchers“ die *Exemplar*-Theorie an, derzufolge wir über eine Menge detaillierter Beschreibungen von paradigmatischen Exemplaren aus der Extension eines bestimmten Begriffs verfügen, die jedes Mal durchgegangen wird, wenn es darum geht, festzustellen, ob ein bestimmter Gegenstand unter einen Begriff fällt. Weist der fragliche Gegenstand ein hinreichendes Maß an Ähnlichkeit zu einer Beschreibung in der Menge der Exemplare auf, wird er entsprechend der Menge klassifiziert.

Etwas elaboriertere Exemplar-Theorien fordern dabei nicht, dass tatsächlich *alle* Exemplar-Beschreibungen durchgegangen werden, sondern nur eine jeweils aktuelle Teilmenge. Was dabei in der aktuellen Teilmenge ist, wie auch der Grad der Ähnlichkeit, der notwendig und hinreichend ist, um zu einer Klassifikation zu führen, wird durch die jüngste Klassifikationsgeschichte des klassifizierenden Subjekts bestimmt (Stich und Weinberg [320], 3).

---

<sup>319</sup> Solche negativen Konsequenzen zieht beispielsweise Morris Weitz in der Kunstphilosophie Weitz [339].

Stich und Weinberg weisen nun darauf hin, dass Exemplare untereinander keineswegs in inferentiellen oder explanatorischen Beziehungen stehen müssen, keine deduktiven oder induktiven Konsequenzen haben und als Exemplare auch keinen Wahrheitswert besitzen, weshalb Exemplare keine Theorien sind.<sup>320</sup>

Nun hängt die Frage, ob hierin ein Problem für die Begriffsanalyse und die Methode des Gedankenexperiments besteht, zum Teil davon ab, ob es sich bei den Gegenständen der Philosophie typischerweise um Begriffe handelt, die ausschließlich als Mengen von Exemplaren repräsentiert sind.

Im hier relevanten Sinne sind Begriffe Gegenstände der Psychologie und es handelt sich um diejenigen Datenstrukturen, die im Langzeitgedächtnis gespeichert sind und standardmäßig in höheren kognitiven Prozessen von Menschen gebraucht werden. Zu diesen „höheren kognitiven Prozessen“ gehört induktives und deduktives Rasonieren, wie auch das Klassifizieren von Gegenständen, etc.<sup>321</sup>

In welcher Beziehung diese mentalen Repräsentationen zu syntaktischen Einheiten der natürlichen Sprache stehen, ob Begriffe also durch Wörter ausgedrückt werden, ist damit noch nicht einmal angesprochen. Gehen wir aber einmal davon aus, dass Philosophen und Psychologen dasselbe meinen, wenn sie von „Begriffen“ reden.<sup>322</sup> In diesem Fall sollte man sich der Frage zuwenden, ob Begriffe eine natürliche Art bilden, und falls sie dies tun, ob sie Eigenschaften haben, die von denjenigen stark abweichen, die Philosophen von ihnen vermuten.

Ob Begriffe eine natürliche Art in diesem Sinne sind, ist eine empirische Frage. Gegenwärtig sprechen die empirischen Befunde dafür, dass es sich bei Begriffen um keine natürliche Art handelt, dass die Datenstrukturen, für die sich Psychologen interessieren, also nicht auf einheitliche Weise mental repräsentiert sind, sondern dass „Begriffe“ (also das, was wir vortheoretisch für einen solchen halten) unterschiedliche mentale Repräsentationen haben, die je nach kognitiver Aufgabenstellung Verwendung finden. Diese Repräsentationen entsprechen (mindestens) Exemplaren (wie oben geschildert), Prototypen und Theorien.

---

<sup>320</sup> Zu ähnlich gelagerten Einwänden vgl. Fodor, Garrett, Walker und Parkes [107], Ramsey [273], Pitt [258].

<sup>321</sup> Vgl. Machery [195].

<sup>322</sup> Es geht hier wohl eigentlich um zwei verschiedene Dinge. In dem einen Fall (Psychologie) geht es darum welche Entitäten postuliert werden müssen, damit in einem empirisch plausiblen Modell des menschlichen Geistes bestimmte funktionale Rollen gespielt werden können. In dem anderen Fall (Philosophie) geht es darum, bestimmte beobachtbare Regelmäßigkeiten in der Verwendung von Ausdrücken in einem idealisierten Sinne zu rekonstruieren. Wenn Philosophen solches „implizites Bedeutungswissen“ durch eine Angabe notwendiger und hinreichender Bedingungen „explizit“ machen, dann gehen sie nicht (notwendigerweise) davon aus, dass es irgendeine Ebene gibt (im menschlichen Unterbewusstsein oder so) auf der diese hinreichenden und notwendigen Bedingungen ebenfalls in propositionaler Form repräsentiert sind. Das Argument von Stich und Weinberg, dass Theorien semantische Eigenschaften haben, die Exemplare nicht besitzen, ist also ein reines Ablenkungsmanöver. Die interessante Frage ist, ob die Tatsache, dass wir manche Begriffe (manchmal) auf Grund von (unsystematischen) Ähnlichkeiten zwischen Dingen verwenden, für die Praxis der Bedeutungsanalyse von Relevanz ist.

Prototypen repräsentieren – in der Prototypentheorie von Begriffen – einen Begriff durch eine gewichtete Eigenschaftsmatrix. Gegenstände werden dann durch einen Prototypen klassifiziert, wenn sie eine gewisse kritische Menge von diesen Eigenschaften besitzen.

Theorien repräsentieren Begriffe – nach der Theorie-Theorie von Begriffen – durch ihre Einbettung in Verallgemeinerungen über das modale, kausale und funktionale Verhalten der fraglichen Gegenstände, die unter einen Begriff fallen. Die Theorie-Theorie geht nicht davon aus, dass Gegenstände aufgrund von äußerlichen Ähnlichkeiten unter einen Begriff subsumiert werden (wie dies die Prototypen- und Exemplartheorien tun), sondern vermutet auf der Basis empirischer Untersuchungen über das Klassifikationsverhalten von Menschen, dass die Klassifikation von Gegenständen unter einen Begriff von denjenigen Theorien geleitet ist, die das klassifizierende Individuum als Hintergrundwissen über den fraglichen Gegenstandsbereich besitzt. Laut der Theorie-Theorie werden Gegenstände dann durch einen Schluss auf die beste Erklärung klassifiziert. D.h. es wird die Menge von Verallgemeinerungen zur Klassifikation eines Gegenstandes herangezogen, die das bekannte Verhalten des Gegenstandes am Besten erklärt (vgl. Prinz [261], Murphy und Medin [215]). Begriffe stehen dabei – ähnlich wie in der umstrittenen Auffassung, dass Begriffe Definitionen sind<sup>323</sup> – untereinander in inferentiellen und explanativen Beziehungen, und können – je nachdem wie kohärent das Hintergrundwissen eines Individuums ist – prinzipiell auch durch die Angabe hinreichender und notwendiger Bedingungen analysiert werden.

Eine Metaanalyse derjenigen empirische Untersuchungen, die jeweils die Frage untersuchten, welche dieser Repräsentationsformen für einen gegebenen Begriff empirisch adäquat ist, kommt zu dem Ergebnis, dass Begriffe vermutlich gar keine natürliche Art darstellen und keines dieser Modelle für einen gegebenen (vortheoretischen) Begriff als alleinige Repräsentation fungiert (vgl. Machery [195]).

Falls man Jacksons Anliegen so versteht, dass es der Begriffsanalyse um die Aufdeckung der impliziten Theorie geht, die unser Klassifikationsverhalten steuert, ist diese Position durch empirische Befunde also *nicht* angegriffen. Solange man davon ausgehen kann, dass die relevanten kognitiven Funktionen aktiviert werden, die mit der Theorien-Repräsentation eines Begriffs in Verbindung stehen, kann man auch davon ausgehen, dass dieses Verfahren zu einer Rekonstruktion der impliziten Theorie führt (zumindest in dem schwachen Sinne, dass man davon ausgehen darf, dass es eine solche gibt, und nach allem, was wir gesagt haben, nichts sonst gegen ihre prinzipielle Rekonstruierbarkeit spricht). Wenn Begriffe *auch* als Theorien repräsentiert sind, sollte diese Repräsentationsweise rekonstruierbar sein. Dass die parallelen Repräsentationen als Exemplare oder Prototypen diesem Verfahren nicht in die Quere kommen, wird man vermutlich dadurch ausgeschlossen haben, dass man den Versuchspersonen (den Adressaten eines Gedankenexperiments) mitteilt, dass es um theoretische Zusammenhänge geht, in denen modale, kausale und funktionale Eigenschaften der fraglichen Dinge zur Debatte stehen. Welche Versuchsumstände am Besten dazu geeignet sind, die relevanten kogniti-

---

<sup>323</sup> Vgl.: Fodor, Garrett, Walker und Parkes [107], Pitt [258].

ven Funktionen zu aktivieren, die eine Rekonstruktion der Alltagstheorie erlauben, ist ansonsten eine Frage der empirischen Psychologie. Macherys Metaanalyse relativiert in jedem Fall zunächst die Bedenken Stichs und Weinbergs, es gäbe gar keine mentale Repräsentation von Begriffen in Gestalt einer Theorie.

Selbst wenn die Ergebnisse dieser Metaanalyse durch neue Befunde in Zweifel gezogen *würden*, und man annehmen dürfte, dass manche Begriffe tatsächlich nur Exemplare sind, ist immer noch nicht ausgemacht, dass auch die philosophischen Begriffe Exemplare sind. Sofern Exemplar-Theorien plausibel sind, beziehen sie sich auf *Beobachtungsbegriffe* bzw. Begriffe, die prinzipiell *ostensiv* erlernbar sind. ‚Wissen‘, ‚Wahrheit‘, ‚freier Wille‘, ‚Bewusstsein‘, ‚logische Folgerung‘ etc. erfüllen diese Bedingung zu einem sehr geringen Grad (wenn überhaupt).

Lassen wir diese Zweifel für das Argument aber ebenfalls vorerst beiseite. Welche Gründe kann man bei einem *bestimmten* philosophischen Begriff für die Annahme haben, dass er durch Exemplare repräsentiert ist?

In der Tat ist für den Personenbegriff wie für den Kunstbegriff schon argumentiert worden<sup>324</sup>, dass er nicht als Theorie, sondern als Exemplar bzw. als Prototyp repräsentiert ist. Genauer gesagt, es wurde argumentiert, dass unsere Repräsentation des Personenbegriffs bzw. des Kunstbegriffs nicht als eine Menge hinreichender und notwendiger Bedingungen rekonstruiert werden kann. Wenn wir etwas als ‚Person‘ oder als ‚Kunstwerk‘ klassifizieren, dann auf der Grundlage von Ähnlichkeitsurteilen, wobei die Basis dieser Ähnlichkeiten eine Liste von Exemplaren, oder eine gewichtete Eigenschaftsmatrix sein mag.

Innerhalb der Kunstphilosophie hat man für diese Auffassung etwa folgendermaßen argumentiert: Die Zukunft der Kunst ist offen, d.h., was in Zukunft als Kunstwerk geschaffen wird, verdankt sich einem essentiell kreativen Prozess, bei dem typischerweise nach Originalität gestrebt wird. Originalität bedeutet, dass mit kunsthistorischen Traditionen gebrochen wird und neue Kunstwerke sich in beliebiger Hinsicht von ihren Vorgängern in der Kunstgeschichte unterscheiden. Wenn dem aber so ist, kann man keine hinreichende und notwendige Menge von Eigenschaften finden, die allen Kunstwerken bis jetzt und in Zukunft gemeinsam wäre (vgl. etwa Weitz [339]).

Dieses Argument ist aus offensichtlichen Gründen schlecht. Zunächst bietet der Verweis auf die essentielle Offenheit der Kunst eine Angriffsfläche. Kunst ist also schon mal *essentiell* offen, damit gibt es zumindest *eine* notwendige Eigenschaft und es müsste mit dem Teufel zugehen, wenn man in dieser Richtung aus Weitz' Überlegungen nicht eine eindeutige Charakterisierung von ‚Kunst‘ gewinnen könnte. Das Argument übersieht nämlich, dass die Offenheit der Kunst sich nicht auf alle *relationalen* Eigenschaften von Kunstwerken ausdehnen muss. Es mag beispielsweise sein, dass Kunstwerke in eine kunsthistorische Tradition gehö-

---

<sup>324</sup> Ich will nicht behaupten, dass in anderen Bereichen so eine Argumentation bisher noch nicht vorgekommen ist. Die Kunstphilosophie und die Problematik diachroner Personenidentität dienen hier als idealtypische Beispiele für zwei verschiedene Argumentationsstrategien.

ren müssen, und dass Künstler schon aus logischen Gründen gegen diese notwendige Bedingung nicht bewusst verstoßen können, etc.<sup>325</sup>

Die Argumentation in Bezug auf den Personenbegriff ist da schon um Einiges plausibler. In diesem Fall hat man tatsächlich Evidenz herangetragen, dass der Personenbegriff aufgrund von Ähnlichkeitsurteilen verwendet wird.

Zunächst können wir mit Gendler (Gendler [119], Gendler [121]) zwei Strategien unterscheiden, wie man mit „Ausnahmefällen“ umgehen kann. ‚Ausnahmefall‘ bezeichnet hierbei einen imaginären Fall eines X (oder nicht-X), der unserem Urteil nach in Widerspruch zu der bisher für wahr gehaltenen Theorie über X steht. In anderen Worten: es geht um Gedankenexperimente und die Frage, wie man als Adressat eines solchen reagieren sollte. Nach der *exception-as-scalpel-strategy* („Ausnahme als Skalpell“, wir reden im Folgenden von ‚EASS‘) besitzt man eine tentative Begriffsanalyse T eines Begriffs C, die einige der Eigenschaften, die die korrekten Anwendungen von C typischerweise begleiten, als begrifflich notwendige und hinreichende Eigenschaften des C-seins auszeichnet. Sei T von der folgenden Form:

$$(T) \quad \forall x (Cx \leftrightarrow (P_1x \wedge P_2x \wedge P_3x \wedge P_4x))$$

T behauptet dann, dass alle und nur die Entitäten, die unter C fallen, die Eigenschaften  $P_1$ - $P_4$  aufweisen. Angenommen wir können uns einen Fall vorstellen, in dem eine Entität  $e$  zwar die Eigenschaften  $P_1$ - $P_3$  besitzt, der  $P_4$  aber fehlt und die dennoch – gemäß unserem Urteil – unter C fällt. In einem solchen Fall sind wir berechtigt, T zu revidieren, da  $P_4$  offenbar nicht länger als eine notwendige Bedingung für C-sein erachtet werden kann:

The exception-as-scalpel strategy uses exceptional cases as a way of progressively narrowing the range of characteristics required for the application of a concept by allowing us to isolate the essential features for concept-application from those which are merely ordinarily correlative. (Gendler [119], 607)

Dabei handelt es sich – laut Gendler – aber nur um *eine* mögliche Weise mit dem neuen Befund umzugehen. Wenn wir die Annahme fallen lassen, dass C mit einer Angabe notwendiger und hinreichender Bedingungen rekonstruiert werden kann, gibt es noch eine andere:

[T]he exception-as-cantilever strategy views the category membership of exceptional cases as essentially reliant on the ordinary instances against which they can be seen as exceptions. (Gendler [119], 607)

Wenn wir nun diese Strategie („Ausnahme als Ausleger“, wir reden im Folgenden von ‚EACS‘) auf den neuen Fall anwenden, dann betrachten wir  $e$  als C, weil es in relevanter Hinsicht den typischeren C-Exemplaren ähnlich ist, die in wahrer Wei-

---

<sup>325</sup> Das Argument übersieht also die Möglichkeit einer gehaltvollen Institutionentheorie der Kunst, oder einer gehaltvollen historischen Theorie (vgl. Carroll 1999).

se durch T beschrieben werden. T hätte demnach eigentlich etwa die folgende logische Form<sup>326</sup>:

$$(T^*) \quad \forall x (Cx \leftrightarrow \exists y ((P_1y \wedge P_2y \wedge P_3y \wedge P_4y) \wedge Sxy))$$

,S' sei hierbei eine zweistellige Ähnlichkeitsrelation, die genau dann besteht, wenn zwischen einem typischen Exemplar und dem neuen Fall ein bestimmter Schwellenwert an Ähnlichkeit erreicht ist.

Falls C so strukturiert ist und S hinreichend liberal, dann könnte jede einzelne der Eigenschaften  $P_1$ - $P_4$  in einem aktualen oder imaginären Fall abwesend sein (wenn vermutlich auch nicht alle gleichzeitig), ohne dass C deswegen auf den fraglichen Fall nicht mehr applizierbar wäre.

Falls wir also Gründe für die Annahme besitzen, dass ein Begriff auf der Grundlage von Ähnlichkeitsurteilen verwendet wird, sollte ein imaginärer Fall, in dem einige der (mutmaßlich) typischen Eigenschaften fehlen, bei dem wir den fraglichen Begriff aber dennoch verwenden würden, uns nicht dazu bewegen, die vermutete Menge der kriteriellen Eigenschaften zu ändern.  $P_1$ - $P_4$  können durchaus in diesem Sinne alle relevant zur Bestimmung der Frage sein, ob etwas unter C fällt, sie müssen aber nicht alle in jedem gegebenen Fall vorkommen. Ausnahmefälle, die zwar C sind aber nicht alle Eigenschaften  $P_1$ - $P_4$  aufweisen, sind eben nur *Ausnahmefälle*.

Warum aber sollte man der Meinung sein, dass der Personenbegriff aufgrund solcher Ähnlichkeitsurteile verwendet wird? Gendler argumentiert für diese These mit einem Gedankenexperiment, das ursprünglich aus Bernard Williams' *The Self and the Future* (Williams [346]) stammt.

Angenommen zwei Personen, Herr A und Herr B, die irgendwie in die Abhängigkeit von einem böswilligen Arzt gelangt sind, bekommen von diesem zum Zeitpunkt  $t_0$  Folgendes mitgeteilt: Beide werden sich zum Zeitpunkt  $t_1$  einer Operation unterziehen müssen, die in einem Wechsel psychischer Kontinuität bestehen wird. Nach der Operation, zum Zeitpunkt  $t_2$ , wird derjenige mit dem jetzigen A-Körper die Erinnerungen, Interessen und Charaktereigenschaften besitzen, die B bis zu  $t_1$  besaß, und umgekehrt. Den beiden wird außerdem mitgeteilt, dass einer von beiden nach  $t_2$  gefoltert wird, während der andere eine Belohnung erhält.

Zu den wenigen guten Nachrichten, die Herr A an diesem Tag erhält, gehört, dass er wählen darf, welcher Körper nach der Operation gefoltert werden soll, und welchem Körper die Belohnung überreicht wird. Würde man uns fragen, wofür sich A entscheiden sollte, würden sicherlich die meisten dazu raten, im wohlverstandenen Eigeninteresse die Folter dem A-Körper angedeihen zu lassen. D.h., dass wir uns intuitiv an einer mentalistischen Auffassung personaler Identität orientieren und der Kontinuität des Körpers in diesem Fall kein Gewicht beilegen.

Nun betrachte man aber den folgenden Fall: Diesmal befindet sich nur Herr A in der Gewalt des böswilligen Arztes, der ihm eröffnet, dass er zum Zeitpunkt  $t_1$

---

<sup>326</sup> Vgl. Cohnitz [70].

eine Operation an ihm durchführen wird, bei der A zunächst sein Bewusstsein verliert, nach der Operation, zu  $t_2$ , aber wieder aufwachen wird, allerdings mit völlig anderen Erinnerungen, Absichten und Charaktereigenschaften. Als sei dies noch nicht genug, würde A nach der Operation außerdem noch gefoltert. Wie sollte A auf diese Neuigkeit reagieren? Nun, intuitiv würden wir urteilen, dass A vermutlich recht schockiert reagieren würde. Nicht nur, dass er die totale Zersetzung seiner Psyche zu erwarten hat, er soll auch noch gefoltert werden.

Wie es scheint, haben wir relativ stabile Intuitionen darüber, wie sich A jeweils verhalten sollte. Es scheint ebenfalls, dass sich As Reaktionen in den beiden Fällen stark unterscheiden sollten. Im ersten Fall sollte A die zukünftige Folter des A-Körpers nicht sonderlich schockieren, während er im zweiten Fall zu Recht sehr besorgt sein sollte. Gendler und andere haben dieses Beispiel als Beleg dafür angesehen, dass ein und dasselbe Gedankenexperiment zu unterschiedlichen Reaktionen führen kann, wenn man es nur auf unterschiedliche Weise erzählt.<sup>327</sup> Was hat dies nun mit der Frage zu tun, wie der Personenbegriff strukturiert ist?

Angenommen wir würden in beiden Fällen die EASS anwenden. Das erste Gedankenexperiment sollte uns dazu bringen, körperliche oder physische Kontinuität nicht zu den notwendigen Bedingungen der diachronen Personenidentität zu zählen. Sofern unsere tentative Begriffsanalyse physische Kontinuität als notwendige Bedingung für diachrone Personenidentität nennt, ist diese Bedeutungsanalyse zu revidieren. (Zumindest auf der Grundlage der Vermutung, dass es uns in Bezug auf Überleben auf personale Identität ankommt. Diese methodologische Prämisse muss man nicht teilen, vgl. Parfit [250].)

Das zweite Gedankenexperiment sollte uns dazu bringen, auch psychischer Kontinuität den Laufpass zu geben. In diesem Fall scheint die Tatsache, dass psychische Kontinuität total verletzt ist, für As Sorge keine Rolle zu spielen, folglich ist psychische Kontinuität keine notwendige Bedingung für personale Identität.

Benutzt man also die EASS im Fall der Analyse des Begriffs diachroner Personenidentität, schneidet man mutmaßlich zu viele notwendige Bedingungen weg. Nach den beiden Gedankenexperimenten sollten wir weder psychische noch physische Kontinuität als notwendige Bedingung diachroner Personenidentität erachten.

Betrachten wir die beiden Ergebnisse aber im Licht der EACS, sollten wir uns eines Urteils bezüglich der notwendigen Bedingungen diachroner Personenidentität enthalten:

But rather than concluding something about the (lack of) necessary and sufficient conditions for the application of the concept 'person', the exception a cantilever strategy tells us to conclude this about our classification of these exceptional cases as cases where diachronic personal identity obtains: our decision about these cases are justified by the rational permissibility of assimilating them to ordinary cases. In the

---

<sup>327</sup> Man bedenke die Vorbehalte, die wir in Kapitel 3 gegenüber der Redeweise geäußert haben, dass es sich hierbei um „ein und dasselbe Gedankenexperiment“ handeln kann. Diese Stelle ist in loser Redeweise formuliert (offenbar trifft nicht mal  $T_3$  die hier intendierte Bedeutung).

[second] scenario, we focus on the similarity that concerns physical continuity; in the [first] scenario, we focus on psychology. But in both cases, we are cantilevering out from the set of generally-obtaining correlations which characterize ordinary cases. (Gendler [119], 608)

Wenn ich Gendlers Ausführungen richtig verstehe, dann fallen wir – nach ihrer Auffassung – bei der Anwendung des Personenbegriffs Ähnlichkeitsurteile, die in einem der beiden imaginären Fälle auf die psychische Kontinuität fokussiert sind und im anderen Fall auf die physische Kontinuität. Unsere Urteile bezüglich imaginärer Fälle hängen also davon ab, in welche Ähnlichkeitsklassen wir die Fälle einordnen, was wiederum davon abhängt, wie uns ein imaginärer Fall präsentiert wird.

Hieraus folgert Gendler, dass imaginäre Fälle nur unzuverlässig darüber Aufschluss geben, wie wir einen Begriff in tatsächlichen Situationen anwenden „würden oder sollten“. Obwohl Gendler an mehreren Stellen verspricht, dass dafür noch ein Argument vorgebracht werden soll und am Ende von Gendler [119] zufrieden feststellt, dass sie dies gezeigt habe, findet sich außer dem bereits Erwähnten nichts, was diese Konklusion rechtfertigen würde. Zunächst ist die Frage, welche Urteile wir in realen Situationen fällen *würden*, eine andere Frage, als die, welche Urteile wir fällen *sollten*. Rekonstruieren wir zunächst, was Gendler damit meinen könnte, dass der obige Fall zeigt, dass Gedankenexperimente uns keinen zuverlässigen Aufschluss darüber geben können, was wir in realen Situationen sagen *würden*.

Hier sollte man eigentlich erwarten, dass wir den Nachweis erhalten, dass die Urteile, die wir im Gedankenexperiment fällen, nicht den Urteilen entsprechen, die wir hinsichtlich realer Fälle tatsächlich treffen. Eine solche empirische Untersuchung wird hier aber offenbar nicht vorgelegt. Dass die Urteile zum imaginären Fall mal so und mal so getroffen werden (nehmen wir einmal an, die Einteilung in verschiedene Ähnlichkeitsklassen wäre unsystematisch), kann mit unserer tatsächlichen Verwendungsweise des Personenbegriffs doch absolut übereinstimmen. Das erste Gedankenexperiment zeigt ja (von Gendler unbestritten), dass psychische Kontinuität ein hinreichendes Kriterium ist und dass psychische Kontinuität kein notwendiges Kriterium ist. Es *ist* eben ein Fehlschluss, daraus zu folgern, dass psychische Kontinuität dann auch automatisch notwendig und psychische Kontinuität automatisch niemals hinreichend ist. Wo ist das Problem? Es kann sehr gut möglich sein, dass wir uns in unseren Urteilen über diachrone Personenidentität im Alltag genau wie im imaginären Fall davon leiten lassen, welche Informationen wir über einen Fall haben und wie er uns präsentiert wird. Dass wir unsere Begriffe so verwenden, mag für einen Philosophen, der in der Alltagstheorie tiefe Wahrheiten vermutet, frustrierend sein, hat aber zunächst keine unmittelbaren Folgen für die Frage nach der Zuverlässigkeit von Gedankenexperimenten.

Anders könnte es bei der Frage danach aussehen, wie wir einen Begriff verwenden *sollen*. Für diese Frage mag es wünschenswert sein, dass unser Urteilen nicht von kontingenten, kontextuellen Faktoren abhängt. Ob jemand dieselbe Person ist wie die, die ein bestimmtes Verbrechen begangen hat, ist die Grundlage für unsere

Zurechnung von Schuld und Verantwortung; diese Urteile sollten in irgendeiner Form dem angemessen sein. In Bezug auf den Personenbegriff ist es aber unklar, was der angemessenere Begriff wäre, wenn wir den Personenbegriff nun mal so kontextabhängig verwenden, wie wir es zu tun scheinen. Ein Beispiel: Aus der empirischen Psychologie sind in der Nachfolge von Kahneman und Tversky viele Untersuchungen bekannt, bei denen gezeigt wurde, dass die intuitiven Urteile von Testpersonen zu wahrscheinlichkeitstheoretischen und logischen Fragen „irrational“ sind. Urteilen Testpersonen beispielsweise über den Erwartungsnutzen eines bestimmten Ereignisses, lassen sie sich häufig von kontextuellen Faktoren beeinflussen, die mit der wahrscheinlichkeitstheoretischen Ermittlung des Erwartungsnutzens überhaupt nichts zu tun haben. (Testpersonen lassen sich beispielsweise davon beeinflussen, ob man die Wahrscheinlichkeit einer negativen Handlungskonsequenz betont, oder die Wahrscheinlichkeit der genau komplementären positiven Handlungskonsequenz).<sup>328</sup> In diesem Fall ist klar, dass wir nicht so urteilen sollen, wie diese Versuchspersonen es intuitiv tun, weil wir eine fertige, anerkannte Theorie haben (die Wahrscheinlichkeitstheorie), die uns sagt, welche Faktoren auf den Erwartungsnutzen einen Einfluss haben und welche nicht und weswegen es irrational ist, sich von anderen Faktoren in seinem Urteil beeinflussen zu lassen. Wir besitzen also eine normative Theorie und können auf deren Grundlage die psychologischen Daten normativ beurteilen: Diese Versuchspersonen verwenden ihren Begriff des Erwartungsnutzens eines Ereignisses nicht so, wie sie sollten.

Über den Personenbegriff haben wir (noch) keine entsprechende Theorie. Da Gendler uns in Bezug auf den Personenbegriff auch keine anbietet, ist es alles andere als klar, warum das „Ergebnis“ des Williamsschen Gedankenexperiments zeigt, dass Gedankenexperimente uns nur unzuverlässig darüber informieren, wie wir den Personenbegriff verwenden *sollen*.

Man könnte einwenden, dass Gendlers Argument hier schlechter wegkommt, als es tatsächlich ist. Schließlich haben wir doch gerade eben selbst zugegeben, dass der Personenbegriff eine wichtige Rolle spielt und es wünschenswert wäre, wenn er konsistenter verwendet würde, als wir es wohl tatsächlich tun. Kann man dann nicht immer noch sagen, dass Gedankenexperimente ein (zwar nicht unzuverlässiges aber doch) *inadequates* Mittel sind, um eine solche konsistente Verwendungsweise aus der tatsächlichen Verwendungsweise herauszuschälen, weil die kontextuellen Faktoren, die uns im Alltag zu unsystematischen Urteilen führen, uns auch im Gedankenexperiment beeinflussen?

Um es vorweg zu sagen: ja, dem ist vielleicht so. Es mag sein, dass unsere Begriffsverwendung in vielen Fällen zu unsystematisch ist, um einer Rekonstruktion zugänglich zu sein. Es ist dann zwar erklärungsbedürftig, wie wir es schaffen, uns mit diesen Begriffen zu verständigen, aber das kann auch Glück sein, oder nur eine Täuschung (vielleicht reden wir ja häufig bei diesen Dingen aneinander vorbei,

---

<sup>328</sup> Vgl. Tversky und Kahneman [331], Bishop und Trout [26], Cohnitz und Rossberg [75], Stich und Nisbett [319], Stich [318], Hoeschen [151]. Die Relevanz dieser Resultate für das Gedankenexperiment diskutiert Horowitz [152].

ohne es (immer) zu merken). Diesen skeptischen Einwand kann man nur schwerlich völlig aushebeln und ich sehe auch keinen Grund, das zu tun. Ich denke, es reicht für eine Rehabilitierung des Gedankenexperimentierens gegen Gendlers Kritik, wenn man zeigen kann, dass es auch in scheinbaren Patt-Situationen wie dem Williams-Gedankenexperiment die Möglichkeit gibt, die Theoriebildung (bzw. Begriffsanalyse) auf rationale Weise weiterzuführen. Wie man sich das vorstellen soll, wird nun kurz erläutert und in Teil 9 in einem größeren Zusammenhang entwickelt.

Das uneinheitliche Urteil in Williams' Gedankenexperiment stellt deswegen ein Problem dar, weil scheinbar unschuldige Varianten in der Erzählweise ein und desselben Falles zu unterschiedlichen Urteilen führen. Ist dem so? Handelt es sich um „unschuldige Varianten“? Das scheint mehr als fraglich.

Dass A im zweiten Fall (nach unserem intuitiven Urteil) schockiert reagieren sollte, hängt zunächst einmal sicherlich davon ab, dass der Fall so präsentiert wird, als ob es bereits ausgemacht wäre, dass A diese Dinge zustoßen werden. Während der erste Fall hinsichtlich der Frage neutral ist, wer nach der Operation in wessen Körper aufwachen wird, ist die Darstellung des zweiten Falles parteiisch. A bekommt mitgeteilt, dass A gefoltert wird und eine Gehirnwäsche bekommt, etc.

Außerdem ist As Reaktion auch unter der Annahme eines psychischen Kriteriums plausibel. Nimmt man eine mentalistische Theorie personaler Identität an, dann droht der Übel wollende Arzt A mit dem Tod. Falls A sich dennoch der Hoffnung hingeben sollte, dass vielleicht nicht seine ganze Psyche dem Verfahren zum Opfer fällt (und das suggeriert die Geschichte aufgrund der Wahl der Personalpronomen durch den Arzt), droht ihm außerdem Folter. Natürlich sollte A sich Sorgen machen.<sup>329</sup>

Williams' Beschreibung ist einfach eine *petitio principii*: es soll gezeigt werden, dass A auch dann dieselbe Person bleibt, wenn As Psyche zerstört wird; die Beschreibung des Gedankenexperiments setzt aber bereits voraus, dass A dieselbe Person bleibt.<sup>330</sup> Der intuitive Befund, der zunächst gegen eine mentalistische Theorie zu sprechen scheint, kann auf diese Weise *wegerklärt* werden, indem auf Fehler im Gedankenexperiment hingewiesen wird. Wie wir in Teil 9 sehen werden, gibt es noch weitere Möglichkeiten, mit solchen scheinbaren oder tatsächlichen Patt-Situationen umzugehen. Zunächst müssen wir uns aber einem fundamentalen Problem zuwenden.

---

<sup>329</sup> Beide Williams-Geschichten sind beispielsweise mit Nozicks „closest-continuer“-Theorie vereinbar, welche mentalistisch ist und sowohl hinreichende wie notwendige Bedingungen für diachrone Personenidentität angibt. Vgl. Nozick [243].

<sup>330</sup> Zu dieser Diagnose kommt auch Kannuck [161], Kannuck [162], wie auch Coleman [77].



## 8. Vorstellbarkeit und Möglichkeit

Wir haben bis jetzt Gedankenexperimente in der Philosophie nur unter unseren Kriterien aus Kapitel 3 betrachtet. Dabei handelte es sich um *Gedankenexperimente zur Überzeugungsänderung*, also Gedankenexperimente, die in ganz bestimmten Argumenten vorkommen und dazu dienen sollen, eine gerechtfertigte Überzeugungsänderung im Adressaten auszulösen. In den letzten Kapiteln haben wir argumentiert, dass die Zuverlässigkeit von Gedankenexperimenten zur Überzeugungsbildung bzw. Revision sehr davon abhängt, welche Art von Überzeugung zur Debatte steht. Man kann vielleicht a priori darüber urteilen, wie man einen Ausdruck (und den damit verbundenen Begriff) in kontrafaktischen Umständen verwenden würde, kann von da aus aber nicht ohne weiteres darauf schließen, dass jeder, der dieselben Ausdrücke benutzt, auch genauso urteilen würde. Man kann auch nicht ohne weiteres darauf schließen, dass der Tatsache, dass wir *Ausdrücke* (bzw. Begriffe) in modalen Kontexten auf bestimmte Weise verwenden, auch eine Tatsache auf der Ebene derjenigen *Dinge* entsprechen muss, die wir mit diesen Ausdrücken bezeichnen. Was uns begrifflich möglich erscheint muss nicht metaphysisch möglich sein (wenn es so etwas gibt), und ob es sich überhaupt systematisch angeben lässt, wann solche Schlüsse erlaubt sind, ist höchst umstritten. Bei Gedankenexperimenten gegen Realdefinitionen ging es aber in jedem Fall darum, eine Aussage darüber zu treffen, was (im jeweils relevanten Sinne) möglich ist.

Wir haben dabei in den letzten Kapiteln unser modales Urteilen als primitiv behandelt.<sup>331</sup> Sei es nun unser modales Urteil über metaphysische oder über begriffliche Möglichkeiten, wir haben bisher in relativ ungeklärter Weise so getan, als wäre die Beurteilung solcher Möglichkeiten eine Frage einer intuitiven Reaktion auf eine bestimmte Proposition. Es wurde dann argumentiert, dass man solchen intuitiven Urteilen mutmaßlich dann vertrauen kann, wenn man begründet annehmen kann, dass sie von dem relevanten Gegenstandsbereich informiert sind. Bei Intuitionen über empirische bzw. metaphysische Zusammenhänge war es eben unklar, in welchem Sinne diese Intuitionen aus diesem Bereich informiert sein sollten, bei Intuitionen in Bezug auf unser Bedeutungswissen schien es plausibel, anzunehmen, dass unsere Intuitionen in relevanter Weise informiert *sind*.

Betrachten wir nun unser modales Urteilen aber etwas genauer. Welche Prozesse gehen einem modalen Urteil voraus?

---

<sup>331</sup> Manche Philosophen sind der Auffassung, dass diese Sichtweise auch hinreichend ist. Vgl. etwa Bealer [16] oder Williamson [348].

In der Philosophie geht man davon aus, dass modale Urteile auf der Grundlage von Vorstellbarkeitstests gefällt werden. Dieses Verfahren ist aber alles andere als unstrittig. Dabei kann man zwei Stoßrichtungen der Kritik unterscheiden. Einerseits kann man an der Auffassung, dass wir aufgrund von Vorstellbarkeitstests über Modales urteilen, grundsätzlich kritisieren, dass wir ein psychisches Vermögen zur Beurteilung von etwas heranziehen, das in keinerlei geklärtem Verhältnis dazu steht. Warum sollte es gerade *dieses* psychische Vermögen sein, das uns über Modales aufklärt. Dieser Einwand ist bereits von Mill vorgetragen worden:

[O]ur capacity or incapacity of conceiving a thing has very little to do with the possibility of the thing in itself; but is in truth very much an affair of accident, and depends on the past history and habits of our minds. (Mill [209], Buch II, Kapitel V, Sektion 6)

Könnte es also nicht sein, dass unser Vermögen uns etwas vorzustellen von dem Bereich des Möglichen und Notwendigen einfach völlig unabhängig ist?<sup>332</sup>

Ein zweiter Einwand würde weniger auf die ungeklärte Verbindung zwischen dem Bereich des Modalen und unseren psychischen Vermögen verweisen, sondern argumentieren, dass wir uns offenbar alles vorstellen können, sei es möglich oder nicht, ohne dass wir dabei in irgendeiner Hinsicht einen Unterschied in der Art der Vorstellung feststellen könnten. Wenn wir uns aber Mögliches so gut wie Unmögliches vorstellen können, ist *diese* Methode, zu modalen Urteilen zu gelangen, völlig unzuverlässig, wie auch immer ihre Verbindung zum Bereich des Möglichen und Notwendigen sonst sein mag. Beiden Kritiken wird nun nachzugehen sein.

---

<sup>332</sup> Mills Einwand ist freilich auf metaphysisch Mögliches gemünzt, kann aber genausogut für begrifflich Mögliches vorgetragen werden. Auch hier kann man fragen, wieso es ausgerechnet *dieses* Vermögen der Vorstellbarkeit sein soll, das uns über begriffliche Möglichkeiten aufklärt.

## 8.1 DIE GRUNDLAGE UNSERES MODALEN WISSENS

Wir hatten in Kapitel 2 und 3 schon gesehen, dass es im Fall der naturwissenschaftlichen Gedankenexperimente schwierig ist, zu erklären, worin unsere Urteile in Bezug auf imaginäre physikalische Sachverhalte eigentlich ihre *Rechtfertigungsgrundlage* haben und ob sich – vielleicht determiniert durch ihre Rechtfertigungsgrundlage – etwas dazu sagen lässt, wie man solche modalen Urteile systematisch verbessern kann. In Bezug auf naturwissenschaftliche (bzw. physikalische) Gedankenexperimente hatten wir dabei mindestens drei verschiedene Auffassungen kennen gelernt. Die Vertreter dieser Auffassungen waren Mach, Popper und Kuhn.

### 8.1.1 METHODOLOGIE DES GEDANKENEXPERIMENTS BEI MACH, POPPER UND KUHN

Da wäre zunächst die Auffassung Ernst Machs (später aufgegriffen durch Roy Sorensen und die Autoren, die sich mit *mental modelling* beschäftigen), derzufolge unsere Urteile im Gedankenexperiment auf physikalischen Intuitionen beruhen, die uns durch die Evolutionsgeschichte mitgegeben sind. Diese Urteile sind zuverlässig, insofern sie durch den kausalen Kontakt mit physikalischen Gegenständen entstanden sind. Für Mach folgen daraus mindestens zwei methodologische Prinzipien. Auf der Seite des Entwicklers des Gedankenexperiments muss dafür Sorge getragen werden, dass es sich um Umstände handelt, die in der Evolutionsgeschichte plausibler Weise eine selektive Rolle gespielt haben. Auf der Seite des Urteilenden muss darauf geachtet werden, dass es sich um ein Urteil handelt, bei dem man sich – in irgendeinem leider nicht näher bestimmten Sinne – bewusst ist, dass man zur Hervorbringung dieses Urteils selbst nichts hinzugefügt hat. Was Mach vermutlich meinte, sind „intuitive Urteile“ im Sinne der Psychologie:

We will call any judgment an *intuitive judgment*, or more briefly an *intuition*, just in case that judgment is not made on the basis of some kind of explicit reasoning process that a person can consciously observe. Intuitions are judgments that grow, rather, out of an underground process, of whatever kind, that cannot be directly observed. (Gopnik und Schwitzgebel [129], 77)

In Bezug auf die Intuitionsbedingung darf man sicherlich Zweifel erheben, inwiefern die Tatsache, dass es sich um eine Überzeugung handelt, die nicht durch Überlegung, sondern sozusagen „spontan“ entstanden ist, darauf schließen lässt, dass sie Informationen aus der Evolutionsgeschichte trägt. Angenommen, die Auffassung der *mental modelling*-Fraktion wäre gar nicht falsch, dass auch unsere neu hinzu gewonnen Überzeugungen in solche Intuitionen Eingang finden können, dann wäre der (von Mach veranschlagte) *zusätzliche* Erkenntnisnutzen des Ge-

dankenexperiments in Frage gestellt, weil am qualitativen Aspekt der Vorstellung kein Unterschied mehr festgestellt werden könnte zwischen „alten“ und „neuen“ Intuitionen.

In Bezug auf die Bedingungen, die an die Beschreibung der imaginären Fälle gestellt werden, ist zu betonen, dass das Gedankenexperiment nach Mach in der Physik nur in einer begrenzten Epoche eine Rolle spielen konnte. Nun, nachdem die Physik die Ebene des Mesokosmos hinter sich gelassen hat (über die physikalischen Eigenschaften mittelgroßer Gegenstände ist praktisch alles gesagt<sup>333</sup>), scheinen Gedankenexperimente keinen rechten Sinn mehr zu machen. Aber was ist dann mit Einstein, der Relativitätstheorie und der Quantenmechanik?

Eine zweite Auffassung hatten wir bei Karl Popper kennen gelernt. Bei ihm ging es beim Gedankenexperiment nur um die Feststellung theorieinterner Widersprüche. Unsere Zuverlässigkeit, diese zu registrieren, erscheint unproblematisch, solange man nur Prognosen einer Theorie für einen bestimmten Fall abzuleiten braucht. Was eher von Interesse ist, ist die Fallbeschreibung. Natürlich sollte es sich um einen physikalisch möglichen Fall aus dem Gegenstandsbereich der fraglichen Theorie handeln. Deswegen ist hier insbesondere darauf zu achten, dass bei der Beschreibung nur solche Annahmen verwendet werden, die die zu kritisierende Theorie auch gestattet. Auch der Einsatzbereich von Poppers Auffassung kam uns zu eingeschränkt vor. Kaum ein Gedankenexperiment deckt in diesem Sinne theorieinterne Widersprüche auf.

#### 8.1.2 METHODOLOGIE UND ERKENNTNISTHEORIE DES GEDANKENEXPERIMENTS

In Kapitel 3 hatten wir gesehen, dass die Wahrheit vermutlich – wie so häufig – irgendwo dazwischen liegt. Manche Gedankenexperimente haben wohl tatsächlich so funktioniert, wie Mach sich das gedacht hat. *Moderne* Gedankenexperimente versuchen (im Allgemeinen) aber nicht mehr unsere physikalischen Intuitionen aus der Evolutionsgeschichte abzufragen, sondern zielen auf begriffliche Probleme mit der physikalischen Theorie („Wann ist eine physikalische Theorie „vollständig“?“) oder auf die Abgrenzung ihres Anwendungsgebiets („Ist eine Theorie der Gravitation eine Theorie der Beschleunigung?“). Was methodologisch bei allen Verwendungsweisen zu beachten bleibt, ist, dass Hintergrundannahmen des Gedankenexperiments – seien sie nun (im engeren Sinne) physikalischer Natur oder allgemeine Erklärungsprinzipien – auf den fraglichen Fall *spezialisiert* unstrittig sein sollten. Darin besteht gerade der besondere Witz des Gedankenexperiments: man kann kein allgemeines Argument präsentieren, weil dazu allgemeine Prinzi-

---

<sup>333</sup> Man denke an den Physikerwitz, bei dem ein Physiker seinen Kollegen – einen String-Theoretiker – fragt, ob es bereits irgendeine empirische Bestätigung für die Theorie gibt. Der String-Theoretiker darauf: „Dinge fallen runter.“ Gegenwärtige Theorienkonflikte in der Physik spielen auf der Beschreibungsebene des Mesokosmos eben fast keine Rolle mehr.

prien auszubuchstabieren wären, die vermutlich noch lange diskussionswürdig blieben. Stattdessen konstruiert man einen imaginären Fall, in dem das allgemeine Prinzip auf eine solch unproblematische Weise instantiiert ist, dass man den Nebenkriegsschauplatz der genauen Formulierung und Etablierung dieses Prinzips zunächst gar nicht zu eröffnen braucht.

Solange unstrittig ist, dass es sich in dem beschriebenen Fall um ein Problem handelt, das in den intendierten Anwendungsbereich der Theorie fällt, ist die Wirklichkeitsnähe des imaginären Falls nur über die Geltung logisch abhängiger Naturgesetze<sup>334</sup> bestimmt.<sup>335</sup> Etwas, was nicht im intendierten Anwendungsbereich liegt, ist nicht relevant. Wie breit oder eng man diesen auffassen will, ist eine Frage, danach, wie man mit seiner Theorie umgeht. Man kann den empirischen Gehalt einer Theorie auch dadurch einschränken, dass man ihren Anwendungsbereich verkleinert. Ist man sehr liberal in Bezug auf den Anwendungsbereich, zieht die zweite Restriktion: Gedankenexperimente, deren Beschreibung bereits die Falschheit der zu diskutierenden Targetthese voraussetzen, begehen eine *petitio principii*.

Insofern tatsächlich intuitive Urteile vorkommen und es sich nicht um Umstände handelt, in denen man annehmen darf, dass es eine robuste physikalische Intuition gibt, die sich evolutionär rechtfertigen lässt (oder es sich um Ereignisse handelt, die man bereits beobachtet hat), dann sind diese entweder mutmaßlich unzuverlässig, oder, falls bezogen auf Erklärungsprinzipien allgemeiner Art, eher ein Ausdruck dessen, was man meint, plausibler Weise von einer physikalischen Theorie über einen bestimmten Gegenstandsbereich verlangen zu dürfen. Letzteres ist vermutlich der Fall im EPR-Gedankenexperiment (vgl. Teil 3).

Wie wir gesehen haben, bleiben Gedankenexperimente auch in den weniger problematischen Fällen *fallibel*. Auch in Bezug auf mittelgroße Gegenstände sind unsere physikalischen Intuitionen oft fehlerhaft; unsere Wünsche in Bezug auf die Erklärungsvollständigkeit einer physikalischen Theorie können schlicht uneinlösbar sein, weil es keine wahren Theorie gibt, die unseren Wünschen entspreche, etc. Gegeben, dass die Methode des Gedankenexperiments innerhalb der Naturwissenschaften nur ein Teil eines weiteren Methodenrepertoires ist, mit dem Urteile über Annehmbarkeit und Ablehnung empirischer Theorien getroffen werden, kann die Fallibilität der Methode des Gedankenexperiments korrigiert werden. Wie wir im EPR-Fall gesehen haben, können Hintergrundannahmen, wie die, dass es eine empirisch äquivalente *hidden-variables* Theorie geben muss, durch empirische Überprüfung als falsch erwiesen werden. Das Problem des EPR-Arguments lag nicht an „falschen“ Gütekriterien für wissenschaftliche Theorien, son-

---

<sup>334</sup> Also von der Target-Theorie „logisch abhängig“. Ein Szenario (im Sinne von  $\Gamma_3$ ), was solchen Gesetzen widerspräche, wäre eine *petitio principii*.

<sup>335</sup> Möglicherweise ist dies auch, was Popper sagen wollte, als er forderte, dass die „Annahmen“ im Gedankenexperiment dem Gegner entgegenkommen müssen. Man denke etwa an die – bereits erwähnte – Kritik Perrys am Zombiargument (Kapitel 5.3.4). Das Zombieszenario wird zurückgewiesen, weil in ihm Annahmen gemacht werden (Epiphänomenalismus), die von der Target-These logisch nicht unabhängig sind.

dern an der falschen empirischen Vermutung, dass es Theorien gibt, die diese Kriterien in höherem Maße erfüllen können.

Kann man Ähnliches auch zu Gedankenexperimenten in der Philosophie sagen?

#### 8.1.2.1 *Methodologie des Gedankenexperiments bei Descartes*

Auf jeden Fall *konnte* man das. Gedankenexperimente sind nicht erst seit die analytische Philosophie auf den Plan getreten ist Bestandteil des philosophischen Methodenrepertoires. Nehmen wir zum Beispiel René Descartes. Descartes machte von Gedankenexperimenten relativ regen Gebrauch. Aus der Tatsache, dass er sich vorstellen konnte, ohne seinen Körper zu existieren, folgerte er, dass Körper und Seele verschiedene Substanzen sind und unabhängig existieren können.

Natürlich war er sich im Klaren darüber, dass eine solche Argumentation eine Begründung braucht. Descartes erhält sie daraus, dass er aus der (ebenfalls intuitiven) Einsicht, dass ein vollkommener Gott existieren muss, die Überzeugung zieht, dass dieser Gott ihn zumindest insofern mit unfehlbaren Intuitionen ausgestattet hat, wie es um Fragen geht, deren Antworten ihm unwillkürlich und evident erscheinen.

Und da ich ja erstens weiß, daß alles, was ich klar und deutlich verstehe, in der Weise von Gott geschaffen werden kann, wie ich es verstehe, so genügt es, eine Sache ohne eine andere klar und deutlich verstehen zu können, um mir die Gewißheit zu geben, daß die eine von der anderen verschieden ist, da wenigstens Gott sie getrennt setzen kann. [...] Da ich ja einerseits eine klare und deutliche Vorstellung meiner selbst habe, sofern ich nur ein denkendes, nicht ausgedehntes Wesen bin, und andererseits eine deutliche Vorstellung vom Körper, sofern er nur ein ausgedehntes, nicht denkendes Wesen ist – so ist, sage ich, soviel gewiß, daß ich von meinem Körper wahrhaft verschieden bin und ohne ihn existieren kann. (Descartes [86], 141)

Die Rechtfertigungsgrundlage des Gedankenexperimentierens wäre in diesem Fall die These, dass Gott uns mit einem partiell zuverlässigen Erkenntnisapparat ausgestattet hat.<sup>336</sup> Die daraus folgende Systematik ist, dass – weil man sich bei diesem partiell zuverlässigen Erkenntnisapparat nur auf die klaren und deutlichen Vorstellungen verlassen kann, die unwillkürlich zu sein scheinen und nicht einer Willensentscheidung bedürfen – man nur dann gedankenexperimentieren darf, wenn es sich um eine Sache handelt, zu der man klare und deutliche Intuitionen hat.

Soweit wir es bis jetzt betrachtet haben, geht es in der Philosophie um Theorien modalen Wissens bestimmter Art. Wir urteilen in Gedankenexperimenten über die begriffliche bzw. logische oder metaphysische Möglichkeit eines bestimmten Sachverhalts. Nähmen wir die soeben gemachten Beobachtungen als Paradigma, sollte eine solche Theorie mindestens zwei Fragen beantworten können:

---

<sup>336</sup> Gott selbst sollte dabei natürlich irgendeinen privilegierten Zugang zu diesen Modalitäten haben. Das Verhältnis in dem Gott zu den Modalitäten steht, ist bei Descartes nicht unbedingt besonders klar. Vgl. etwa Bennett [21].

- (1) Gibt es einen genauer angebbaren kognitiven Akt, der (mutmaßlich) auf zuverlässige Weise zu wahren modalen Urteilen führt?
- (2) Kann man die Existenz und die Zuverlässigkeit dieses Aktes (in Bezug auf die besondere Natur modaler Tatsachen) erklären?

Descartes' Antwort auf die erste Frage wäre positiv und bestünde in der Theorie, dass wir solchen Vorstellungen, die sich uns auf „klare und deutliche“ Weise präsentieren, vertrauen dürfen, weil sie nicht nur zuverlässig zu wahren modalen Urteilen führen, sondern sogar unfehlbar (und deshalb unbezweifelbar) sind.

Auf die zweite Frage würde Descartes antworten, dass die Zuverlässigkeit dieses Vorstellungsaktes auf die Güte Gottes zurückgeführt werden kann sowie die Tatsache, dass Gott die Grundlage alles Möglichen und Notwendigen ist. Gott hat uns mit einem Erkenntnisapparat ausgestattet, der zwar nur begrenzt infallibel ist, dessen Grenzen wir aber an den qualitativen Aspekten unserer Vorstellungsakte immer feststellen können (was „klar und deutlich“ ist, liegt innerhalb der Grenze).

#### 8.1.2.2 Non-Kognitivismus bezüglich modaler Urteile

Man kann sich vorstellen, dass manche Theorien modalen Wissens von einer besonderen Auffassung modaler Tatsachen ausgehen und in gewissem Sinne einer Beantwortung der zweiten Frage einer Antwort auf die erste Frage den Vorschub leisten: Non-Kognitivisten in Bezug auf modale Urteile, bzw. Konventionalisten in Bezug auf modale Tatsachen, würden unser modales „Wissen“ als Überzeugungen auffassen, die sich in Übereinstimmung mit denjenigen sozialen Praktiken befinden, die wir zur intersubjektiven Abstimmung modaler Urteile entwickelt haben. Wenn die Natur modaler Tatsachen auf diese Weise geklärt ist, ergibt sich eine Beantwortung der ersten Frage von selbst: unabhängig von den qualitativen Aspekten irgendwelcher Vorstellungsakte sind (trivialer Weise) insbesondere diejenigen Verfahrensweisen zur Erreichung modaler Urteile zuverlässig (und führen zu „wahren“ modalen Urteilen), die denjenigen sozialen Praktiken folgen, die erfolgreich festlegen, was die modalen Tatsachen *sind*.

Hierbei muss die *Erklärungsrichtung* davon unterschieden werden, wie für diese Position *argumentiert* wird. Während der konventionelle bzw. non-kognitive Charakter modaler Tatsachen bzw. Urteile *erklärt*, wieso bestimmte Verfahren zu „wahren“ modalen Urteilen führen, kann diese Position sehr wohl mit den Eigenheiten des Verfahrens zur Erreichung wahrer modaler Urteile *begründet* werden. Dies ist z.B. dann der Fall, wenn eine konventionalistische bzw. non-kognitivistische Position zur Natur modaler Urteile über einen Schluss auf die beste Erklärung für unsere modale Epistemologie etabliert wird.<sup>337</sup> Es wird also für den Non-Kognitivismus argumentiert, weil sich aus unserer Methode zur Etablierung modaler Urteile sonst kein Sinn machen lässt.

Auf diese Weise argumentiert beispielsweise André Fuhrmann für einen Non-Kognitivismus in Bezug auf Modalitäten. Er geht aus von der Frage, wie sich die

---

<sup>337</sup> Vgl. Sidelle [297], Fuhrmann [112].

Praxis des philosophischen Gedankenexperimentierens rechtfertigen lässt. Existiert so etwas wie eine allgemein zuverlässige Beziehung zwischen dem Bereich des Möglichen und dem Bereich des Vorstellbaren? Genauer gesagt untersucht er die Maxime (M),

(M)  $\forall x (B(x) \rightarrow \forall p (p \text{ ist möglich} \leftrightarrow x \text{ kann sich } p \text{ vorstellen}))$

und fragt danach, wie sich die Bedingungen B auf nicht-zirkuläre Weise angeben lassen. Fuhrmann unterscheidet hierbei zunächst zwei Interpretationen der Maxime. Eine „sokratische“ und eine „euthyphronische“ Interpretation.<sup>338</sup> Nach der ersten (der „sokratischen“) Interpretation ist (M) folgendermaßen zu interpretieren (vgl. Fuhrmann [112]):

- (a) Es gibt eine Beziehung R so, dass ein Vorstellender, indem er sich einen Sachverhalt p vorstellt, in der Beziehung R zu p steht.
- (b) Es gibt ferner Bedingungen B so, dass wenn ein Vorstellender  $x$  die Bedingung B erfüllt und  $x$  in der Beziehung R zu p steht, p möglich ist.
- (c) Die Bedingungen B sind normalerweise erfüllt oder können absichtsvoll herbeigeführt werden.

Diese Interpretation kann leicht an der Parallele zur visuellen Wahrnehmung erläutert werden: So steht bei der visuellen Wahrnehmung eines Sachverhalts ein visuell Wahrnehmender in einer – in Bezug auf die Funktionsweise des Wahrnehmungsapparats explizierbaren – kausalen Relation zu aktuellen Gegenständen und Relationen, die den fraglichen Sachverhalt konstituieren (a), der Wahrnehmende muss Bedingungen der Normalsichtigkeit, etc. erfüllen, und das Vorliegen dieser Bedingungen garantiert, dass der Sachverhalt auch aktual besteht (b). Außerdem können diese Bedingungen absichtsvoll herbeigeführt werden, indem man sich um günstige Lichtverhältnisse, ein Mikroskop, etc. bemüht.

Das Problem der sokratischen Interpretation von (M) besteht darin, dass die Beziehung R nicht in zirkelfreier Weise erläutert werden kann. Entweder man bestimmt R einfach als diejenige Beziehung, in der ein Vorstellender zu einem Sachverhalt p genau dann steht, wenn p möglich ist, oder R kann bestenfalls negativ als eine Beziehung expliziert werden, die nicht kausal ist und folglich nicht nach dem Modell der Wahrnehmung gedacht werden kann, weil modale Tatsachen – nach Fuhrmann – qua modaler Tatsache nicht in kausaler Beziehung zu uns stehen können.

---

<sup>338</sup> In Anlehnung an das so genannte „Euthyphron-Dilemma“. In Platons Dialog „Eutyphron: Über die Frömmigkeit“ diskutiert Sokrates gegen die These seines Gesprächspartners Eutyphron, dass alles gut sei, was gottgefällig ist. Nach der Auffassung, die Sokrates verteidigt, muss es einen von Gottgefälligkeit unabhängigen Maßstab für das Gute geben. Vgl. Wright [355], Wright [357].

Nach der zweiten (der „euthyphronischen“) Interpretation ist (M) hingegen parallel zur Reaktionsabhängigkeitsanalyse von Farben zu verstehen. Wie in (F)

- (F)  $x$  ist rot  $\leftrightarrow$  jeder Normalsinnige (dem  $x$  unter normalen Bedingungen präsentiert wird) würde auf  $x$  mit einer Rotempfindung reagieren.

würden die Bedingungen B in (M) also die Normal- oder Idealbedingungen angeben, in denen ein Vorstellender sich befinden muss, um zuverlässig modale Urteile abzugeben.

In dieser euthyphronischen Variante bleibt die genaue Beziehung des Vorstellenden zum Vorgestellten unerwähnt, da die „Objektseite“ und ihre Beschaffenheit in der Reaktionsabhängigkeitsanalyse zunächst nicht auftritt. Es ist aber so, dass solange modale Urteile kognitiv aufgefasst werden, (M) auch in dieser Interpretation leer bleibt. Betrachten wir (F), dann stellen wir fest, dass (F) nur dann eine gehaltvolle Erläuterung unseres Farburteilens darstellt, wenn die Normalitätsbedingungen rechts vom ‚ $\leftrightarrow$ ‘ nicht einfach dadurch angegeben werden, dass man eben dann in Normalbedingungen ist, wenn man rote Gegenstände als rote Gegenstände wahrnimmt. Bei Farburteilen wissen wir zumindest soweit über ihr Zustandekommen gut genug Bescheid, um Umstände, die mit dem angenommenen kausalen Prozess, der zu zuverlässigen Urteilen führt, zu stark konfligieren, angeben zu können: Auf LSD und durch eine bunte Brille sollte man in einem Raum mit farbiger Beleuchtung keine voreiligen Urteile über die Farbe der Tapete abgeben. Eine solche Ätiologie ist aber in Bezug auf modale Urteile nicht auf dieselbe Weise möglich. Zwar kann man darauf hinweisen, dass wir manche modale Urteile zurückweisen und uns dabei darauf berufen, dass sie unter nicht-idealen Umständen zustande gekommen sind, allerdings ist diese Praxis der Kritik zwar ein Faktum aber damit nicht automatisch gerechtfertigt, solange eben nicht angegeben werden kann, warum solche Umstände zu unzuverlässigen modalen Urteilen führen, was nicht geht, weil man kein unabhängiges Kriterium für zutreffende modale Urteile hat, außer unseren Urteilen und der Tatsache, dass manche der Kritik standhalten.

Fuhrmann schlägt stattdessen vor, die Annahme, dass modale Urteile kognitiv sind, aufzugeben. „Kognitiv“ bedeutet hierbei, dass jeder Streit um ein Urteil einer bestimmten Art auf einen kognitiven Mangel auf Seiten einer der beiden streitenden Parteien zurückgeführt werden kann.<sup>339</sup> Natürlich kann es auch dann noch „Fehlleistungen“ geben, nur sind diese dann nicht kognitiver Art, sondern bestehen schlicht darin, ein modales Urteil gefällt zu haben, das unseren Praktiken der Kritik modalen Urteile nicht standhält.

Fuhrmann argumentiert vom prinzipiellen Scheitern einer gehaltvollen modalen Epistemologie zur Aufgabe bestimmter Annahmen in Bezug auf die Natur modalen Urteile bzw. Tatsachen. Stimmt man dieser Auffassung modalen Urteile bzw. Tatsachen dann zu, ist die dazu passende Erkenntnistheorie trivialerweise gerechtfertigt.

---

<sup>339</sup> Vgl. Wright [355], Wright [357].

## 8.2 VORSTELLBARKEIT ALS KRITIKGRUNDLAGE AN DER TARGETTHESE

So sympathisch diese Auffassung auch ist, die Meisten halten modale Urteile für kognitive Urteile, die wahr sind, wenn sie mit objektiven modalen Tatsachen übereinstimmen und werden Fuhrmanns Argumentation nicht folgen wollen.<sup>340</sup> Es ist auch nicht klar, ob man seinem Argument folgen muss. Könnte es nicht sein, dass man unsere beiden Fragen an eine Theorie modalen Wissens nicht vielleicht trennen und so Fuhrmanns Problematik umgehen kann? Fuhrmann scheint davon auszugehen, dass eine Rechtfertigung unserer Praktiken zur Aufstellung und Kritik modaler Urteile Bezug auf ihre besondere Natur nehmen *muss*. An irgendeiner Stelle muss man angeben können, warum die Bedingungen für gute Gedankenexperimente und richtiges Vorstellen auch tatsächlich die *richtigen* Bedingungen sind.

Dieser Auffassung muss man nicht folgen. Es scheint durchaus auch folgende Strategie prinzipiell möglich zu sein: Man ignoriert zunächst die zweite Frage und konzentriert sich auf die erste. Kann man in gehaltvoller (nicht-zirkulärer) Weise Bedingungen angeben, unter denen man modale Urteile fällen kann, von denen es keinen Grund gibt, anzunehmen, dass sie unzuverlässig sind? Natürlich ist so eine Verteidigung nicht in letzter Hinsicht befriedigend. Wir haben aber in Kapitel 5 schon gesehen, dass die Feststellung, dass es keine Gründe gibt, einer Methode zu misstrauen, als Rechtfertigung einer Erkenntnismethode hinreichend erscheint, insbesondere wenn es sich bei der fraglichen Erkenntnismethode um ein *Instrument der Kritik* handelt.

Letzterer Strategie folgen in der Tat viele Ansätze in der modalen Epistemologie, also dem Bereich der Erkenntnistheorie, der zu klären versucht, woher unser Wissen von Notwendigkeit und Möglichkeit stammt.<sup>341</sup> Dabei geht es dann hauptsächlich darum, in klarer Weise anzugeben, was man genau meint, wenn man behauptet, dass etwas „vorstellbar“ oder „intuitiv“ ist und dann zeigt, dass Urteile, die zustande gekommen sind, weil sie in einem relevanten Sinne einen Sachverhalt repräsentieren, der „vorstellbar“ ist, oder weil sie „intuitiv“ waren, nicht nachweislich unzuverlässig sind. Warum dem so ist, wird dann entweder offen gelassen, oder durch eine unabhängige Theorie erklärt. So lässt beispielsweise Stephen Yablo, der diese Strategie in Bezug auf ‚Vorstellbarkeit‘ verfolgt, die Frage absichtlich offen. Yablo gesteht zwar zu, dass es (a) keine vom Vorstellbarkeitstest

---

<sup>340</sup> Solange man ‚objektiv‘ nicht als ‚unabhängig von unseren Praktiken‘ versteht, ist die Objektivität modaler Tatsachen durch den Non-Kognitivismus nicht in Frage gestellt. Relevant ist hier ‚kognitiv‘.

<sup>341</sup> Die Fragestellung der modalen Epistemologie ist nicht (unmittelbar) die Frage danach, woher wir Wissen von denjenigen Objekten haben können, über die in der Mögliche-Welten-Semantik quantifiziert wird. Diese Fragen befinden sich zunächst auf zwei verschiedenen Ebenen der Analyse. Zur Frage danach, wie wir von möglichen Welten Wissen haben können, vgl. O’Leary-Hawthorne [245] und Lewis [187].

unabhängige Evidenz für den Vorstellbarkeitstest gibt, und dass es (b) auch nahezu ausweglos erscheint, unser Wissen von Modalem in eine kausale Erkenntnistheorie einzupassen, bezweifelt aber, dass diese Gründe schon hinreichen können, den Vorstellbarkeitstest als solchen in Verruf zu bringen:

Taken in a suitably flat-footed way, these claims are again true enough. But the same could be said about various other faculties, notably logical and mathematical intuition; and to judge by our reaction there, they constitute a reason less for mistrusting them than for reconsidering either the nature of the target facts or our nature of access to them. (Yablo [359], 3-4)

Paul Tidman, der zunächst die These verteidigt, dass Gedankenexperimente auf Intuitionen fußen (und nicht auf Vorstellbarkeit oder Logik<sup>342</sup>) und diese als zuverlässige Dispositionen identifiziert, beantwortet die Frage, *warum* wir solche Dispositionen haben, auf überraschend cartesische Weise:

The point that emerges [...] is [that] the Theist is in particularly good shape when it comes to explaining how we have come to have modal intuitions. The Theist can attribute our possession of these intuitions as due to the benevolent workings of our creator who desires that we be able to come to know the truth in modal matters. (Tidman [325], 199)

Tidman weist allerdings ebenfalls darauf hin, dass eine solche Antwort auf Frage (2) nicht nötig ist zur Fundierung und Rechtfertigung der relevanten Urteile:

Although it is, I think, genuinely puzzling to see how one can naturalistically explain our having a reliable faculty of intuition, it does not follow that only the Theist can adequately explain how we have modal knowledge. For even if it is just a bit of luck that we have these intuitions, the point is that we do have them. (Tidman [325], 200)

Diese Problematik, wie sich unsere modalen Urteile rechtfertigen lassen, ist insbesondere in den letzten Jahren zu einem eigenständigen Gebiet der modernen Erkenntnistheorie ausgewachsen. Eine umfassende Diskussion würde daher eine eigene Untersuchung erfordern. Schließlich geht es dann nicht mehr nur um die Methodologie der Philosophie, sondern um eine Form von Wissen, das wir in allen möglichen Lebensbereichen verwenden, sei es zum Einparken oder um akzidentelle von naturgesetzlichen Verallgemeinerungen zu scheiden.

Was uns hier interessieren soll, ist die Frage, ob sich aus dieser Debatte um modale Epistemologie methodologische Hinweise gewinnen lassen, die uns darüber aufklären können, wie sich die Methode des Gedankenexperiments systematisch

---

<sup>342</sup> Die Auffassung, dass sich eine modale Epistemologie am Besten über den Intuitionsbegriff entwickeln lässt, werden wir in dieser Arbeit etwas unter den Tisch fallen lassen. Der Intuitionsbegriff erscheint einfach viel zu schwammig, um halbwegs fruchtbar zu sein. Man vgl. aber Bealer [14], Bealer [15], Bealer [16] sowie Tidman [325], Tidman [326], Tidman [327], Tidman [328].

verbessern lässt. Wir betrachten daher nur diejenigen Ansätze genauer, aus denen solche Hinweise angeblich folgen sollen.

Wir wollen dabei zunächst diskutieren, wie der Vorstellbarkeitstest unter einer phänomenologischen Beschreibung gerechtfertigt werden kann. Dabei gilt es, zu präzisieren, worin Vorstellbarkeit eigentlich besteht, um sodann nachzuweisen, dass Vorstellbarkeit (soweit wir wissen) ein zuverlässiger Indikator für Möglichkeit darstellt. Wir werden feststellen, dass eine solche Verteidigung letztlich als *Rechtfertigung* des Vorstellbarkeitstests unbefriedigend bleiben muss.

In 8.2.1-8.2.4 wird zunächst erläutert, worin das Projekt einer solchen Verteidigung des Vorstellbarkeitstests genau besteht. Es soll darum gehen, eine Präzisierung von ‚es ist vorstellbar, dass p‘ zu finden, die (a) als Interpretation philosophischer Verwendungsweisen dieser Ausdrucksweise adäquat ist, und (b) einen zuverlässigen Indikator für Möglichkeit darstellt. (a) soll dadurch gewährleistet werden, dass notwendige Wahrheitsbedingungen<sup>343</sup> formuliert werden, die für die philosophische Verwendungsweise gelten (8.2.2, 8.2.3). *Zuverlässigkeit* bemisst sich daran, wie es dann gelingt, Beispiele für die angebliche Unzuverlässigkeit des Vorstellbarkeitstests wegzuerklären (8.2.4).

Dabei besteht die Argumentationsstrategie darin, prima facie Gegenbeispiele gegen den Vorstellbarkeitstest als Gegenbeispiele gegen eine Verwendung von ‚vorstellbar‘ zu erweisen, die die relevanten notwendigen Wahrheitsbedingungen nicht erfüllt (8.2.5-8.2.7). Am Ende der Untersuchung steht dann eine Präzisierung des Vorstellbarkeitstests, die alle bekannten Gegenbeispiele abzuweisen vermag (8.2.8), die aber weder für sich beanspruchen kann, eine nicht-psychologistische Analyse des Vorstellbarkeitstests zu sein, noch für sich beanspruchen kann, den Vorstellbarkeitstest als Indikator für Möglichkeit *gerechtfertigt* zu haben (8.2.9).

### 8.2.1 PHILOSOPHISCHE VORSTELLBARKEIT?

Die bekannteste Verteidigung des Vorstellbarkeitstests gegen den Vorwurf, er sei unzuverlässig, stammt von (dem eben bereits erwähnten) Stephen Yablo. In ‚Is Conceivability a Guide to Possibility?‘<sup>344</sup> versucht er (a) zu klären, was genau unter ‚vorstellbar‘ von Philosophen verstanden wird oder werden sollte<sup>345</sup>, wenn sie die Vorstellbarkeit einer Proposition als Evidenz für deren (metaphysische) Möglich-

---

<sup>343</sup> Die „Wahrheitsbedingungen“ eines Satzes sind diejenigen notwendigen und hinreichenden Bedingungen dafür, dass der fragliche Satz wahr ist. „Notwendige Wahrheitsbedingungen“ sind die Bedingungen, die erfüllt sein *müssen*, damit er wahr ist.

<sup>344</sup> Vgl. Yablo [359] Wenn nicht anders angegeben, folgen die Ausführungen in diesem Abschnitt diesem Text von Yablo.

<sup>345</sup> Der Auffassung, dass es so einen ausgezeichneten Sinn von ‚vorstellbar‘ gibt, sind außerdem Hetherington [147], Woolhouse [354], Chalmers [58], Van Cleve [332], und natürlich Descartes. Dass man verschiedene Sinne von ‚vorstellbar‘ in der modalen Epistemologie auseinanderhalten muss, ist auch die Auffassung von Reid [277], der allerdings gegenüber dem Vorstellbarkeitstest nichtsdestotrotz eher skeptisch eingestellt ist.

keit heranziehen, (b) zu untersuchen, ob sich gegen diese Vorgehensweise Gründe finden lassen, d.h., ob sich nachweisen lässt, dass die Beurteilung der Möglichkeit einer Proposition auf Grundlage ihrer Vorstellbarkeit (in diesem Sinne) in der Regel unzuverlässig ist. Entsprechend wird die Frage danach, *weshalb* Vorstellbarkeit ein verlässlicher Indikator für Möglichkeit ist, falls es ein verlässlicher Indikator ist, zunächst zurückgestellt.

Yablos Projekt, den „philosophischen Sinn“ von ‚vorstellbar‘ zu analysieren kann dabei auf dreierlei Weise interpretiert werden: (a) als Bedeutungsanalyse eines tatsächlichen Gebrauchs von ‚vorstellbar‘ in philosophischen Kontexten, (b) als Begriffsexplikation, also als Vorschlag, wie ‚vorstellbar‘ von Philosophen verwendet werden *sollte*, wenn sie für die Möglichkeit eines bestimmten Sachverhalts argumentieren, (c) als Entpsychologisierung der modalen Epistemologie, die unser Wissen von Möglichkeiten als Schlüsse aus Vorstellbarkeitsurteilen rekonstruiert.

In Bezug auf (a) kann man Yablo vorwerfen, dass es fraglich ist, ob es einen spezifisch „philosophischen“ Sinn von ‚vorstellbar‘ überhaupt gibt. Yablo kann hier zur Erwidern das *principle of charity* heranziehen und darauf hinweisen, dass viele Philosophen, die mit „Vorstellbarkeit“ für Möglichkeit argumentieren, ziemlich schlechte Argumente vorgetragen hätten, wenn sie nicht ‚vorstellbar‘ in dem Sinne gemeint hätten, wie Yablo es versteht. Interpretiert man diese Philosophen wohlwollend, dann haben sie alle das gemeint, was Yablo schließlich als „philosophischen“ Sinn von ‚vorstellbar‘ aus dem Hut zaubert. Ob man Yablo letztlich unter einer dieser möglichen Interpretationen seines Projekts zum Erfolg gratulieren darf, wird uns in 8.2.9 beschäftigen.

Zunächst besteht die Strategie, die Yablo verfolgt, darin, festzustellen, was zu den *notwendigen Wahrheitsbedingungen* von ‚p ist vorstellbar‘ im philosophischen Sinn gehört und dann zu untersuchen, ob die Einwände, die sich *gegen* Vorstellbarkeit als verlässliches Mittel zur Beurteilung von Möglichkeit erheben lassen, sich auf einen Sinn von ‚p ist vorstellbar‘ beziehen lassen, der diese notwendigen Wahrheitsbedingungen erfüllt. Falls dem *nicht* so ist, ist das Yablosche Projekt scheinbar erfolgreich, da es dann zumindest keine klaren Gründe gibt, warum Philosophen der Vorstellbarkeit in ihrem Sinn nicht trauen sollten (unabhängig davon, ob sie erklären können, warum sie verlässlich sein sollte). Yablo beginnt seine Untersuchung also mit einer Analyse der notwendigen Wahrheitsbedingungen von ‚p ist vorstellbar‘ im philosophischen Sinn.

Betrachtet man die berühmte Humesche Maxime etwas genauer, so scheint Hume mit

whatever the mind clearly conceives, includes the idea of possible existence, or in other words, [...] nothing we imagine is absolutely impossible. (Hume [154], 32)

eigentlich zweierlei zu behaupten. Zum einen behauptet er, dass dasjenige, was wir uns auf klare Weise vorstellen können, uns als möglich erscheint, zweitens die Behauptung, dass das, was vorstellbar ist, auch möglich *ist*. Die erste Behauptung könnte man sicherlich als eine notwendige Wahrheitsbedingung für das Vorliegen

von Vorstellbarkeit im philosophischen Sinne verstehen. Sowie eine Erfahrungserinnerung sich so präsentiert, als habe man an dem Ereignis teilgenommen, welches der Gegenstand der Erfahrungserinnerung ist, oder so wie eine Wahrnehmung den Gegenstand der Wahrnehmung als wirklich, so präsentiert sich das Vorgestellte bei philosophischen Formen des Vorstellens als möglich. Aus diesen Überlegungen extrahiert Yablo dann entsprechend zwei verschiedene notwendige Wahrheitsbedingungen:

(WB1) ‚p ist vorstellbar‘ (im philosophischen Sinn) ist nur dann wahr, wenn, wenn die Vorstellung von p zutreffend ist, p möglich ist. Genauer:

$$\forall x \forall p ((\text{«}x \text{ stellt sich vor, dass } p \ll = \text{wahr}) \rightarrow ((\text{«}x \text{ stellt sich zutreffend vor, dass } p \ll = \text{wahr}) \rightarrow \diamond p))$$

(WB2) ‚p ist vorstellbar‘ (im philosophischen Sinn) ist nur dann wahr, wenn das vorstellende Subjekt durch die Vorstellung von p dazu motiviert ist, zu glauben, dass p möglich ist.

(WB1) behauptet dabei wohlgernekt lediglich, dass die Bedingungen für das Zutreffen einer Vorstellung zu den Wahrheitsbedingungen von (dem relevanten Sinn von) ‚vorstellbar‘ gehören, wie zu den Wahrheitsbedingungen von ‚wahrnehmen‘ gehört, dass eine zutreffende Wahrnehmung auch ein tatsächliches Objekt hat.

Yablos Strategie besteht nun darin, die Kritiken an der Zuverlässigkeit von Vorstellbarkeitstests durchzumustern und jeweils zu überprüfen, ob die *unzuverlässigen* Vorstellbarkeitstests ‚p ist vorstellbar‘ in einem solchen Sinne verstehen, dass die beiden notwendigen Wahrheitsbedingungen *nicht* erfüllt sind.

### 8.2.2 DER KONFUSIONSVORWURF UND DIE GLAUBBARKEIT VON P

Eine nahe liegende und von Philosophen häufig vertretene Auffassung ist, dass ‚vorstellbar, dass p‘ im Grunde bedeutet, dass man über keine Informationen verfügt, die p ausschließen würden. Yablo zitiert äquivalente Formulierungen dieser Auffassung, wonach man diese Art von Vorstellbarkeit auch ausdrücken kann als: p ist „glaubbar“, oder „nicht unglaublich“. Obwohl diese Auffassung von Vorstellbarkeit zwar verbreitet scheint, hält Yablo sie aus zwei Gründen für zweifelhaft:

- (i) Schlüsse auf die Möglichkeit von p aufgrund der Unkenntnis irgendwelcher Informationen, die p ausschließen würden, sind (laut Yablo) nicht zu rechtfertigen.

- (ii) Solche Schlüsse sind darüber hinaus unzuverlässig, da man leicht Gegenbeispiele anführen kann, die zwar in diesem Sinne vorstellbar, aber nicht möglich sind.

Beide Gründe werden im nächsten Unterkapitel in einem etwas anderen Zusammenhang genauer zu betrachten sein. An dieser Stelle soll nur referiert werden, welche Überlegungen die beiden Gründe gegen Vorstellbarkeit in diesem Sinne stützen. Bezeichnen wir Vorstellbarkeit im Sinne von Glaubbarkeit als ‚Vorstellbarkeit<sub>b</sub>‘ (mit dem Subindex ‚b‘ für ‚believability‘):

**Definition 8.2-1 (vorstellbar<sub>b</sub>):**  $p$  ist vorstellbar<sub>b</sub> für  $S =_{\text{Df}}$   $S$  verfügt über keine Informationen  $I$ , so dass  $I$  mit  $p$  inkompatibel wäre.

Zu (i): Schlüsse auf die Möglichkeit von  $p$  auf Grundlage der Vorstellbarkeit<sub>b</sub> von  $p$  scheinen nun deshalb nicht zu rechtfertigen zu sein, da es so aussieht, als würde man substantielle modale Folgerungen aus dem eigenen Nichtwissen erschließen: Je schlechter es um die Kenntnis der Wahrheit oder Falschheit von  $p$  gestellt zu sein scheint, desto besser scheint es um die Kenntnis des modalen Status von  $p$  bestellt zu sein. Wäre man gewiss, dass  $\neg p$  der Fall ist, so könnte man ausschließen, dass  $p$  überhaupt möglich ist. Besäße man dieses Wissen nicht, sollte man sich in seinem Urteil, dass  $p$  möglich ist, deutlich sicherer sein. Wie zuverlässig scheint ein solcher Schluss zu sein?

Zu (ii): In der Regel werden zwei berühmte Beispiele angeführt, um die Unzuverlässigkeit des Schlusses von ‚vorstellbar<sub>b</sub>  $p$ ‘ zu ‚möglich  $p$ ‘ zu demonstrieren. Das erste Beispiel macht von Saul Kripkes Überlegungen zu so genannter „metaphysischer Notwendigkeit“ Gebrauch, auf die wir im letzten Kapitel eingegangen sind. Das zweite Beispiel ist weniger voraussetzungsreich. Intuitiv sollten allerdings beide Beispiele nachvollziehbar sein.

Folgt man Gottlob Frege und Saul Kripke<sup>346</sup>, dann wird es Menschen geben oder gegeben haben, die nicht über die Information verfügen (bzw. verfügten), dass der Morgenstern mit dem Abendstern identisch ist. Entsprechend ist (bzw. war) es für sie vorstellbar<sub>b</sub>, dass der Morgenstern den Abendstern überdauern wird. Nun sind Morgenstern und Abendstern aber identisch, weshalb der Abendstern den Morgenstern nicht überdauern kann, da nichts sich selbst überdauern kann. Für Menschen, die nicht über die Information verfügen, dass der Morgenstern mit dem Abendstern identisch ist, wäre daher eine Proposition vorstellbar<sub>b</sub>, die in Wahrheit unmöglich ist.

Das zweite Beispiel bezieht sich auf die Tatsache, dass die Goldbachsche Vermutung bisher noch nicht bewiesen wurde. In einem Brief an Leonhard Euler äußerte der Mathematiker Christian Goldbach (1690–1764) am 7. Juni 1742 seine

---

<sup>346</sup> Das Beispiel stammt zwar von Frege, allerdings geht es ihm um die Informativität solcher Identitätsaussagen, nicht um ihre Notwendigkeit, die erst von Kripke im relevanten Zusammenhang betont wurde.

berühmte Vermutung, dass sich jede gerade Zahl  $> 4$  als Summe zweier Primzahlen ausdrücken lässt. Dies wird heute als „binäre Goldbachsche Vermutung“ bezeichnet. Obwohl man davon ausgeht, dass die Vermutung stimmt, konnte bisher kein Beweis gefunden werden.

Bei dieser Vermutung könnte es sich sehr wohl um eine mathematische Wahrheit handeln, deren Beweis wir bloß noch nicht gefunden haben. Wenn es sich allerdings um eine mathematische Wahrheit handelt, dann ist sie – wie die Standardauffassung von mathematischen Wahrheiten behaupten würde – *notwendigerweise* wahr. Entsprechend ist die Goldbachsche Vermutung, wenn sie falsch ist, auch *notwendigerweise* falsch. Da die Goldbachsche Vermutung aber noch nicht bewiesen ist, so könnte man nun argumentieren, ist es sowohl vorstellbar<sub>v</sub>, dass sie wahr ist, wie auch vorstellbar<sub>v</sub>, dass sie falsch ist. *Wäre* Vorstellbarkeit<sub>v</sub> ein verlässlicher Indikator für Möglichkeit, müsste es möglich sein, dass die Goldbachsche Vermutung wahr ist, und möglich sein, dass die Goldbachsche Vermutung falsch ist. Da es sich bei mathematischen Wahrheiten aber (nach der Standardauffassung) um notwendige Wahrheiten handelt, ist nur eines von beidem wirklich eine Möglichkeit. Folglich ist Vorstellbarkeit<sub>v</sub> unzuverlässig.

Wollte man die Kritik an Vorstellbarkeit<sub>v</sub> auf einen Fehlschluss zurückführen, so würde man in diesem Falle sagen, dass eine epistemische Verwendung von „möglich“ und „könnte der Fall sein“ mit einer metaphysischen Verwendung dieser Ausdrücke äquivoliert wird. In der Tat scheint es neben dem Sinne von Vorstellbarkeit als Vorstellbarkeit<sub>v</sub> auch einen Begriff von Möglichkeit im Sinne von Vorstellbarkeit<sub>v</sub> zu geben. Dies bemerkte bereits George Edward Moore:

We, very, very often use expressions of the form ‘It is possible that p’ in such a way that by using such an expression we are making an assertion of our own ignorance on a certain point – an assertion namely that we do not know that p is false. This is certainly one of the very commonest uses of the word ‘possible’; it is a use in which what it expresses is often expressed instead by the use of the word ‘may’. (Moore [212], 232)

Aber ist es wirklich so, dass die philosophische Verwendung von ‚vorstellbar p‘ im Sinne von ‚vorstellbar<sub>v</sub> p‘ zu verstehen ist? Zunächst könnte man mit Verwunderung feststellen, dass in der Tat noch niemand auf die Idee gekommen ist, die Goldbachsche Vermutung mit einem Vorstellbarkeitsargument als falsch zu erweisen. Eine mögliche Erklärung für diese Zurückhaltung könnte einfach darin bestehen, dass Philosophen, die ansonsten Vorstellbarkeitsargumente vorbringen, ‚vorstellbar‘ bei ihren eigenen Argumenten nicht im Sinne von ‚vorstellbar<sub>v</sub>‘ verstehen.

Eine zweite Motivation der Annahme, dass ‚vorstellbar‘ im philosophischen Sinne anders als ‚vorstellbar<sub>v</sub>‘ zu interpretieren ist, legt, laut Yablo, das *principle of charity* nahe: Philosophische Argumente, die sich auf Vorstellbarkeit beziehen, wären so offensichtlich schlecht, würde man ‚vorstellbar‘ mit ‚vorstellbar<sub>v</sub>‘ interpretieren, dass allein das Wohlwollensprinzip uns bereits motivieren sollte, nach alternativen Interpretationen Ausschau zu halten.

Entsprechend argumentiert Yablo dafür, dass zumindest in den meisten philosophischen Kontexten, in denen Vorstellbarkeitsargumente vorgebracht

werden, ‚Vorstellbarkeit‘ nicht im Sinne von ‚Vorstellbarkeit<sub>v</sub>‘ zu verstehen ist. Ein Argument dafür ist das oben bereits erwähnte Problem, dass in allen Fällen, in denen man sich einer Sache gewiss wäre, man sich nicht vorstellen<sub>v</sub> könnte, dass es auch anders sein könnte.

Ein Argument gegen die umgekehrte These, dass alles was ‚vorstellbar<sub>v</sub>‘ ist, auch vorstellbar im philosophischen Sinne ist, ist schon schwieriger zu finden. Ein direktes Gegenbeispiel, so bemerkt Yablo zu Recht, würde erfordern, dass man etwas für glaubbar hält, was bekanntermaßen unmöglich ist. Ein direktes Gegenbeispiel ist vermutlich schon daher kaum zu finden, da, wenn man irgendein *p* im relevanten Sinn für glaubbar hält, man ja gerade *keine* Gründe besitzt, die es ausschließen, dass *p* in irgendeiner möglichen Welt der Fall ist.

Was sich aber finden lässt, sind indirekte Gegenbeispiele, in denen – gegeben alle im Augenblick verfügbare Evidenz –, der modale Status von *p* unentscheidbar ist. Bei diesen Beispielen ist es zwar glaubbar, dass *p* (da man über keine Evidenz verfügt, die *p* ausschließen würde), aber eben auch glaubbar, dass nicht-*p*, in Bezug auf Propositionen die *notwendigerweise* wahr oder notwendigerweise falsch sind. Ein solches Beispiel wäre der bereits angesprochene Fall der Goldbachschen Vermutung. Yablos Argument scheint hier darin zu bestehen, dass Vorstellbarkeit<sub>v</sub> nicht WB2 erfüllt, damit also andere Wahrheitsbedingungen besitzt als der philosophische Sinn von ‚vorstellbar‘. Es für vorstellbar<sub>v</sub> zu halten, dass *p*, motiviert keineswegs immer *prima facie* dazu, zu glauben, dass *p* möglich ist.

### 8.2.3 DER PETITIO-VORWURF

Es scheint jedoch, als ließe sich gegen den Vorstellbarkeitstest folgender Petitio-Einwand in ähnlicher Stoßrichtung vorbringen:

(PV) Wenn es dafür, eine Proposition für vorstellbar zu halten, nur darauf ankommt, nicht zu wissen, dass sie unmöglich ist, dann sind Unmöglichkeiten, da sie immer oder sehr häufig unbemerkt bleiben, genauso oft vorstellbar. Um sich also in einem besonderen Fall auf den Vorstellbarkeitstest verlassen zu können, braucht man einen Grund, zu glauben, dass *in diesem Fall* *p*'s Vorstellbarkeit indiziert, dass *p* in der Tat vorstellbar ist und nicht bloß indiziert, dass man nicht weiß, dass *p* unmöglich ist. D.h., man braucht einen Grund um leugnen zu können, dass:

(\*) obwohl man nicht bemerkt, dass *p* unmöglich ist, ist *p* unmöglich

Da das erste Konjunkt von (\*) aber wahr ist, und man dies ja weiß, kann man (\*) nur dann vernünftigerweise ablehnen, wenn man Gründe besitzt, das zweite Konjunkt abzulehnen. Dies behauptet aber, dass *p* unmöglich ist. D.h., man muss bereits wissen, dass *p* möglich ist, bevor man aufgrund von *p*'s Vorstellbarkeit<sub>v</sub> darauf schließen kann, dass es möglich ist.

Dieser Einwand ist sicherlich einer der stärksten, den man gegen die Vorstellbarkeitsthese vorbringen kann. Die Frage ist, ob er auch gerechtfertigt ist. (PV) beruht an zentraler Stelle auf einer anscheinend empirisch gehaltvollen Hypothese; nämlich, dass unerkannte Unmöglichkeiten oft bzw. fast immer vorstellbar sind.

Ob (PV) durchschlagend ist, hängt davon ab, ob sich nicht doch ein Sinn von ‚vorstellbar‘ angeben lässt, der mögliche positive Instanzen von vorstellbaren Unmöglichkeiten wegerklären kann. Mustern wir solche Interpretationen einmal durch:

Wenn man Vorstellbarkeit so versteht, dass es glaubbar ist, dass  $p$  möglich ist, gewinnt die in (PV) enthaltene empirisch scheinende Hypothese einiges an Plausibilität. Unter einer solchen Definition von Vorstellbarkeit (mit ‚bp‘ für ‚believable that  $p$  is possible‘)

**Definition 8.2-2 (vorstellbar<sub>bp</sub>):**  $p$  ist vorstellbar<sub>bp</sub> =<sub>Def</sub> Es ist glaubbar, dass  $p$  möglich ist.

wird die Hypothese, dass jemand, der nicht weiß, dass  $p$  unmöglich ist, die Möglichkeit von  $p$  für glaubbar halten wird, sogar fast zu einer analytischen Wahrheit (und eine Angabe confirmierender Instanzen erledigt sich). Vorstellbarkeit<sub>bp</sub> ist aber schwächer als Vorstellbarkeit<sub>v</sub>, die Yablo schon als Kandidaten für einen philosophischen Sinn von ‚vorstellbar‘ ausgeschaltet hat. Wenn wir aber stärkere Interpretationen von Vorstellbarkeit zu Rate ziehen, ist die Frage, ob wir Gegenbeispiele gegen die Vorstellbarkeitsthese finden können. Wie wir bereits sagten, sind in der Literatur zu diesem Thema so genannte Kripke-Fälle (also a posteriori Notwendigkeiten) die Hauptgegenbeispiele gegen eine wie auch immer geartete Vorstellbarkeitsthese. Das entsprechende Argument für unsere Hypothese wäre Folgendes (Yablo [359], 21):

- (G1) Immer wenn  $p$  a posteriori falsch ist, finde ich es vorstellbar dass  $p$ , unabhängig davon, ob  $p$  möglich ist oder nicht.
  - (G2) A posteriori Falschheiten sind oft unmöglich.
- 
- (G3) Folglich sind a posteriori Falschheiten oft vorstellbar, obwohl sie unmöglich sind.

Nimmt man aber die Existenz von a posteriori Notwendigkeiten zum Zwecke des Arguments einmal an, ist nicht klar, wie ein solches Argument funktionieren sollte. *Innerhalb* einer kausalen Referenztheorie kann ich – wie im letzten Kapitel erläutert – zwischen zwei Bedeutungsbestandteilen unterscheiden: einmal dem extern festgelegten Bestandteil (sekundäre Proposition), einmal dem intern festgelegten Bestandteil (primäre Proposition). Wenn ich die Proposition  $p$  in dem Sinne verstehe, dass ich mich auf den extern festgelegten Bedeutungsbestandteil beziehe, dann kann ich mir nicht vorstellen, dass  $p$ , wenn  $p$  in der Tat unmöglich ist. Auch die Babylonier konnten sich vermutlich nicht vorstellen, dass  $Venus \neq Venus$ . Im Fall von

**Definition 8.2-3 (vorstellbar<sub>ib</sub>):** p ist vorstellbar<sub>ib</sub> =<sub>Df</sub> Ich kann mir ausmalen, wie ich p auf wahre Weise glaube.

kann ich, wenn p eine a posteriori Falschheit ist, mir p nicht vorstellen. Beziehe ich mich auf den internen Bedeutungsaspekt und verstehe Vorstellbarkeit in diesem Sinne, also entweder als

**Definition 8.2-4 (vorstellbar<sub>ib</sub>):** p ist vorstellbar<sub>ib</sub> =<sub>Df</sub> Ich kann mir ausmalen<sup>347</sup>, gerechtfertigt zu glauben, dass p.

oder

**Definition 8.2-5 (vorstellbar<sub>ap</sub>):** p ist vorstellbar<sub>ap</sub> =<sub>Df</sub> Ich kann mir ausmalen, wie ich etwas Wahres mit meinem gegenwärtigen p-Gedanken denke.

kann ich mir p in diesem Sinne zwar vorstellen, was mir aber nur versichert, dass ich mit meinem gegenwärtigen Gedanken<sup>348</sup> vielleicht eine Wahrheit denken *könnte*, mir aber nicht den Eindruck gibt, dass die Proposition, die ich mit meinem Gedanken in der Tat meine, auch möglich ist (zumindest ist so Yablos Intuition). Definition 8.2-5 zeigt sich also (scheinbar) unbeeindruckt von Putnam/Kripke-Gegenbeispielen. Ist dies die gesuchte Bedeutung von ‚vorstellbar‘?

#### 8.2.4 „POSITIVE“ VORSTELLBARKEIT

In der Tat schlägt Yablo schließliche folgende Analyse als endgültige Lösung vor:

**Definition 8.2-6 (vorstellbar im philosophischen Sinn):** p ist vorstellbar im philosophischen Sinn =<sub>Df</sub> Ich kann mir eine Welt ausmalen, die ich als Verifikation von p ansehen/akzeptieren würde.

Wie Yablo zugibt, können modale Irrtümer natürlich auch bei dieser Art von Vorstellbarkeitstest auftreten. Es muss auch schließlich nicht gezeigt werden, dass der Vorstellbarkeitstest infallibel ist. Dass etwas in Yablos Sinne vorstellbar ist, ist zwar *prima facie* Evidenz für seine metaphysische Möglichkeit, aber es kann na-

---

<sup>347</sup> Yablo expliziert ‚conceivability‘ an dieser Stelle mit ‚imaginability‘ (was hier mit ‚ausmalen‘ übersetzt wird – dies soll zunächst aber nur ein Platzhalter sein). Da beide Ausdrücke in einem technischen, philosophischen Sinn oft synonym verwendet werden (Gedankenexperimente fangen typischerweise damit an, dass der Leser aufgefordert wird, sich einen bestimmten Sachverhalt „vorzustellen“, wobei ‚vorstellen‘ abwechselnd als ‚to conceive‘ oder ‚to imagine‘ wiedergegeben wird), könnte man einen starken Zirkularitätseinwand gegen diese Erklärung vorbringen. Wir werden darauf zurückkommen.

<sup>348</sup> ‚Gedanke‘ im Sinne Freges bzw. im Sinne einer primären Proposition/Intension eines Satzes.

türlich sein, dass ich beim Vorstellungstest für  $p$  Fehler gemacht habe. Ein solcher Fehler mag folgende Ursachen haben: Es gibt eine Proposition  $q$ , so dass

- (i)  $q$  wahr ist,
- (ii) wenn  $q$ , dann  $\Box\neg p$

und

- (iii) dass ich  $p$  für vorstellbar gehalten habe, kann dadurch erklärt werden, dass ich (i) oder (ii) für falsch gehalten habe, oder mir der Wahrheit von (i) oder (ii) nicht bewusst war.

$q$  ist dann ein „defeater“ für  $p$ , wenn (i), (ii) und (iii) wahrscheinlich wahr/sehr plausibel sind. Auf der Grundlage dieses Modells folgen zwei methodologische Einsichten, die uns im Abschlusskapitel noch einmal beschäftigen werden:

(M1) Modale Dispute in der Philosophie haben folgende typische Form:

- (1.)  $Y$  stellt die These  $T$  auf.
- (2.)  $X$  versteht  $T$  so, dass  $T \rightarrow \Box\neg p$ ; findet  $p$  vorstellbar und behauptet, dass  $p$  möglich ist und  $T$  daher falsch sein muss.
- (3.) Wenn  $Y$  diese Argumentation zurückweisen will, kann  $Y$  entweder versuchen nachzuweisen, dass  $T$  durch  $X$  missverstanden wurde und ‚ $T \rightarrow \Box\neg p$ ‘ falsch ist, oder einen Defeater  $q$  präsentieren, der  $X$ s Vorstellbarkeitsfehler *erklärt*.
- (4.) Wenn  $X$  diese Erklärung nicht akzeptieren will, muss  $X$  entweder  $q$  in Zweifel ziehen oder  $Y$ s Behauptung, dass  $q$  die Vorstellbarkeit von  $p$  in Frage stellt.

(M2) In einem solchen Dialog gilt die vermutliche Unmöglichkeit von  $p$  nicht als potentieller Defeater.

Natürlich müssen wir an (M2) festhalten, wenn wir einen Streit nicht in einer Sackgasse enden lassen wollen. Ansonsten könnte ein Möglichkeitsurteil niemals zur Kritik an einer Notwendigkeitsbehauptung herangezogen werden. Die Tatsache, dass es eine These *gibt*, die mit der Möglichkeitsbehauptung inkompatibel ist, wäre sonst ein Grund in der umstrittenen Notwendigkeitsbehauptung einen Defeater zu sehen. Ein Defeater muss also ein unabhängiger Grund sein, der die Intuition, dass  $p$  vorstellbar ist, in Frage stellt. Yablo gibt auch eine erkenntnistheoretische Begründung:

Part of my point here is just that ignorance of the fact that  $p$  is impossible does not itself do much to explain why I would conceive it as possible. But that is not all.

Even if a fuller explanation is provided, it carries little dialectical force if it depends on the prior concession that my intuition has a significant chance of being false. (With an equal plausibility one could explain away my perceptual impression of ducks by saying that they were produced by decoy ducks, these being the usual explanation of erroneous duck-impressions.) Only if there is independent reason to suspect that my refusal of some relevant proposition really does put me out of touch with the facts, does that refusal call my intuition into question. (Yablo [359], 36)

Obwohl Yablos Analyse überzeugende methodologische *Konsequenzen* hat (auf die wir noch zurückkommen werden), kann die Analyse als solche vermutlich nur diejenigen überzeugen, die Vorstellbarkeitstests schon vorher für zuverlässig gehalten haben.

#### 8.2.5 MÖGLICHE KRITIK AN EINER SOLCHEN CHARAKTERISIERUNG POSITIVER VORSTELLBARKEIT

Problematisch an Yablos Analyse sind mindestens die folgenden Punkte:

(1.) Wenn man nicht versteht, was mit ‚ausmalen‘ in den obigen Definitionen gemeint ist, kann einem durch Yablos Analyse nicht geholfen werden. Einige Philosophen werden sicherlich leugnen, dass sie verstehen, was mit einer objektualen nicht-perzeptuellen Vorstellung gemeint ist (vgl. Cohnitz [72]).

(2.) Dieser Ansatz bleibt psychologistisch. D.h., es wird ein bestimmtes psychisches Phänomen phänomenologisch identifiziert und dann argumentiert, dass dieses Phänomen in gewisser Beziehung zu bestimmten wahren Aussagen steht (es indiziert diese auf nicht-unzuverlässige Weise). Dies erschwert mögliche Erklärungen, *warum* dem so sein sollte – schließlich gibt es keine offensichtliche Verbindung zwischen dem Bereich der modalen Tatsachen und unserer Psyche (jedenfalls nicht in der Standardauffassung modaler Tatsachen).

(3.) Putnam scheint die Auffassung zu vertreten, dass wir bei a posteriori Unmöglichkeiten nicht einmal zu wissen brauchen, dass sie einen notwendigen Wahrheitswert haben. Wenn dies bei allen a posteriori Unmöglichkeiten immer so wäre, könnte folgendes der Fall sein:

(PV\*) Fast jede a posteriori Unmöglichkeit habe ich für vorstellbar (im philosophischen Sinne) und deshalb für möglich gehalten, bevor ich erfahren habe, dass es sich um eine Unmöglichkeit handelt.

Yablos Analyse scheint davon auszugehen, dass a posteriori Unmöglichkeiten fast immer mindestens unentscheidbar sind. Die Tatsache, dass die „Entdeckung“ von metaphysischer Notwendigkeit durch Kripke soviel Aufsehen erregt hat, spricht

wohl dagegen.<sup>349</sup> Damit die Redeweise von *a posteriori* Notwendigkeiten Sinn macht, sollten sie unentscheidbar sein. Wären sie es nicht, wäre es ja *a priori*, dass sie notwendig wahr sind. Wenn sie aber unentscheidbar sind, ist es *a priori*, wenn es sich bei einer Aussage *nicht* um eine *a posteriori* Notwendigkeit handelt.

Bei Putnam haben wir eine Position kennen gelernt, die leugnen würde, dass wir *a priori* wissen können, dass es sich bei einer Aussage nicht um eine *a posteriori* Notwendigkeit handelt. Putnams Beispiel sind Bleistifte, die sich als Organismen entpuppen können:

Verweilen wir kurz bei Namen für Artefakte, bei Wörtern wie „Bleistift“, „Stuhl“, „Teller“ etc. Nach der herkömmlichen Meinung sind diese Wörter ganz gewiß durch Konjunktion oder möglicherweise auch Bündel von Eigenschaften definiert; was alle in der Konjunktion aufgeführten Eigenschaften [...] hat, muß dann ein Bleistift, Stuhl, Teller, etc. sein. Außerdem sind, so die allgemeine Meinung, [...] einige Eigenschaften aus dem Bündel *notwendig* [...]. So wird angenommen, daß die Eigenschaft, ein Artefakt zu sein, notwendig sei, ebenso die, einem gewissen Standardzweck zu dienen; z.B. werden „Bleistifte sind Artefakte“ und „Bleistifte sind normalerweise zum Schreiben da“ für notwendig gehalten. Schließlich besteht die Meinung, daß es sich bei dieser Sorte von Notwendigkeit um *epistemische* Notwendigkeit und sogar um Analytizität drehe. (Putnam [263], 57)

Nach der Auffassung, die Putnam hier skizziert, gehen viele („die herkömmliche Meinung“) offenbar davon aus, dass es sich bei den aufgelisteten Wahrheiten um Notwendigkeiten handelt, die man im Fall von Bleistiften *a priori* wissen kann. Bleistifte sind keine natürlichen Arten – so glauben wir *a priori* – und so ist es auch (in Yablos Sinne) unproblematisch, dass wir ihre modalen Eigenschaften nicht für unentscheidbar, sondern für *a priori* bestimmbar halten. Was aber, wenn sich diese *a priori* Überzeugung als *a posteriori* falsch erweist?

Bemühen wir noch einmal die Science-Fiction, diesmal mit einem Beispiel, das sich Roger Albritton ausgedacht hat. Man imaginiere, wir entdecken eines Tages, daß *Bleistifte Lebewesen sind*. Wir sezieren sie, betrachten sie unter dem Elektronenmikroskop und erblicken ein nahezu unsichtbares Gespinnst aus Nerven und anderen Organen. Wir spionieren ihnen bis zu ihren Laichplätzen nach und sehen, wie sie Eier legen und wie diese zu ausgewachsenen Bleistiften heranreifen. Wir müssen feststellen, daß diese Lebewesen nicht andere (künstlich hergestellte) Bleistifte imitieren; außer diesen Lebewesen gab und gibt es keine Bleistifte. Gewiß, es mutet absonderlich an, daß viele dieser Lebewesen *Aufschriften* tragen – z.B. A. W. FABER-CASTELL 9000 HB –, aber wer weiß, womöglich ist das Camouflage, womöglich sind's Lebewesen. (Es gilt auch zu klären, wieso niemand es je unternehmen hat, Bleistifte zu fabrizieren, und so weiter; doch in einem gewissen Sinn ist das eindeutig eine mögliche Welt.)

Wenn das denkbar ist – und das ist es, darin schließe ich mich Albritton an –, dann ist es epistemisch möglich, daß sich *Bleistifte als Lebewesen entpuppen*. Daraus

---

<sup>349</sup> Diese Vorwürfe treffen auf dieselbe Weise David Chalmers Konzeption positiver Vorstellbarkeit, vgl. Chalmers [58]. Chalmers Konzeption negativer Vorstellbarkeit wird uns im Folgenden beschäftigen.

folgt, daß die Aussage „Bleistifte sind Artefakte“ nicht epistemisch notwendig im striktesten Sinne und a fortiori nicht analytisch ist. (Putnam [263], 57-58)

Folgt man Putnams Auffassung (und sieht man davon ab, wie hier mit einem Vorstellbarkeitsargument eigentlich die Methode des Vorstellbarkeitstests unterminiert wird), scheint Putnam zu vertreten, dass es a posteriori sein kann, welche Ausdrücke natürliche Arten bezeichnen (und dann als rigide Designatoren in a posteriori Notwendigkeiten fungieren können). Wenn wir nun sehr häufig oder fast immer solche epistemischen Möglichkeiten fälschlicherweise ausschließen, werden wir entsprechend in vielen Fällen, in denen wir zu dem Urteil kommen sollten, dass der modale Status eigentlich unentscheidbar ist, urteilen, dass es in der einen oder anderen Weise entschieden werden kann und somit möglicherweise ein modales Fehlurteil produzieren. Ob dies ein guter Einwand ist, hängt davon ab, ob man die Auffassung teilen möchte, dass wir (zumindest häufig) nicht a priori wissen können, ob ein Ausdruck, den wir verwenden, nicht unbekannterweise rigide eine natürliche Art bezeichnet.

### 8.3 KONSISTENTISMUS UND „NEGATIVE“ VORSTELLBARKEIT<sup>350</sup>

Angesichts dieser Probleme mit Yablos „Analyse“ könnte man sich fragen, ob es nicht möglich ist, ‚vorstellbar‘ in einem nicht-psychologistischen Sinne zu explizieren, der nicht davon abhängt, ob man in der Introspektion ein bestimmtes psychisches Phänomen wieder erkennen kann, noch davon abhängt, dass auf wunderbare Weise seine Zuverlässigkeit durch die spektakuläre Entdeckung eines inneren Auges für die modale Struktur der Wirklichkeit fundiert werden wird (oder davon, dass sich modale Urteile letztlich doch als nicht-kognitiv herausstellen).

Ein solcher alternativer Ansatz, den wir als ‚Konsistentismus‘ bezeichnen werden, ist schon häufig vorgeschlagen worden, wurde aber genauso oft zurückgewiesen. Nach dem Konsistentismus ist dasjenige „vorstellbar“ (in Chalmers [58] „negativ vorstellbar“), was keinen nachweisbaren Widerspruch enthält. Vorstellbarkeit betrifft dann eine Frage logischer Analyse, keine Frage von Intuition. Das wäre doch sicherlich wünschenswert.

Claims concerning what is or is not possible abound in contemporary philosophy. The epistemology of such claims, however, remains largely unexplored. Anything imaginable is possible, we are told, with the proviso that imagination be governed by logic. Many who defend this methodology argue that logic frees us from recourse to some mysterious a priori faculty of intuition. Anything is possible so long as it does not contain a contradiction—and we don't need intuition to tell us what is contradictory, just logic. Logic is thus supposed to provide us with a non-mysterious, intuition-free, modal epistemology. (Tidman [327])

---

<sup>350</sup> Der folgende Teil enthält einige der Ideen, die ich in Cohnitz [69] und Cohnitz [72] ausführlicher dargestellt habe.

Das ist natürlich zu schön, um wahr zu sein. Wie wir sehen werden, ist es vermutlich zu schön, um logisch möglich zu sein, wenn man eine klassische Auffassung von Logik vertritt. Bevor wir aber die Hoffnung aufgeben, sollten wir zuerst die Konzeption gegen einige zu vorschnelle Widerlegungsversuche in Schutz nehmen.

In 8.3.1-8.3.4 werden wir zunächst einige vorschnelle Einwände gegen die Möglichkeit einer rein logischen Rekonstruktion unseres Wissens von Möglichkeiten zurückweisen. In 8.3.5 werden wir erläutern, wie die These von der rationalen Rekonstruierbarkeit unseres Wissens von Möglichkeiten auf der Grundlage der Logik genau zu verstehen ist. In 8.3.6.1-8.3.6.4 werden wir dann diese These mit formalen Mitteln ausdrücken und beweisen, dass sie für die Standardlogik falsch ist. Der Konsistentismus kann unser Wissen von Möglichkeiten weder erklären, noch rechtfertigen.

In 8.3.6.5 werden wir die Möglichkeit diskutieren, wie ein modifizierter Konsistentismus aussehen müsste. Wie wir sehen werden, kann eine rationale Rekonstruktion mit den Mitteln nichtklassischer Logiken erreicht werden. Ob eine solche Rekonstruktion aber die gewünschte Erklärungs- und Rechtfertigungsleistung erbringen kann, ist damit noch nicht beantwortet.

### 8.3.1 FORMALE LOGIK IST UNZUREICHEND?

Paul Tidman hat darauf hingewiesen, dass, wenn wir den Konsistentismus als die These verstehen, dass dasjenige möglich ist, was nicht mit den Mitteln der Prädikatenlogik erster Stufe als widersprüchlich erwiesen werden kann, *bona fide* Unmöglichkeiten als Möglichkeiten ausgewiesen würden:

[O]n a purely formal conception of consistency, a great many propositions that we clearly want to deny are possible will turn out to be consistent. An example would be the familiar "Joe is bachelor and Joe is married" or "Figure A is both a square and a circle." Viewed purely in terms of consistency in classical logic, such propositions seem clearly consistent. (Tidman [327])

Um ein späteres Problem nicht vorwegzunehmen, wollen wir zunächst einmal die monadische Prädikatenlogik betrachten. Stimmt es, dass in ihr Sätze konsistent zu sein „scheinen“, die eigentlich unmöglich sind? In anderen Worten: gibt es wohlgeformte Formeln der Prädikatenlogik erster Stufe, die etwas Unmögliches bedeuten, die aber von der entsprechenden Beweistheorie<sup>351</sup> als konsistent eingestuft werden?

---

<sup>351</sup> Unter ‚Beweistheorie‘ verstehe ich hier dasselbe wie unter ‚deduktives System‘. Beide Ausdrücke bezeichnen das Regelsystem, das in einer bestimmten Logik die erlaubten Beweisschritte festlegt. Unter ‚Modelltheorie‘ verstehe ich entsprechend dasselbe wie unter ‚Semantik‘. In der Literatur werden die beiden Ausdrücke manchmal auch zur Bezeichnung jeweiliger Metatheorien verwendet.

Relativ zur Semantik der monadischen Prädikatenlogik erster Stufe ist das natürlich *falsch*. Wie man mit Theorem 8.3-4, das wir später noch diskutieren werden, zeigen kann<sup>352</sup>, wäre eine monadische Prädikatenlogik, die allein mit syntaktischen Mitteln die Konsistenz ihrer Sätze beurteilt, *korrekt*. D.h., wenn mit logischen Mitteln ein Satz als konsistent beurteilt wird, drückt er garantiert *keine* Unmöglichkeit aus.

Die Prädikatenlogik ist also streng genommen nur dann unzuverlässig in Tidmans Sinn, wenn man ihr eine *andere* Semantik unterschiebt als die eigentliche modelltheoretische Semantik, indem man eine syntaktische Identifikation von Prädikatbuchstaben, Konstanten und Namen der Logik mit Ausdrücken der natürlichen Sprache vornimmt und entsprechend den Ausdrücken der Prädikatenlogik dann die Semantik der natürlichen Sprache zuweist. Kein Wunder, dass man damit gegen die Wand fährt. ‚ $p \wedge q$ ‘ ist aussagenlogisch auch dann konsistent, wenn es die Formalisierung von ‚Alle Menschen sind sterblich und es gibt einen Menschen, der unsterblich ist.‘ darstellt, was *prädikatenlogisch* modelliert inkonsistent ist.<sup>353</sup> Mit denselben Überlegungen könnte man sonst argumentieren, dass die Aussagenlogik *unvollständig* sein muss, weil in ihr ‚Der Urmeter in Paris ist 1m lang.‘ nicht abgeleitet werden kann, wenn man diesen Satz mit ‚p‘ ‚übersetzt“. In anderen Worten: unsere Logik ist ein limitiertes Modell einer Sprache. Wenn wir uns für Inkonsistenzen in der natürlichen Sprache interessieren, sollten wir ein adäquates Modell der natürlichen Sprache nehmen, kein inadäquates. Die Prädikatenlogik ist inadäquat als Modell für die natürliche Sprache, da sie – unter anderem – keine Bedeutungspostulate enthält, die eine Zuordnung von Prädikaten von Umgang- und formaler Sprache überhaupt erst möglich machen.

### 8.3.2 BEDEUTUNGSPOSTULATE?

Nun ist Tidman außerdem der Überzeugung, dass auch dann, wenn Bedeutungspostulate der Sprache „hinzugefügt“ worden sind, immer noch Unmöglichkeiten durch den Konsistenztest als Möglichkeiten beurteilt werden müssten:

Of course, counterexamples remain for even this expanded account. Those more inclined to rationalism propose that we have synthetic a priori knowledge of some necessary truths. Many argue, for example, that appeal to mere definitional truths cannot satisfactorily account for our knowledge that nothing can be red and green all over at the same time. It does not seem plausible, in particular, to suggest that it is the linguistic meaning of color words like “red” and “green” which provides us with the information that something which is red all over cannot possibly be green

<sup>352</sup> Theorem 8.3-4 bezieht sich auf eine Axiomatisierung einer modalen Aussagenlogik, die eine Beweistheorie enthält, die Konsistenzbeurteilungen mit logischen Mitteln zulässt. Eine parallele Axiomatisierung könnte man auch für die monadische Prädikatenlogik durchführen.

<sup>353</sup> Wen dieses Beispiel nicht überzeugt, der konstruiere sich eine Inkonsistenz der Prädikatenlogik erster Stufe, die in monadischer Prädikatenlogik nicht ausdrückbar ist.

all over at the same time. Other widely discussed examples include the principle of sufficient reason and the claim that God exists necessarily. [...] Recent developments in the philosophy of language have provided even those with empiricist leanings reason to question the account at hand. [...] Due to the widespread acceptance of such claims [wie ‚Wasser ist H<sub>2</sub>O‘] as being both a posteriori and necessary, the view that necessity is merely a function of logic and linguistic meaning clearly no longer holds the position of dominance in contemporary philosophy it once enjoyed. (Tidman [327], 393)

Über die von Tidman präsentierten synthetischen „Wahrheiten“ a priori braucht man wohl nicht viel sagen. Die eine ist analytisch, die anderen beiden sind falsch. Ob metaphysische Notwendigkeiten durch die Bedeutungspostulate erfasst sind, wird davon abhängen, ob man der Meinung ist, dass es Teil der Bedeutung von Wasser ist, dass es notwendig H<sub>2</sub>O bezeichnet. Ob man solche Notwendigkeiten der Menge der Bedeutungspostulate hinzufügt, ist schließlich eine pragmatische Frage, die davon abhängt, welche Art von Notwendigkeit bzw. Möglichkeit man im Blick hat. Wichtig ist im Grunde doch nur, dass es eine Frage der Logik sein soll, ob etwas als vorstellbar und deshalb als möglich angesehen wird. Wenn ich wissen möchte, was metaphysisch möglich ist und der Meinung bin, dass es weniger metaphysische Möglichkeiten gibt als es begriffliche Möglichkeiten gibt (man denke hier wieder an die Zwiebel der Notwendigkeiten bzw. der Möglichkeiten), dann sollte ich ein System betrachten, das neben den Bedeutungspostulaten eben *auch* metaphysische Notwendigkeiten zu seinen Axiomen zählt. Ansonsten ist das betrachtete System eben für den anvisierten Zweck inadäquat, was aber nichts an der konsistentistischen Idee ändert. Diese besteht ja schlicht darin, dass es – gegeben einen bestimmten Bestand an Wahrheiten – eine Frage der Logik ist, ob etwas möglich ist oder nicht.

### 8.3.3 WELCHE LOGIK?

Schließlich bemängelt Tidman, dass der Konsistentismus deswegen unbestimmt bliebe, weil es schließlich mehr als eine Logik gebe und zwischen diesen nur auf der Grundlage unserer Intuitionen gewählt werden könne. Wenn unsere Intuitionen aber bestimmen, was für eine Logik gelten soll, dann dreht der Konsistentismus die epistemische Reihenfolge doch einfach um, wenn er das Problem, welche Intuitionen zuverlässig sind, bzw. welchen Möglichkeitsurteilen wir trauen sollen, umgekehrt auf die Logik abwälzen will. Dieses Problem haben wir in Kapitel 3.1 schon kennen gelernt.

Diese Kritik übersieht erstens, dass wir adäquate Logiken durchaus durch das Verfahren eines angenäherten reflektierten Gleichgewichts festlegen können. Dass zweitens ein solches Verfahren nicht auf vitiöse Weise zirkulär ist, liegt daran, dass wir in der Regel ja nicht die fragliche Proposition als Postulat der Logik hinzufügen oder wegnehmen, wann immer es uns beliebt, sondern generelle Prinzipien betrachten. Dass ‚Der Morgenstern wird den Abendstern überdauern.‘ als inkon-

sistent erwiesen werden kann, mag daran liegen, dass ‚Wenn  $x = y$ , dann  $\Box(x = y)$ ‘ und ‚Wenn  $x \neq y$  überdauert, dann  $x \neq y$ .‘ als Bedeutungspostulate dem System genauso angehören wie die Angabe der Referenzgleichheit von ‚Abendstern‘ und ‚Morgenstern‘.

### 8.3.4 FALLIBILITÄT?

Ein interessanterer Einwand gegen die konsistentistische Auffassung, dass Vorstellbarkeit nur bedeutet, dass sich kein Widerspruch finden lässt, ist von Yablo und Van Inwagen vorgetragen worden:

Do I acquire evidence in favor of a proposition's possibility, by finding myself without evidence against its truth? That would be very strange, to say the least. Among other things it would have the result that there is a necessary limit on how bad my epistemological position can get: the poorer my evidence for  $p$ 's truth, the better my evidence for its possibility. (Yablo [359], 8)

What I dispute is the contention that if a concept or state of affairs is not logically impossible, then it is "logically possible." It hardly follows that, because a certain thing cannot be proved to be impossible by a certain method, it is therefore possible in any sense of 'possible' whatever. (Van Inwagen [333], 71)

Yablos Gründe für seine Auffassung haben wir weiter oben bereits ausgeführt. Betrachten wir, weshalb Van Inwagen glaubt, dass ein konsistentistischer Ansatz eine Fehlrekonstruktion der modalen Epistemologie sein soll:

Suppose that the infallible Standard Atlas marks many islands as uninhabitable, none as inhabitable, and makes no claim to completeness in that matter. We could, if we liked, say that the islands marked 'uninhabitable' in the Standard Atlas were "cartographically uninhabitable." In doing this, we should be calling attention to the fact that our knowledge that these islands were uninhabitable had a certain source. But would there then be any sense in saying that an island was "cartographically inhabitable" just in the case it was not cartographically uninhabitable? Very little; I should think. (Van Inwagen [333], 71-72)

Van Inwagen scheint hier zu argumentieren, dass wir (in diesem Beispiel) nicht *deshalb*, weil wir nur wissen, dass einige Inseln unbewohnbar sind, gerechtfertigt sind zu glauben, alle anderen seien bewohnbar. Aus dieser Beobachtung folgert Van Inwagen zunächst einen modalen Skeptizismus („[we] do not and cannot know (at least by the exercise of our own unaided powers) modal propositions like [crucial premises in modal philosophical reasoning]“ Van Inwagen [333], 69). Darüber hinaus zieht Van Inwagen aber auch ontologische Konsequenzen bezüglich der Existenz solcher Möglichkeiten („[...] there is no such thing as logical possibility“ Van Inwagen [333], 71). Mir scheint Van Inwagen in seinem Argument „Rechtfertigung“ so eingeschränkt zu verwenden, dass nurmehr Überzeugungen,

die absolut gewiss sind, rechtfertigbar sein können, Überzeugungen, die potentiell falsch sind, aber grundsätzlich nicht.<sup>354</sup> Das ist alles andere als ein gutes Argument gegen einen konsistentistischen Ansatz. Schließlich sind alle unsere Überzeugungen potentiell falsch, nicht nur die, die wir in Bezug auf Möglichkeiten haben.

Vom Standpunkt eines konsistentistischen Ansatzes aus gibt es (genau wie im Fall von *actualia*) auch eine Möglichkeit zur Rechtfertigung von Überzeugungen aus der Abwesenheit widersprechender Evidenz, vorausgesetzt, dass es (a) systematisch möglich ist, zu bestimmen was die Klasse der relevanten Evidenz auszeichnet (beispielsweise logische Notwendigkeiten bei der Beurteilung von logischen Möglichkeiten), und (b) man tatsächlich (möglichst systematisch) untersucht hat, ob solche widersprechende Evidenz existiert.

Kann man (a) und (b) plausibel machen, liegt zumindest kein epistemologischer Salto Mortale mehr vor, zumindest keiner, der nicht auch einem Falsifikationisten vorgeworfen werden könnte. Hier eröffnet sich also durchaus eine Möglichkeit, Vorstellbarkeit durch die Nicht-Feststellbarkeit von Inkonsistenzen zu explizieren. Man kann also durchaus zugeben, dass der konsistentistische Ansatz zeitweilig zu viele Möglichkeiten produzieren kann (wenn wir noch nicht alle Notwendigkeiten kennen, halten wir aufgrund des konsistentistischen Vorstellbarkeitstests zunächst zu vieles für möglich). Es ist aber keinesfalls klar, warum weitere Untersuchungen die bisher noch unbekanntes Notwendigkeiten nicht ans Tageslicht holen sollten, die die „Überproduktion“ von Möglichkeiten wieder eindämmt. Dann würde eben auch in der modalen Epistemologie Wissensfortschritt auf „herkömmliche“ Weise ablaufen.

Dabei ist es keinesfalls nötig anzunehmen, dass wir zu irgendeinem gegebenen Zeitpunkt wissen bzw. wissen *müssen*, dass wir jetzt über die vollständige Kenntnis aller relevanten Notwendigkeiten verfügen. Es reicht für die Rechtfertigung unserer Schlüsse, zu wissen, dass wir ernsthaft nach relevanten Notwendigkeiten geforscht haben. Eine solche Verteidigungsstrategie schwebt auch Bob Hale vor, der (wie wir gleich noch sehen werden) ein ganz analoges Problem lösen muss:

[If a consistentist] is to answer the charge that her account of the basis of our beliefs about possibility simply confuses having grounds to believe a proposition with lacking grounds to disbelieve it, she needs to give decent sense to the idea of a well-conducted search for countervailing necessities. (Hale [139], 15)

Diese Argumentation erscheint auf den ersten Blick plausibel. Wie wir gleich sehen werden, ist sie auf den zweiten Blick leider nicht ohne weiteres zu verteidigen.

---

<sup>354</sup> Auch in Bezug auf a priori-Wissen gibt es durchaus noch fallibilistische Strategien von „Wissen“ reden zu können, ohne Gewissheit bemühen zu müssen. Vgl. Edidin [98].

### 8.3.5 VERSCHIEDENE EBENEN DER BETRACHTUNG

Wir hatten in Teil 4 im Zusammenhang mit Wilkes Kritik an der Methode des Gedankenexperiments bereits bemerkt, dass es eigentlich die Dialektik modaler Dispute auf den Kopf stellt, wenn man davon ausgeht, dass Möglichkeiten durch die Abwesenheit ausschließender Notwendigkeiten *explizit* gerechtfertigt werden. Yablo hat außerdem insofern völlig Recht, dass unsere Möglichkeitsurteile phänomenologisch gesehen *nicht* dadurch getroffen werden, dass wir eine Reihe potentiell relevanter Notwendigkeiten durchgehen und formal feststellen, dass sie nicht in Widerspruch zur fraglichen Möglichkeitsaussage stehen. Wäre der Konsistentismus ein Modell der expliziten Rechtfertigung von Möglichkeitsaussagen, hätte er insofern einen schweren Stand, da er unserer tatsächlichen Praxis des modalen Argumentierens nicht zu entsprechen scheint. Wie wir bei Wilkes gesehen haben, könnte er nicht erklären, warum wir Möglichkeiten als Gegenbeispiele gegen Notwendigkeiten einsetzen können. Wenn Notwendigkeiten eigentlich das epistemisch primäre sind, beurteilen wir eben das, was möglich ist, auf der Ebene dessen, was wir für notwendig halten, nicht aber umgekehrt. Der Konsistentismus scheint außerdem nicht damit übereinzustimmen, was wir für die richtige Phänomenologie unseres modalen Urteilens halten (wenn wir modale Urteile auf der Grundlage von Vorstellbarkeitstests fällen, wie wir es in 8.2 erläutert haben).

Man kann den Konsistentismus aber auch anders interpretieren und diese Probleme umgehen: Man könnte zunächst einmal zugeben, dass Yablo auf einer *phänomenologischen Ebene* einen Sinn von ‚vorstellbar‘ identifiziert hat, der in zuverlässiger Weise Auskunft über Möglichkeit gibt. Diese phänomenologische Betrachtung *erklärt* noch nicht, *warum* ein solcher Vorstellbarkeitstest zuverlässig ist. Dies geschieht auf der *Ebene der rationalen Rekonstruktion*. Auf dieser Ebene wird erklärt, wie durch logische Schlussfolgerungsprozesse Wissen von Möglichkeiten aus Wissen von Notwendigkeiten gewonnen wird, wobei sowohl das relevante Notwendigkeitswissen implizites Wissen sein kann, wie auch der Schlussfolgerungsprozess nicht bewusst sein muss. Die Ebene der rationalen Rekonstruktion *erklärt* und *rechtfertigt* die Praxis, die wir auf der phänomenologischen Ebene identifiziert haben. Sie erklärt sie, indem sie uns darüber Auskunft gibt, warum bestimmte phänomenologische Aspekte des Vorstellbarkeitstests vorhanden sind: so wäre auf der Ebene der rationalen Rekonstruktion bei einem Möglichkeitsurteil festzustellen, ob das fragliche Urteil mit dem impliziten Wissen von Notwendigkeiten kompatibel ist. Kompatibilität wird dadurch festgestellt, dass man keinen Widerspruch in der Aussagenmenge finden kann, die das fragliche Urteil wie das gesamte relevante Hintergrundwissen enthält. Auf der phänomenologischen Ebene besteht die Angabe eines solchen Modells in der Imaginierung einer Welt, die man zur Verifikation des fraglichen Urteils akzeptieren würde. Die Erklärungsleistung ist also, dass die von Yablo festgestellte Zuverlässigkeit auf die Zuverlässigkeit des (mutmaßlich logischen) Schlusses auf der Ebene der rationalen Rekonstruk-

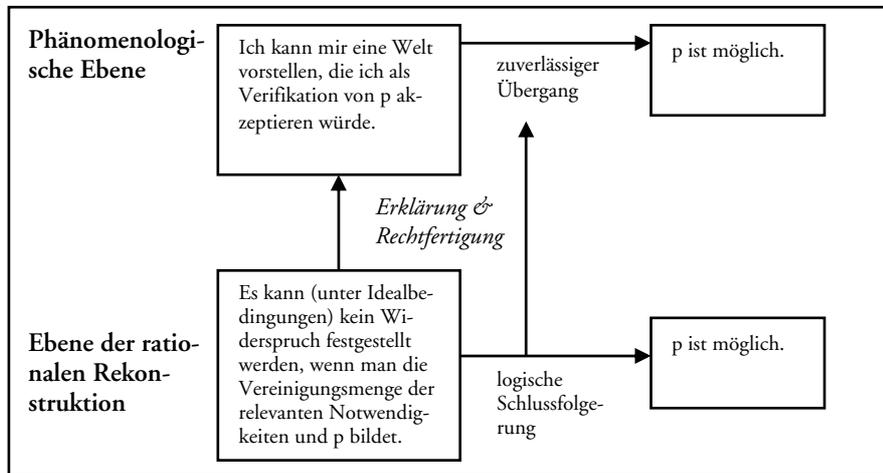


Abbildung 8.3-1

tion zurückgeführt werden kann. Der Übergang von positiver Vorstellbarkeit auf die Möglichkeit von  $p$  ist gerechtfertigt, *weil* es sich auf der Ebene der rationalen Rekonstruktion (unter Idealbedingungen) um einen logischen Schluss handelt.

Eine ähnliche Erklärung kann man dann auch für Yablos Modell modalen Irrtums angeben. Defeater weisen auf Notwendigkeiten hin, die im impliziten Schlussfolgerungsprozess übersehen worden sind (vielleicht weil sie nicht bekannt waren). Wenn diese aufgedeckt werden, wird das Möglichkeitsurteil revidiert, *weil* nun ein Widerspruch nachgewiesen werden kann, der zwischen der neuen Information und dem fraglichen modalen Urteil besteht. In diesem Fall wären „positive“ und „negative“ Vorstellbarkeit nicht eigentlich zwei Arten von Vorstellbarkeit. Es ginge *in jedem Fall* darum, festzustellen, ob eine bestimmte Menge von Aussagen einen Widerspruch enthält, der Unterschied zwischen den beiden „Vorstellungsarten“ bestünde nur darin, ob alle der relevanten Notwendigkeiten explizit bekannt wären oder nicht.<sup>355</sup> Ob man diese Rekonstruktion aufrechterhalten kann, hängt davon ab, ob auf der Ebene der rationalen Rekonstruktion tatsächlich davon ausgegangen werden kann, dass es sich beim Schluss auf die Möglichkeit unter Idealbedingungen um eine logische Schlussfolgerung handelt (vgl. Abbildung 8.3-1).

Akzeptiert man diese Betrachtungsweise, würde sich auch ein Problem Tidmans erledigen. Tidman hatte am Konsistentismus bemängelt, dass dieser für unser tatsächliches modales Urteilen eine bestimmte Logik festlegen müsse, die die logische Struktur unserer Überzeugungen exakt wiedergibt. Geht man aber davon aus, dass der Konsistentismus nicht hauptsächlich eine Theorie darüber darstellt, wie wir *explizit* rasonieren, sondern nur (oder hauptsächlich) eine rationale Rekon-

<sup>355</sup> Vgl. Chalmers [58] für die hier verwendete Terminologie.

struktion impliziter Schlussfolgerungsprozesse sein will, muss er keine bestimmte Logik angeben. Es reichte völlig hin, könnte man zeigen, dass es *prinzipiell möglich* ist, eine rationale Rekonstruktion unseres modalen Schließens anzugeben. Dafür wäre jede Logik geeignet, die ausdrucksstark genug ist, um der Struktur unseres modalen Rasonierens gerecht zu werden. Wir müssen in Einzelfällen ja *deswegen* keine explizite logische Rekonstruktion betreiben, weil wir uns – auf der phänomenologischen Ebene – schon unabhängig der Tatsache versichert haben, dass der Vorstellungstest alleine hinreichend zuverlässig für unsere Zwecke ist. In anderen Worten: *intern* verwenden wir ohnehin die „richtige“ Logik, wenn (a) Yablo Recht hat und (b) eine rationale Rekonstruktion in diesem Sinne prinzipiell möglich ist.

### 8.3.6 WARUM DER KONSISTENTISMUS NICHT FUNKTIONIEREN KANN<sup>356</sup>

Die Rede von „Idealbedingungen“ ist deswegen vorhin in unsere Formulierung geflossen, weil modales Urteilen natürlich fehlbar sein kann. Es kann sein, dass wir bestimmte relevante Notwendigkeiten nicht kennen, es kann sein, dass wir beim Schlussfolgern Fehler machen. Für die Zwecke der rationalen Rekonstruktion kann man von diesen Dingen absehen. Was die rationale Rekonstruktion im Konsistentismus leisten muss, ist lediglich für bestimmte Idealbedingungen nachzuweisen, dass es sich beim Schluss auf die Möglichkeit einer Proposition aus der Tatsache, dass kein Widerspruch zwischen dieser Proposition und der Menge der relevanten Notwendigkeiten gefunden werden kann, um einen logischen Schluss handelt. Wie wir gleich sehen werden, ist es unmöglich, einen solchen Nachweis zu erbringen, wenn man bestimmte Minimalannahmen über die logische Struktur unseres modalen Rasonierens macht und sich auf den Standpunkt der klassischen Logik stellt. Es scheint, als sei dies nicht nur für unser Zwei-Ebenen-Modell fatal, sondern auch für die Auffassung, dass sich unser modales Wissen insgesamt auf Wissen von Notwendigkeiten basieren lässt (notwendigkeitsbasierter Asymmetrieansatz modalen Wissens), die Auffassung, dass es sich bei idealer negativer Vorstellbarkeit um einen zuverlässigen Indikator für Möglichkeit handelt (Modaler Rationalismus) und schließlich die Auffassung, dass sich eine Reaktionsabhängigkeits-Analyse für die Modalitäten angeben lässt. Gehen wir diese Dinge nach und nach an, beginnend mit Chalmers' modalem Rationalismus.

#### 8.3.6.1 Eine weitere Enthaltenseins-These

Wie bereits erwähnt, diskutiert David Chalmers ähnlich wie Yablo verschiedene Bedeutungen von ‚vorstellbar‘ in der Absicht, eine Bedeutung aufzuweisen, die einen zuverlässigen Indikator für Möglichkeit abzugeben vermag. Neben „positiver Vorstellbarkeit“, die ähnlich der Yabloschen Analyse konzipiert ist, identifiziert Chalmers auch eine Form der „negativen Vorstellbarkeit“:

---

<sup>356</sup> Vgl. Cohnitz [72].

The central sort of negative conceivability holds that  $S$  is negatively conceivable when  $S$  is not ruled out a priori, or when there is no (apparent) contradiction in  $S$ . [...] And we can say that  $S$  is ideally negatively conceivable when it is not a priori that  $\neg S$ . (Chalmers [58], 149)

„Ideal“ bedeutet bei Chalmers ‚cannot be trumped by better reasoning‘ in dem Sinne, dass ein besserer Denker die Rechtfertigung,  $\neg S$  nicht a priori zu finden, rational zurückweisen könnte. Zunächst wollen wir diese (problematische<sup>357</sup>) Interpretation von ‚ideal‘ zurückstellen und stattdessen von einem ‚idealen Denker‘ reden, unter der Annahme, dass ein solcher Begriff kohärent explizierbar ist. „Ideale Denker“ sind einfach universale Turing-Maschinen mit unendlicher Speicherkapazität, die sowohl mit den jeweils relevanten Informationen als auch den relevanten Ableitungsregeln ausgestattet sind. ‚Es ist a priori, dass  $\neg S$ ‘ bedeutet, dass die Turing-Maschine einen Beweis für  $\neg S$  produzieren kann. Dies klingt zunächst durchaus kohärent.

Chalmers ist nun der Auffassung, dass das folgende sehr wahrscheinlich wahr ist:

(NC) Ideale negative Vorstellbarkeit enthält Möglichkeit.

Im Grunde argumentiert Chalmers für die Wahrheit von (NC), indem er eine Reihe möglicher Gegenbeispiele und so genannte ‚Inscrutabilities‘ durchmustert. Bei Letzteren handelt es sich um notwendige Wahrheiten, die a priori nicht wissbar sind und daher verhindern, dass wir für einige  $S$  a priori herausfinden, dass tatsächlich  $\neg S$  der Fall ist. Wir haben dieses Problem für den modalen Rationalismus bereits in 7.2.1.3.2 besprochen. Wie Chalmers im Einzelnen für (NC) argumentiert, soll uns aber hier auch nicht kümmern.

Betrachten wir zunächst (NC) etwas genauer. (NC) ist offenbar eine These, die eine bestimmte Beziehung zwischen unseren (idealisierten) mentalen Fähigkeiten schlussfolgernd zu denken und dem Bereich der Möglichkeiten behauptet. Nehmen wir einmal an, dass sich unsere Fähigkeiten, schlussfolgernd zu denken, als ein deduktives System repräsentieren lassen, d.h. als eine Menge syntaktischer Regeln, die gewisse Schlüsse bzw. Ableitungen gestatten. Insbesondere erlauben diese Regeln die Ableitung von Möglichkeiten aus der Tatsache ihrer negativen Vorstellbarkeit. Der Bereich der Möglichkeiten kann dann als semantisches Gegenstück dieses deduktiven Systems angesehen werden. Chalmers' These nun, dass negative Vorstellbarkeit Möglichkeit enthält, wäre dann die These, dass das deduktive System korrekt ist, d.h., dass unter den angesprochenen Idealbedingungen gilt, dass, wenn

$\vdash \diamond S$

---

<sup>357</sup> Vgl. die Diskussion dieses Begriffs in Anhang I von Cohnitz [72].

in Übereinstimmung mit den Regeln negativer Vorstellbarkeit (wie auch immer diese aussehen), es auch der Fall ist, dass

$$\models \diamond S$$

(NC) kann also als folgende These wiedergegeben werden:

(NC')  $\forall S(\text{»}\diamond S\text{« wurde unter Idealbedingungen in Übereinstimmung mit den Ableitungsregeln negativer Vorstellbarkeit abgeleitet} \rightarrow \text{»}\diamond S\text{« ist gültig})$ .

(NC') ist aber nicht nur eine zentrale These des modalen Rationalismus, sondern auch eine Annahme, die so genannte „notwendigkeitsbasierte Asymmetrieansätze modalen Wissens“ machen müssen. Diese Betrachtungsweise ist im Grunde eine Erfindung Bob Hales. Nach Hale ist die modale Epistemologie eigentlich mit zwei Fragen beschäftigt:

(Q1) Woher wissen wir für ein gegebenes  $p$ , dass  $\Box p / \neg \diamond \neg p$ ?

(Q2) Woher wissen wir für ein gegebenes  $p$ , dass  $\diamond p / \neg \Box \neg p$ ?

Nach Hale kann die modale Epistemologie fruchtbar so betrieben werden, dass man zunächst einer der beiden Fragen Priorität zuweist und davon ausgeht, dass sich die jeweils andere Frage im Nachhinein als überflüssig herausstellt:

To put the idea roughly and suggestively, we might think of possibility as more revealingly characterised as just absence of necessity, so that knowledge of possibilities is primarily knowledge of the absence of relevant necessities – or oppositely, we may view necessity as just absence of possibility, and knowledge of necessity as primarily knowledge of the absence of any relevant possibility. (Hale [139], 5)

Ein modalepistemologischer Ansatz, der der ersten Frage den Vorzug gäbe, hieße ‚notwendigkeitsbasiert‘, ein Ansatz, der unser modales Wissen auf unserem Möglichkeitswissen basieren lassen will, entsprechend ‚möglichkeitsbasiert‘. Notwendigkeitsbasierte Ansätze betrachten unser Wissen von Notwendigkeiten als „dominant“ und unsere Wissen von Möglichkeiten als „rezessiv“, Umgekehrtes gilt für möglichkeitsbasierte Ansätze. Ein asymmetrischer Ansatz muss nun in jedem Fall nachweisen, dass er nicht auf das jeweils rezessive „Wissen“ aufgrund von Ignoranz urteilt, ansonsten wären asymmetrische Ansätze epistemologisch höchst fragwürdig. Dies lässt sich leicht erläutern, wenn man annimmt, dass es bei Wissen (auch) um *gerechtfertigte* Überzeugungen geht. Wenn unser Wissen von (beispielsweise) Möglichkeiten analysiert würde als Wissen, dass durch die bloße Ignoranz in Bezug auf potentielle widersprechende Notwendigkeiten zustande kommt, wäre die Überzeugung, dass etwas möglich ist, auf die ungerechtfertigte Überzeugung gegründet, dass das Gegenteil nicht notwendig ist. Eine ungerechtfertigte Überzeugung kann aber keine Basis für eine gerechtfertigte Überzeugung sein.

Wie wir oben schon sahen, besteht Hales Strategie zur Lösung dieses Problems darin, aufzuzeigen, dass zumindest für notwendigkeitsbasierte Ansätze gezeigt wer-

den kann, dass es Sinn macht, von einer gerichteten systematischen Suche nach relevanten Notwendigkeiten zu reden. Falls es so etwas gäbe, sollte der Übergang von ‚wir haben keinen Grund zu glauben, dass  $\neg p$ ‘ auf ‚wir haben Grund zu glauben, dass  $p$ ‘ gestattet sein (wohlgemerkt als eine in allen Einzelfällen fehlbare Schlussregel).

Damit eine solche Verteidigungsstrategie von Erfolg sein kann, muss Hale aber doch offenbar annehmen, dass zumindest unter idealen Umständen, der gerade diskutierte Übergang auch *gültig* sein sollte. „Ideale Umstände“ wären dabei solche, in denen wir alle relevanten Notwendigkeiten kennen, wir wissen, dass wir sie kennen, die fraglichen Schlussregeln kennen, und kognitiv nicht limitiert sind. In anderen Worten, auch Hale muss davon ausgehen, dass (NC') wahr ist. Falls dem nicht so wäre, es also keine Umstände gäbe unter denen (NC') wahr wäre, würde es auch keinen Sinn machen, von einer systematischen Suche nach relevanten Notwendigkeiten zu reden. *Systematizität* bedeutet hier doch, dass es so etwas wie Konvergenz geben muss. Wenn es aber nicht einmal Idealbedingungen gibt, unter denen diese „systematische Suche“ ein Ende gefunden hat, hängt die Redeweise von Systematizität völlig in der Luft. Falls selbst in eben beschriebenem Idealfall der Schluss auf Mögliches immer noch fehlgehen könnte, wäre unser Wissen von Möglichkeiten nicht in unserem Wissen von Notwendigkeiten begründet, sondern davon einfach logisch unabhängig. Hale sollte also folgendermaßen argumentieren: Gerade *weil* (NC') wahr ist, ist der Übergang von ‚wir haben keinen Grund zu glauben, dass  $\neg \diamond p$ ‘ auf ‚wir haben Grund zu glauben, dass  $\diamond p$ ‘ auch unter Normalbedingungen gestattet. „Normalbedingungen“ sind dabei diejenigen Bedingungen, in denen wir nicht gewiss sein können, dass wir tatsächlich alle relevanten Notwendigkeiten kennen, sondern uns nur versichert haben, dass wir ordentlich nach den relevanten Notwendigkeiten gesucht haben.

### 8.3.6.2 Was ist Möglichkeit?

(NC') ist folglich ein zentraler Bestandteil von mindestens zwei verschiedenen modalepistemologischen Ansätzen. (NC') ist aber nicht nur von Relevanz für Fragen der Epistemologie, sondern auch für solche der Explikation der modalen Begriffe überhaupt.

Eine Möglichkeit, einen bestimmten Begriff zu explizieren, besteht darin, eine reduktive Analyse anzugeben, die den fraglichen Begriff auf etwas Bekannteres zurückführt. (Eine solche „reduktive“ Analyse der Modalitäten ist beispielsweise von David Lewis versucht worden.) Das so genannte „Blackburn-Dilemma“ stellt allgemein in Frage, ob eine solche Explikation möglich ist.

Die kritische Aussage des Blackburn-Dilemmas besteht kurz gesagt darin, dass es absolut unklar ist, wie wir jemals eine befriedigende reduktive Analyse für irgendeine Form der (sei es natürlichen, moralischen oder logischen) Notwendigkeit erreichen können. Falls die Frage darin besteht, warum A (sei es natürlich, moralisch oder logisch) der Fall sein *muss*, könnten wir natürlich einen Beweis von A aus B geben, der aber nur dann eine Erläuterung darstellt, wenn wir bereits wissen, warum B (sei es natürlich, moralisch oder logisch) der Fall sein *muss*. Uns ging es aber um die Erläuterung von Notwendigkeit generell, nicht um die Notwendig-

keit von A im Speziellen. Erläutern wir die Notwendigkeit von A also aus der Notwendigkeit von irgendetwas anderem, ist es uns nicht gelungen, tatsächlich anzugeben, worin Notwendigkeit besteht. Erläutern wir die Notwendigkeit von A aber aus bloß kontingenten Umständen, ist die Notwendigkeit weniger erläutert oder erklärt als vielmehr unterminiert:

Either the explanandum shares the modal status of the original and leaves us dissatisfied, or it does not, and leaves us equally dissatisfied. (Blackburn [27], 54)

Sehen wir einmal davon ab, dass Blackburn im obigen Zitat ‚explanandum‘ mit ‚explanans‘ verwechselt, da das „original“ ja schließlich dasjenige ist, dessen modaler Status durch die reduktive Analyse *zu erklären* ist, es also nicht unbefriedigend sein kann, dass es den modalen Status mit sich selbst gemeinsam hat, und stellen außerdem die Frage zurück, ob es sich hierbei tatsächlich um ein Dilemma handelt. In jedem Fall war diese Problematik Grund genug, sich nach anderen als reduktiven Lösungen für diese klassischen philosophischen Probleme umzusehen.

#### 8.3.6.2.1 Response-Dependence

Im Gegensatz zu „reduktiven“ Analysen (wie beispielsweise dem bereits angesprochenen modalen Realismus von David Lewis<sup>358</sup>, der modale Begriffe auf Aussagen über von unserer Welt kausal isolierte mögliche Welten zurückführt) wird bei reaktionsabhängigen Ansätzen keine Reduktion auf nicht-modale Begriffe angestrebt. Modale Begriffe und Aussagen werden vielmehr als primitiv und nicht weiter analysierbar aufgefasst, und als ihre Wahrmacher werden schlicht modale Tatsachen angenommen. Die Analyseleistung dieser Ansätze besteht dann darin, eine Erklärung zu geben, wie modale Aussagen und Begriffe von uns verwendet werden und welche notwendigen Bedingungen erfüllt sein müssen, um über modale Begriffe zu verfügen. Unter *Reaktionsabhängigkeit* wird verstanden, dass ein Begriff eine menschliche Reaktion impliziert, so wie dies beispielsweise Begriffe von sekundären Qualitäten tun. Oder präziser:

**Definition 8.3-1 (Reaktionsabhängigkeit):** Der Begriff einer Eigenschaft C ist *reaktionsabhängig*, gdw. es eine (sensorische, affektive oder kognitive) Reaktion  $R_C$  gibt, so dass ein Bikonditional der folgenden Art a priori wahr ist:

$x$  ist C  $\leftrightarrow$   $x$  besitzt die Disposition, die Reaktion  $R_C$  in geeigneten Subjekten unter geeigneten Umständen auszulösen.

Man sieht leicht, dass solche Analysen nicht als reduktive Analyse erfolgreich sein können, da die rechte Seite dieser Bikonditionale entweder das Analysandum selbst oder einen eng verwandten Begriff enthält, und zwar an der Stelle, an welcher die Reaktion  $R_C$  spezifiziert wird. Standardbeispiel für solche reaktionsabhängigen Analysen sind *Farbbegriffe*. So kann man ROT (wie bereits erwähnt) analysieren als ‚ $x$  ist rot  $\leftrightarrow$   $x$  ist dazu disponiert, in normalen Beobachtern unter Stan-

<sup>358</sup> Vgl.: Lewis [187].

dardbedingungen Rotempfindungen auszulösen (rot zu erscheinen)'. Eine solche Analyse ist dann insofern eine Angabe der Bedingungen für das *Verfügen* über den Begriff ROT, als normale Subjekte für gewöhnlich Farbwahrnehmungen haben, die sie als primäre Kriterien zur Anwendungen von Farbbegriffen heranziehen, diese Farbwahrnehmungen aber nicht als maßgebend auffassen, wenn intertemporale oder interpersonale Unstimmigkeiten auftreten. In solchen Fällen werden entweder einige wahrnehmende Subjekte oder bestimmte Wahrnehmungsumstände durch korrigierende Praktiken als anormal ausgesondert.

Anstatt einer Analyse des fraglichen Begriffs liefern diese Bikonditionale also eine Kurzfassung der besonderen Merkmale der jeweiligen Reaktion, sowie der jeweiligen korrigierenden Praktiken, die als notwendige Bedingungen für Begriffskompetenz aufgefasst werden. Subjekte, die beispielsweise Farbbegriffe aufgrund von Schmerzempfindungen anwenden, würden dementsprechend genauso wenig über Farbbegriffe *verfügen* wie Subjekte, die sich in einem stockdunklen Zimmer über die Farbe der Tapete streiten, den Disput aber letztlich auf unterschiedliche persönliche Präferenzen zurückführen und beilegen, ohne es als hilfreich anzusehen, das Licht einzuschalten.

Wenn man eine solche Analyse fruchtbar auf modale Begriffe anwenden will, muss man davon ausgehen, dass die Verfügungsbedingungen für modale Begriffe denjenigen von Farbbegriffen hinreichend ähnlich sind. Man wird daher erstens behaupten wollen, dass unser primäres Kriterium zur Anwendung von modalen Begriffen die primitive Reaktion der Vorstellbarkeit ist, und dass wir zweitens korrigierende Praktiken besitzen, die dazu geeignet sind, Reaktionen, die wir als valide Indikatoren für Modalität erachten, immer weiter zu verfeinern. Die Parallele zum Verfügen über Farbbegriffe wäre vollständig, wenn man nun auf die Redeweise von „Normalbedingungen“ zurückgreifen könnte. Da wir uns im Bereich der Vorstellbarkeit allerdings mit komplexen und anspruchsvollen geistigen Operationen befassen, scheinen kognitive Limitationen eher die Regel als die Ausnahme darzustellen. Gerade innerhalb eines reaktionsabhängigen Ansatzes zur Modalität wird die Vorstellbarkeit einer Aussage bzw. Aussagenmenge aber als die Unmöglichkeit aufgefasst, einen Widerspruch aus dieser Aussage abzuleiten:

What exactly is conceivability? What does the mental ability to conceive something consist in? [...] [T]he mental ability to conceive of something is really a complex ability, consisting in the ability to suppose that the state of affairs holds without being able to reduce this supposition to absurdity. Clearly, this complex ability presupposes a number of other more complex abilities: first, the ability to entertain suppositions; and secondly, the ability to infer other propositions, in particular absurd propositions, from suppositions. (Menzies [208], 265)

Bekanntlich sind die Fähigkeiten für solche komplexeren logischen Übungen nicht besonders glücklich verteilt. Statt von Normalbedingungen zu reden, weicht man daher auf *Idealbedingungen* aus, dementsprechend auch auf „ideale Denker“ und „ideales Denken“. Ein idealer Denker wird dann als ein Subjekt verstanden, das unter keiner Form kognitiver Beschränktheit leidet, die durch unsere korrigierenden Praktiken jemals als kognitive Beschränkungen ausgezeichnet werden

könnten (Bedingungen, unter denen – gemäß unseren korrigierenden Praktiken – Vorstellbarkeit nicht als veritabler Indikator angesehen würde).<sup>359</sup> Gegeben den Begriff eines idealen Denkers, können dann dem Farbbegriffsbeispiel ähnliche Bikonditionale für modale Begriffe angegeben werden:

Es ist möglich, dass  $p \leftrightarrow$  ein idealer Denker kann sich vorstellen, dass  $p$ .

Es ist notwendig, dass  $p \leftrightarrow$  ein idealer Denker kann sich nicht vorstellen, dass nicht- $p$ .

#### 8.3.6.2 Reaktionsabhängigkeit und NC'

Die soeben vorgestellte Analyse der Modalbegriffe leidet allerdings unter einem fundamentalen Problem, da auch sie zu (NC') verpflichtet zu sein scheint. Wir erinnern uns an Menzies Analyse des Möglichkeitsbegriffs:

Es ist möglich, dass  $p \leftrightarrow$  Ein idealer Denker kann sich vorstellen, dass  $p$ .

Betrachten wir von diesem Bikonditional zunächst nur die Richtung von rechts nach links:

Ein idealer Denker kann sich vorstellen, dass  $p$ .  $\rightarrow$  Es ist möglich, dass  $p$ .

Diese Teilthese ist offenbar äquivalent mit (NC) und daher auch mit (NC'). Menzies versteht wie (wir gesehen haben) „Vorstellbarkeit“ als formale Fähigkeit, einen Widerspruch zu entdecken, und unter einem „idealen Denker“ ebenfalls dasselbe wie wir. Wenn der ideale Denker dazu berechtigt ist, auf die Möglichkeit von  $p$  zu schließen, weil er feststellt, dass  $p$  negativ vorstellbar ist, muss (NC') auch wahr sein: Möglichkeit ist ja (angeblich) a priori nichts anderes als für einen idealen Denker negativ vorstellbar zu sein.

#### 8.3.6.3 Widerlegung von NC'

(NC') ist also eine zentrale These modaler Epistemologien sowie von gewissen Reaktionsabhängigkeitsansätzen. Wir werden nun sehen, dass (NC') allerdings in seiner intendierten Bedeutung falsch ist, und dass man dies aufgrund bloßer formallogischer Überlegungen feststellen kann.

Nachzuweisen, dass etwas aufgrund bloßer formallogischer Überlegungen falsch ist, ist keine besonders einfache Übung, wenn die fragliche These nicht selbst in einer formalen Version gegeben ist. Jedes Argument, dass ein Prinzip X in einem formalen Modell M falsch ist, kann immer dadurch ausgehebelt werden, dass man schlicht behauptet, M sei eben nicht das einzige adäquate Modell. Einen solchen Einwand kann man prinzipiell auch gegen unser nun folgendes Argument erheben. Allerdings ist das formale Modell, in dem wir (NC') betrachten wollen, eines, das die Intentionen und die Äußerungen von Chalmers, Hale und Menzies auf die natürlichste Weise interpretiert. Wenn diese Ansätze (NC') in einem ande-

<sup>359</sup> Vgl. hierzu insbesondere Menzies [208].

ren Sinne verstanden wissen wollen, als wir es nun rekonstruieren werden, sollten sie ein alternatives formales Modell angeben. Solange kein anderes vorliegt, sollte man das natürlichste nehmen. Wenn kein anderes vorgelegt wird, begehen diese Ansätze bloße Augenwischerei bei der Ablehnung unseres Modells.

Unser Problem besteht zunächst darin, alles, was bisher gesagt worden ist, in einen formalen Rahmen zu übersetzen, der uns eine formale Evaluierung von (NC') erlaubt. Wir haben oben bereits erwähnt, dass (NC') eigentlich eine Korrektheitsbehauptung bezüglich einer bestimmten Ableitungsregel ist. Aber welche Ableitungsregel kommt in Frage? Im Allgemeinen werden Ableitungsregeln als Paare einer Menge von Prämissenformeln und einer Konklusionsformel dargestellt. Folgende Regel ist die Disjunktionseinführungsregel der Prädikatenlogik erster Stufe:

$$\begin{array}{ll} \vdash P_i & \text{Prämissenformel} \\ \cdot & \\ \cdot & \\ \cdot & \\ \vdash P_i \vee \dots \vee P_j \vee \dots \vee P_n & \text{Konklusionsformel} \end{array}$$

Diese Regel gibt in schematischer Weise Prämissenformeln an, die an irgendeiner Stelle im Beweis ableitbar sein können. In diesem Fall sei die Formel  $P_i$  ableitbar an einer bestimmten Stelle. Die Disjunktionseinführungsregel sagt nun, dass wenn eine Instanz von  $P_i$  ableitbar ist, dann auch eine Instanz der Konklusionsformel ableitbar ist. In diesem Falle dürften wir auf eine Disjunktion schließen, die  $P_i$  als ein Disjunktionsglied enthält.

Wie steht es nun mit der Regel, über die (NC') zu reden scheint? Wir wissen, dass nach der fraglichen Regel die Möglichkeit von  $p$  aus seiner negativen Vorstellbarkeit geschlossen werden soll. Wir wissen also schon einmal, wie die Konklusionsformel aussehen muss:

$$\begin{array}{ll} ? & \text{Prämissenformel} \\ ? & \\ ? & \\ \cdot & \\ \vdash \diamond P_i & \text{Konklusionsformel} \end{array}$$

Die Idee hinter negativer Vorstellbarkeit war ja, dass es uns nicht gelingt, einen Widerspruch in  $P_i$  zu finden, und wir deshalb berechtigt sind, auf  $\diamond P_i$  zu schließen. Dass es uns aber *nicht* gelingt, etwas festzustellen, kann nicht in eine Ableitbarkeitsbehauptung, sondern nur in eine Nicht-Ableitbarkeitsbehauptung übersetzt werden, genauer gesagt in die Nicht-Ableitbarkeit von  $P_i$  Negation. Wenn  $P_i$  negative Vorstellbarkeit bedeutet, dass es fehlschlägt, eine Inkonsistenz in  $P_i$  zu finden, ist das äquivalent damit, dass es nicht gelingt die Negation von  $P_i$  abzuleiten:

$\nVdash \neg P_i$	Prämissenformel
.	
.	
.	
$\vdash \diamond P_i$	Konklusionsformel

Wir nennen diese Regel ‚Möglichkeitseinführungsregel‘ oder ‚ $\diamond$ Intro‘. Wir werden annehmen, dass sie die Idee wiedergibt, die hinter negativer Vorstellbarkeit steht.

$\diamond$ Intro soll in unserem Rahmen zur Evaluierung von (NC') eine zweite Regel zur Seite gestellt bekommen, die zur Einführung von Notwendigkeitsoperatoren dient.  $\diamond$ Intro modelliert den Fall, dass wir die Möglichkeit von  $P_i$  annehmen können, weil es uns nicht gelingt, die *Notwendigkeit* von  $\neg P_i$  zu zeigen. In anderen Worten: wenn wir  $P_i$  ableiten können in einem Beweis, sollte dies erlauben, auf die Notwendigkeit von  $P_i$  zu schließen. Diese Regel werden wir ‚Notwendigkeitseinführungsregel‘ oder ‚ $\Box$ Intro‘ nennen:

$\vdash P_i$	Prämissenformel
.	
.	
.	
$\vdash \Box P_i$	Konklusionsformel

Die zweite Regel ist natürlich eine bekannte Regel in normalen Modallogiken. Es handelt sich um die Nezesitationsregel aus Hughes und Cresswell [153] bzw. um Kripkes „R2“ aus Kripke [171]. Die erste Regel aber, die wir für negative Vorstellbarkeit vorgeschlagen haben, ist nicht besonders geläufig in normalen Modallogiken. Dies liegt daran, dass normale Modallogiken *monoton axiomatisierbar* sind, was für eine Logik mit  $\diamond$ Intro nicht der Fall ist<sup>360</sup>:

**Definition 8.3-2 ((Nicht)-Monotonie von Regeln):** Wir nennen eine Regel *monoton*, wenn die Ableitbarkeit der Konklusionsformel nur von der Ableitbarkeit der Prämissenformeln abhängt, bzw. wenn die Nicht-Ableitbarkeit der Konklusionsformel nur von der Nicht-Ableitbarkeit von Prämissenformeln abhängt.

Wir nennen eine Regel *nicht-monoton*, wenn die Ableitbarkeit ihrer Konklusionsformel auch von der Nicht-Ableitbarkeit ihrer Prämissenformeln abhängen kann, bzw. wenn die Nicht-Ableitbarkeit ihrer Konklusionsformel auch von der Ableitbarkeit ihrer Prämissenformeln abhängen kann.

---

<sup>360</sup> Zur folgenden Terminologie vgl. Schurz [292].

**Definition 8.3-3 (Monotone Axiomatisierbarkeit):** Eine Logik ist *monoton axiomatisierbar* gdw. sie eine Axiomatisierung besitzt, die nur aus monotonen Regeln besteht, die nur Ableitbarkeitsbehauptungen enthalten.

Da  $\diamond$ Intro eine nicht-monotone Regel ist, kann eine Logik, die diese Regel enthält, auch nicht monoton axiomatisierbar sein.  $\diamond$ Intro ist das beweistheoretische Pendant zu negativer Vorstellbarkeit. Ohne diese Regel könnten wir keine nicht-trivialen Möglichkeiten ableiten. Eine Möglichkeit ist in diesem Sinne *trivial*, wenn die Formel im Skopus des Möglichkeitsoperators auch dann gültig ist, wenn sie neccessitiert ist. In normalen Modallogiken gibt es solche nicht-trivialen Möglichkeitstheoreme gar nicht. Wenn wir eine Logik negativer Vorstellbarkeit konstruieren möchten, brauchen wir sie aber natürlich. Schließlich geht es bei Vorstellungstests um nicht-triviale Möglichkeitsurteile. Wenn wir solche Modalitäten in unserer Logik repräsentieren wollen, müssen wir auf syntaktischer wie auch semantischer Seite von normalen Modallogiken etwas abweichen. Betrachten wir zunächst die Seite der Semantik:

In Kripkes ‚A Completeness Theorem in Modal Logic‘ (Kripke [171]) werden nur diejenigen Sätze als Theoreme behandelt, die gültig relativ zu *jedem* Unterraum  $\mathbb{W} \subseteq \mathbb{W}$  von möglichen Welten sind. Betrachten wir zunächst die modale Aussagenlogik. Nehmen wir einen Satz der Form  $\diamond A$ , wobei  $A$  eine nicht-modale Formel von  $L_{\text{prop}}$  sei, der propositionalen Modallogik. Wie wir sagten, taucht in der Kripke-Semantik dieser Satz nur dann als Theorem auf, wenn er relativ zu jedem Unterraum  $\mathbb{W} \subseteq \mathbb{W}$  von möglichen Welten gültig ist; aber dann kann  $A$  nur eine Tautologie sein, da nur diese für jeden Unterraum der möglichen Welten garantiert sind.<sup>361</sup>

**Theorem 8.3-1:** Für jede nicht-modale Formel  $A \in L_{\text{prop}}$  und normale propositionale Modallogik  $L$ : wenn  $\diamond A \in L$ , dann  $\Box A \in L$  und  $A$  ist eine wahrheitsfunktionale Tautologie, vorausgesetzt, dass  $L$  konsistent ist.

Die Standardsemantik ist also inadäquat, wenn wir die Logik negativer Vorstellbarkeit modellieren wollen. Wir brauchen für unsere Zwecke nicht-triviale Möglichkeiten als gültige Formeln und als Theoreme, andererseits würde negative Vorstellbarkeit (falls korrekt) uns diejenigen Sätze als möglich präsentieren, die notwendig sind (und dann enthielte negative Vorstellbarkeit *trivial* Möglichkeit).

Anstatt also einen variablen Unterraum von  $\mathbb{W}$  für die Bestimmung der Gültigkeit einer Formel heranzuziehen, nehmen wir einen fixierten Raum  $\mathbb{W}$  an, der alle möglichen Interpretationen einer Sprache enthält, und identifizieren diese Interpretationen mit möglichen Welten. Zum Nachweis, dass dies problemlos funktioniert, und um den Unterschied zu normalen Modallogiken deutlich zu ma-

---

<sup>361</sup> Für den Beweis dieses Theorems vgl. Schurz [292]; Cohnitz [72], Appendix II.

chen, betrachten wir das Beispiel einer Sprache der propositionalen Modallogik (aus Schurz [292]<sup>362</sup>).

Unsere Logik  $C_{\text{prop}}$  bestehe aus einer abzählbaren Menge  $P$  von Aussagenvariablen  $p, \dots, q, \dots$ , sowie den Standardjunktoren und Operatoren  $\neg, \wedge, \vee, \supset, \top, \perp, \Box, \Diamond$ . Wir nehmen Standardinterpretationen (Wahrheitsbewertungen)  $I: P \rightarrow \{0, 1\}$  an, sowie eine klassische Semantik für die nicht-modalen Formeln. Wie wir sagten, identifizieren wir Interpretationen mit möglichen Welten,  $W$  ist die Menge aller möglichen Welten (aller Interpretationen  $I: P \rightarrow \{0, 1\}$ ). Ein Satz ist *logisch wahr* gdw. er in allen Welten wahr ist. Jetzt können wir auch die Semantik für modale Aussagen angeben:  $\Box A$  ist wahr (in einer gegebenen Welt) gdw.  $A$  in allen möglichen Welten wahr ist, gdw.  $A$  logisch wahr ist. Die Wahrheit nezesstierter Sätze ist damit eindeutig logisch determiniert: wenn sie wahr sind, sind sie logisch wahr, wenn sie falsch sind, logisch falsch:

Wenn  $\models_c A$ , dann  $\models_c \Box A$ .

Wenn  $\not\models_c A$ , dann  $\not\models_c \neg \Box A (= \Diamond \neg A)$ .

Wenn wir die Semantik auf diese Weise angeben, erhalten wir sofort alle Theoreme von  $S5$ , aber auch, dass für jede Aussagenvariable  $p \in P$  gilt, dass  $\Diamond p$  ein Theorem von  $C_{\text{prop}}$  ist. Dasselbe gilt für jeden  $C_{\text{prop}}$ -konsistenten Satz  $A$ . Dies ist allerdings so, wie es auch sein sollte, da es sich ja hierbei genau um die nicht-trivialen Möglichkeiten handelt, die wir von unserer Semantik haben wollten.

Allerdings müssen wir für diese Eigenschaft an anderer Stelle einen Preis bezahlen.  $C_{\text{prop}}$  ist nicht unter homomorpher Substitution geschlossen, wie man leicht sehen kann: Betrachten wir  $\models_c \Diamond p$ . Wie wir sagten gilt dies, weil  $p$  von mindestens einer der  $I$  mit ‚wahr‘ bewertet wird. Nun ist aber  $\Diamond(p \wedge \neg p)$  eine Substitutionsinstanz von  $\Diamond p$ . Klarerweise gilt aber nicht  $\models_c \Diamond(p \wedge \neg p)$ , da jede  $I$   $p \wedge \neg p$  falsch macht.

In anderen Worten: wir werden keine unrestringierte Substitutionsregel in unserem deduktiven System haben können, wenn dieses System korrekt sein soll. Nichtsdestotrotz ist  $C_{\text{prop}}$  geschlossen unter syntaktisch isomorpher Substitution sowie semantisch isomorpher Substitution (vgl. Cohnitz [72], Schurz [292]).

**Theorem 8.3-2:**  $C_{\text{prop}}$  ist unter allen syntaktisch isomorphen und semantisch isomorphen Substitutionen geschlossen.

Man kann durchaus dafür argumentieren, dass diese Eigenschaften für eine „echte“ Modallogik ausreichen (vgl. Schurz [291], Schurz [292] vs. Makinsons [199]), dies braucht aber hier nicht unsere Sorge zu sein. Dass  $C_{\text{prop}}$  unter homomorpher Substitution nicht geschlossen ist, hängt mit der Frage der monotonen Axiomatisierbarkeit zusammen. Wenn eine Logik unter homomorpher Substituierbarkeit

---

<sup>362</sup> Das Folgende entstand ursprünglich als Rekonstruktion Carnaps „eigentlicher“ Vorstellung von Modallogik. Vgl. hierzu Carnap [49], Carnap [53], Feys [105], Gottlob [130], Gottlob [131], Hendry und Pokriefka [146], Schurz [291], Schurz [292].

nicht geschlossen ist, dann ist sie nicht schematisch axiomatisierbar oder nicht monoton axiomatisierbar (vgl. Schurz [292]).

**Theorem 8.3-3:** Wenn eine Logik schematisch und monoton axiomatisierbar ist, dann sind ihre Theoreme geschlossen unter homomorpher Substitution.

Da  $C_{\text{prop}}$  – wie wir gesehen haben – nicht unter homomorpher Substitution geschlossen sein kann, werden wir entweder keine monotone Axiomatisierung angeben können, oder die Axiomatisierung nicht schematisch halten können. Da wir aber eine Logik modellieren möchten, die  $\diamond$ Intro enthält, können wir alle Regeln und Axiome schematisch halten. Zur Veranschaulichung erfolgt hier eine Axiomatisierung  $A_c$  von  $C_{\text{prop}}$ , die schematisch, korrekt, vollständig und entscheidbar ist (für den Beweis vgl. Cohnitz [72], Appendix II, Schurz [292];  $A_c$  stammt von Schurz [292]):

Die Axiomatisierung von  $C_{\text{prop}}$  erfolgt in Form einer simultanen Axiomatisierung, indem wir die Regeln des natürlichen Schließens verwenden, die in beiden Richtungen gültig sind und in einer Richtung Komplexität reduzieren. Aus Gründen der Einfachheit beschränken wir die betrachteten Junktoren auf die Negation und die Konjunktion. Zunächst werden die (links-nach-rechts) Ableitbarkeitsregeln angegebene, dann die (rechts-nach-links) Nicht-Ableitbarkeitsregeln:

$\Rightarrow$

$$\begin{aligned}
& L_1, \dots, L_n \vdash_{\text{CPROP}} L \text{ gdw. } L \in \{L_1, \dots, L_n\} \text{ oder } L_1, \dots, L_n \vdash_{\text{CPROP}} \perp \\
& L_1, \dots, L_n \vdash_{\text{CPROP}} \perp \text{ gdw. } \exists L, L'' \in \{L_1, \dots, L_n\}: L' = \neg L'' \\
& \Gamma \vdash_{\text{CPROP}} \neg\neg A \text{ gdw. } \Gamma \vdash_{\text{CPROP}} A \\
& \Gamma, \neg\neg A \vdash_{\text{CPROP}} B \text{ gdw. } \Gamma, A \vdash_{\text{CPROP}} B \\
& \Gamma \vdash_{\text{CPROP}} A \wedge B \text{ gdw. } \Gamma \vdash_{\text{CPROP}} A \text{ und } \Gamma \vdash_{\text{CPROP}} B \\
& \Gamma, A \wedge B \vdash_{\text{CPROP}} C \text{ gdw. } \Gamma, A, B \vdash_{\text{CPROP}} C \\
& \Gamma \vdash_{\text{CPROP}} \neg A \wedge B \text{ gdw. } \Gamma, A \vdash_{\text{CPROP}} \neg B \\
& \Gamma, \neg A \wedge B \vdash_{\text{CPROP}} C \text{ gdw. } \Gamma, \neg A \vdash_{\text{CPROP}} C \text{ und } \Gamma, \neg B \vdash_{\text{CPROP}} C \\
& \Gamma \vdash_{\text{CPROP}} \Box A \text{ gdw. } \vdash_{\text{CPROP}} A \text{ oder } \Gamma \vdash_{\text{CPROP}} \perp \\
& \Gamma, \Box A \vdash_{\text{CPROP}} B \text{ gdw. } \Gamma \vdash_{\text{CPROP}} B \text{ oder } \not\vdash_{\text{CPROP}} A \\
& \Gamma \vdash_{\text{CPROP}} \neg \Box A \text{ gdw. } \not\vdash_{\text{CPROP}} A \text{ oder } \Gamma \vdash_{\text{CPROP}} \perp \\
& \Gamma, \neg \Box A \vdash_{\text{CPROP}} B \text{ gdw. } \Gamma \vdash_{\text{CPROP}} B \text{ oder } \vdash_{\text{CPROP}} A
\end{aligned}$$

$\Leftarrow$

$$\begin{aligned}
& L_1, \dots, L_n \not\vdash_{\text{CPROP}} L \text{ gdw. } L \notin \{L_1, \dots, L_n\} \text{ und } L_1, \dots, L_n \not\vdash_{\text{CPROP}} \perp \\
& L_1, \dots, L_n \not\vdash_{\text{CPROP}} \perp \text{ gdw. } \neg \exists L, L'' \in \{L_1, \dots, L_n\}: L' = \neg L'' \\
& \Gamma \not\vdash_{\text{CPROP}} \neg\neg A \text{ gdw. } \Gamma \not\vdash_{\text{CPROP}} A \\
& \Gamma, \neg\neg A \not\vdash_{\text{CPROP}} B \text{ gdw. } \Gamma, A \not\vdash_{\text{CPROP}} B \\
& \Gamma \not\vdash_{\text{CPROP}} A \wedge B \text{ gdw. } \Gamma \not\vdash_{\text{CPROP}} A \text{ oder } \Gamma \not\vdash_{\text{CPROP}} B \\
& \Gamma, A \wedge B \not\vdash_{\text{CPROP}} C \text{ gdw. } \Gamma, A, B \not\vdash_{\text{CPROP}} C \\
& \Gamma \not\vdash_{\text{CPROP}} \neg A \wedge B \text{ gdw. } \Gamma, A \not\vdash_{\text{CPROP}} \neg B \\
& \Gamma, \neg A \wedge B \not\vdash_{\text{CPROP}} C \text{ gdw. } \Gamma, \neg A \not\vdash_{\text{CPROP}} C \text{ und } \Gamma, \neg B \not\vdash_{\text{CPROP}} C
\end{aligned}$$

$$\begin{aligned} \Gamma \not\vdash_{C_{PROP}} \Box A \text{ gdw. } \vdash_{C_{PROP}} A \text{ und } \Gamma \not\vdash_{C_{PROP}} \perp \\ \Gamma, \Box A \not\vdash_{C_{PROP}} B \text{ gdw. } \Gamma \not\vdash_{C_{PROP}} B \text{ und } \vdash_{C_{PROP}} A \\ \Gamma \not\vdash_{C_{PROP}} \neg \Box A \text{ gdw. } \vdash_{C_{PROP}} A \text{ und } \Gamma \not\vdash_{C_{PROP}} \perp \\ \Gamma, \neg \Box A \not\vdash_{C_{PROP}} B \text{ gdw. } \Gamma \not\vdash_{C_{PROP}} B \text{ und } \vdash_{C_{PROP}} A \end{aligned}$$

**Theorem 8.3-4:** Die Axiomatisierung  $A_C$  von  $C_{PROP}$  ist schematisch, korrekt, vollständig und entscheidbar.

Diese Logik formalisiert negative Vorstellbarkeit in dem oben besprochenen Sinne. Da  $C_{PROP}$  korrekt ist, ist (NC') für diese Logik auch wahr. Jeder Satz, der in  $C_{PROP}$  als negativ vorstellbar gilt und daher den Schluss auf eine Möglichkeit erlaubt, drückt auch eine Möglichkeit aus. Darüber hinaus gilt auch die Gegenrichtung von (NC') in dieser Logik, da  $C_{PROP}$  vollständig ist. D.h. das negativ Vorstellbare und das Mögliche fallen zusammen, wie es bei Menzies Reaktionsabhängigkeitsanalyse auch intendiert war. Da dies für den aussagenlogischen Fall zunächst nach einer formalen Möglichkeit aussieht, eine „echte“<sup>363</sup> Modallogik anzugeben, sollten wir uns stärkeren Systemen zuwenden. Wie Carnap<sup>364</sup> schon bemerkte:

Any system of modal logic without quantification is of interest only as a basis for a wider system including quantification. If such a wider system were found to be impossible, logicians would probably abandon modal logic entirely. (Carnap [53], 196)

Wir nennen eine Logik eine *C-modale Erweiterung*,  $C_L$ , eines logischen Systems  $L$  mit einer gegebenen interpretationalen Semantik, wenn  $L$  um die beiden Modaloperatoren und deren jeweilige Semantik erweitert ist. Bisher haben wir eine C-modale Erweiterung der Aussagenlogik betrachtet. Betrachten wir als nächstes eine C-modale Erweiterung der Prädikatenlogik erster Stufe,  $C_{FOL}$ .

Natürlich interessieren wir uns auch hier für die metalogischen Eigenschaften einer solchen Erweiterung. Betrachten wir als erstes die Frage der Vollständigkeit einer solchen Logik. Vollständigkeit wäre zunächst die Gegenrichtung von (NC'), demnach also eigentlich nicht relevant für die Frage einer modalen Epistemologie. Es mag sehr wohl sein, dass es so etwas wie eine komputationale Grenze dafür gibt, was wir a priori wissen können, bzw. was wir nur auf der Grundlage der Kenntnis der logischen Notwendigkeiten über die Menge der logischen Möglichkeiten wissen können. Für den oben skizzierten Ansatz einer Reaktionsabhängigkeitsanalyse der Modalitäten ist die Vollständigkeit von  $C_{FOL}$  allerdings relevant. Falls es – wie Menzies behauptet – a priori so ist, dass genau dasjenige, was ideal negativ vorstellbar ist, auch möglich ist, sollte die Logik idealer Vorstellbarkeit

<sup>363</sup> Schurz [291], Schurz [292] verteidigt  $C_{PROP}$  als einzige echte Modallogik, da sie nicht-triviale Möglichkeitstheoreme enthält.

<sup>364</sup> In Anbetracht der Tatsache, dass Carnaps Idee einer Modallogik, wie sie Schurz [292] zu rekonstruieren meint, in der Tat nicht prädikatenlogisch erweitert werden kann (zumindest nicht ohne einschneidende Abweichungen, wie wir noch sehen werden), wirkt diese Aussage Carnaps fast tragisch.

auch vollständig sein, oder das Bikonditional ist in einer Richtung falsch. Wenn die Logik nicht vollständig ist, gibt es Möglichkeiten, die nicht negativ vorstellbar sind.

Es ist allerdings leicht zu sehen, dass die erste Stufe-Version unserer Logik negativer Vorstellbarkeit unvollständig ist, wenn man annimmt, dass es die Logik ist, mit der ein idealer Denker Möglichkeiten erschließt. Um diesen Punkt so einsichtig wie möglich zu machen, präsentieren wir hier einen intuitiv leicht nachvollziehbaren Beweis. Nehmen wir – für eine *reductio* – einmal an  $C_{\text{FOL}}$  sei vollständig und algorithmisch. Letzteres stellt sicher, dass die Theoreme von  $C_{\text{FOL}}$  durch eine deterministische Turing Maschine aufgezählt werden können. Wir machen diese Annahme, da es uns ja um ein epistemologisches Modell des Rasonierens geht. Ein nicht-algorithmisches Verfahren zu postulieren wäre (für unsere Zwecke jedenfalls) uninteressant. Es geht ja gerade darum, durch die Logik der negativen Vorstellbarkeit eine Explikation davon zu erhalten, wie wir Wissen von nicht-trivialen Möglichkeiten haben können.

Konstruieren wir uns zu diesem Zweck einmal einen idealen Denker, Hale 9000, einen Supercomputer, den wir mit den Axiomen der polyadischen Prädikatenlogik erster Stufe (FOL) programmiert haben. Von dieser Basisklasse von logischen Notwendigkeiten wissen wir bereits, dass sie vollständig ist, weil wir wissen, dass FOL vollständig ist. Darüber hinaus verfügt Hale 9000 auch über unsere erweiterten Ableitungsregeln. Im Grunde sollte Hale 9000 unsere Idealbedingungen also erfüllen. Hale 9000 leidet unter keinerlei kognitiven Beschränkungen oder Informationsdefiziten. Hale 9000 sollte in der Lage sein, Möglichkeiten aufzuspüren und Beweise durchzuführen. Lassen wir Hale 9000 einmal die Theoremeigenschaft wohlgeformter Formeln testen.

Hale 9000 wird so programmiert, dass er für jede wohlgeformte Formel  $A$ , die wir auf ihren Theoremstatus prüfen wollen, zwei Beweise startet, die er simultan ausführt. Im ersten Beweis versucht er zu beweisen, dass  $\Box A$ , im zweiten Beweis versucht er zu beweisen, dass  $\Diamond \neg A$ . Nun wissen wir, dass für jede Formel  $A$  von  $C_{\text{FOL}}$  entweder  $\Box A$  oder  $\Diamond \neg A$  eine gültige Formel ist. Daher sollte, wenn  $C_{\text{FOL}}$  vollständig ist, Hale 9000 nach endlich vielen Beweisschritten garantiert mit einem Beweis für eine der beiden Formeln fertig werden. Wäre dem so, hätten wir einen rein mechanischen Weg gefunden, die Theoremeigenschaft für alle Formeln der polyadischen Prädikatenlogik zu entscheiden. Die polyadische Prädikatenlogik ist ein echter Teil von  $C_{\text{FOL}}$  und demnach sollte die Entscheidbarkeit von Formeln von  $C_{\text{FOL}}$  eine Frage endlicher Analyse sein. Wir wissen aber unabhängig durch den Beweis von Church aus dem Jahre 1936, dass die polyadische Prädikatenlogik in diesem Sinne unentscheidbar ist. Daher ist  $C_{\text{FOL}}$  unvollständig oder nicht algorithmisch und dann keine Logik, die ein idealer Denker dazu verwenden könnte, Wissen von nicht-trivialen Möglichkeiten zu erhalten. Q.E.D.

**Theorem 8.3-5:** Eine C-modale Erweiterung einer unentscheidbaren Logik  $L$ ,  $C_L$ , ist unvollständig oder nicht algorithmisch.

Wie wir schon beobachtet haben, ist Unvollständigkeit zunächst kein gravierendes Problem für die modale Epistemologie. Hale 9000 mag nicht jede logische Möglichkeit wissen können, es kann durchaus ein hinreichendes Ergebnis sein, wenn Hale 9000 einige dieser nicht-trivialen Möglichkeiten auf zuverlässigem Wege erschließen kann. (NC') war keine Behauptung darüber, dass Möglichkeit negative Vorstellbarkeit enthält, sondern darüber, dass negative Vorstellbarkeit Möglichkeit enthält: dass dasjenige, was wir unter Idealbedingungen als negativ vorstellbar und deshalb als möglich betrachten, tatsächlich möglich *ist*.

Untersuchen wir also als nächstes die folgende Frage: Angenommen, dass  $C_{\text{FOL}}$  *unvollständig* ist, ist es dann *korrekt*, Möglichkeiten mit  $\diamond$ Intro abzuleiten, bzw. Notwendigkeiten mit  $\Box$ Intro? Theorem 8.3-5 unterrichtet uns darüber, dass unsere Logik entweder nicht vollständig ist oder nicht für epistemologische Zwecke taugt. Das lässt ja noch hoffen, dass sie unvollständig ist, aber trotzdem korrekt, so wie die Prädikatenlogik zweiter Stufe zwar unvollständig, aber natürlich korrekt ist. Doch auch hier können wir sehr leicht ein vernichtendes Ergebnis beweisen: Korrektheit ist ausgeschlossen, wenn wir annehmen, dass die zusätzlichen Ableitungsregeln in  $C_{\text{FOL}}$  erlaubt sind.

Intuitiv bedeutet die Unvollständigkeit einer C-modalen Erweiterung einer Logik L, dass es eine Formel A von  $C_L$  gibt, so dass  $\models_C A$  aber  $\not\models_C A$ . Dank  $\diamond$ Intro können wir aber von  $\not\models_C A$  auf  $\vdash_C \diamond\neg A$  übergehen. Die Umformungsregeln<sup>365</sup> für die Modaloperatoren erlauben uns auf den äquivalenten Ausdruck  $\vdash_C \neg\Box A$  überzugehen. Da  $\models_C A$ , wissen wir aus der Semantik von  $C_{\text{FOL}}$ , dass dann auch  $\models_C \Box A$ . Unvollständigkeit von  $C_{\text{FOL}}$  würde also gleichzeitig die Nicht-Korrektheit von  $C_{\text{FOL}}$  implizieren.

Umgekehrt impliziert die Nicht-Korrektheit von  $C_{\text{FOL}}$  übrigens auch die Unvollständigkeit (falls  $C_{\text{FOL}}$  konsistent ist). Nehmen wir einmal Nicht-Korrektheit an und für die reductio auch Vollständigkeit. Dann gibt es eine Formel A von  $C_{\text{FOL}}$ , so dass  $\not\models_C A$  aber  $\vdash_C A$ . Durch  $\Box$ Intro können wir auf  $\vdash_C \Box A$  übergehen, während aufgrund der Semantik der Modaloperatoren  $\models_C \diamond\neg A$  gilt. Letzteres ist wieder äquivalent mit  $\models_C \neg\Box A$ . Wäre  $C_{\text{FOL}}$  vollständig, müsste dann auch  $\vdash_C \neg\Box A$  gelten. dann wäre  $C_{\text{FOL}}$  aber inkonsistent. Q.E.D.

Dieses Resultat kann man folgendermaßen verallgemeinern (Beweis in Schurz [292]):

**Theorem 8.3-6:** Für jede konsistente Axiomatisierung einer Logik des Typs  $C_L$ , in der die Regeln  $\diamond$ Intro und  $\Box$ Intro uneingeschränkt zulässig sind, gilt Folgendes:  $C_L$  ist semantisch korrekt gdw.  $C_L$  semantisch vollständig ist.

In anderen Worten: sobald wir es mit dem modalen Status von Sätzen zu tun haben, die in inferentiellen Beziehungen zueinander stehen, die die Ausdrucksressourcen der monadischen Prädikatenlogik übersteigen und deshalb eine nicht-ent-

<sup>365</sup> Diese werden typischerweise schon von vornherein so festgelegt, damit nur ein Operator dem Vokabular angehört, während der andere Operator nur als Abkürzung definiert ist.

scheidbare Basislogik erfordern, kann negative Vorstellbarkeit Möglichkeit nicht mehr enthalten.

Falls FOL (oder ein anderes unentscheidbares System) in richtiger Weise unseren Begriff der logischen Wahrheit einfängt, kann Wissen von logischen Möglichkeiten nicht aus dem Wissen von logischen Notwendigkeiten *erschlossen* werden.

Falls FOL zusammen mit Bedeutungspostulaten eine Explikation von analytischer Wahrheit darstellt, kann unser Wissen von begrifflichen Notwendigkeiten nicht ausreichen, um uns Wissen von begrifflichen Möglichkeiten zu verschaffen, etc. In all diesen Fällen enthält negative Vorstellbarkeit nicht Möglichkeit.

#### 8.3.6.4 Ist der Konsistentismus widerlegt?<sup>366</sup>

Wir sind von der Überlegung ausgegangen, dass eine rationale Rekonstruktion des Vorstellbarkeitstests dann in einer wissenschaftstheoretisch und erkenntnistheoretisch befriedigenden Weise erfolgt ist, wenn die vermutete Zuverlässigkeit dieses phänomenologisch beschriebenen Tests auf einer anderen Beschreibungsebene als deduktiver logischer Schluss erwiesen werden könnte, der unter Idealbedingungen zu wahren Ergebnissen führt. Diese Idealbedingungen sind – wie bei logischen Schlüssen üblich – wahre (und hier auch vollständige) Prämissen, sowie ideale Berechnungskapazitäten. ‚Wahre Ergebnisse‘ impliziert, da es sich ja um einen logischen Schluss handeln soll, die *Korrektheit* der fraglichen Schlussregel.

In unserem modalen Rasonieren ziehen wir Schlüsse aus den impliziten und expliziten Überzeugungen, die wir haben. Manchmal kommen wir zu der Überzeugung, ein bestimmter Sachverhalt sei möglich, weil wir unter unseren Überzeugungen keine finden können, die diesen Sachverhalt ausschließen würde, wir können nicht zeigen, dass der fragliche Sachverhalt unmöglich ist. Natürlich kann dabei eine Menge schief laufen. Es kann sein, dass wir nicht alle relevanten Notwendigkeiten kennen und dass eine dieser noch unbekanntem Notwendigkeiten mit dem fraglichen Sachverhalt inkompatibel ist. Es mag sein, dass wir zwar alle fraglichen Notwendigkeiten kennen, aber nicht bemerken, dass sie alle relevant sind, oder nicht merken, dass sie in Widerspruch zum fraglichen Sachverhalt stehen, etc. Wie dem auch sei, die Tatsache, dass unser Begriff von logischer Möglichkeit so ist, dass dasjenige logisch möglich ist, was durch die Gesetze der Logik nicht ausgeschlossen ist<sup>367</sup>, macht es mehr als plausibel anzunehmen, dass dies auch der Weg ist, wie wir von diesen Möglichkeiten Wissen haben. Zumindest unter idealen Umständen, in denen wir alle fraglichen Gesetzmäßigkeiten kennen und keinen Fehler im Ableiten begehen, sollte negative Vorstellbarkeit zur Wahrheit führen. In solchen Fällen sollten die inferentiellen Abläufe in unserem Geist mit den ontologischen Zusammenhängen im Bereich des Möglichen und Notwendigen harmonieren. Wenn unser Geist in diesem Sinne als Beweistheorie re-

<sup>366</sup> Für eine etwas ausführlichere Diskussion vgl. Cohnitz [72]. Dort sind insbesondere einige der Reaktionen von Hale und Chalmers diskutiert.

<sup>367</sup> Wie auch unser Begriff von physikalischer, metaphysischer, technischer, etc. Möglichkeit so ist, dass dasjenige physikalisch, metaphysisch, begrifflich, etc. möglich ist, was durch die Gesetze der Physik, Metaphysik, Sprache, etc. nicht ausgeschlossen ist. Vgl. Bremer und Cohnitz [34] und die Zwiebel von Notwendigkeiten und Möglichkeiten in Kapitel 5.2.

konstruiert werden kann, dann sollte eine idealisierte Version davon erschließen können, was tatsächlich möglich ist. Diese starke Intuition steht sicherlich im Hintergrund von Menzies' Konzeption und (mit gewissen Abstrichen) auch hinter Chalmers' und Hales Erkenntnistheorie. Wie wir gesehen haben, ist sie unter bestimmten sehr plausiblen Annahmen nicht zu verteidigen.

Negative Vorstellbarkeit enthält nur dann Möglichkeit, wenn wir unser modales Rasonieren auf einen Teilbereich unseres modalen Wissens einschränken, nämlich denjenigen, der in einer entscheidbaren Basislogik ausgedrückt werden kann. Eine solche Einschränkung ist *ad hoc* und inadäquat. Für Menzies Reaktionsabhängigkeitsansatz kann das nicht gehen, da es natürlich logische Unmöglichkeiten gibt, die nicht durch ein entscheidbares Fragment der Prädikatenlogik ausgeschlossen werden. Falls wir das a priori wissen können, kann Menzies Bikonditional nicht a priori wahr sein. Außerdem schlussfolgern wir über den modalen Status polyadisch logischer Sätze, sonst hätten wir nie eine unentscheidbare Prädikatenlogik entwickelt.

Das hier vorgeführte Resultat gilt allgemein: wir haben ein Modell modalen Rasonierens angegeben, das den natürlichsten Annahmen über logische Notwendigkeit, Möglichkeit und negative Vorstellbarkeit folgt. Negative Vorstellbarkeit impliziert nicht etwa deswegen nicht Möglichkeit, weil es irgendwelche nicht wissbaren Notwendigkeiten gibt, sondern einfach weil es auf einer zu naiven Annahme darüber aufgebaut ist, was es heißt, die Möglichkeit von  $p$  daraus abzuleiten, dass man nicht in der Lage ist,  $p$  auszuschließen. ‚Nicht in der Lage sein,  $p$  auszuschließen' ist eben eine Behauptung über die Nicht-Ableitbarkeit von  $p$ , und solche Behauptungen sind mit Vorsicht zu genießen, wenn die Zusammenhänge ein bisschen komplizierter werden.

#### 8.3.6.5 Ein Ausweg aus dem Problem?

Dass unser Resultat in Bezug auf den Konsistentismus uneingeschränkt gilt, liegt an den Erklärungsansprüchen, die wir mit einer rationalen Rekonstruktion verbinden. Erklärungsansprüche – das ist zumindest eine Denkweise, die uns bisher begleitet hat – sind eine Frage von festgesetzten Kriterien.<sup>368</sup> Ob etwas erklärt ist, hängt davon ab, ob es letztendlich verstanden worden ist, nicht aber davon, ob man alles dazu gesagt hat, was sich dazu sagen lässt. Letzteres hat damit zu tun, was man erklären *kann*. Eventuell kann man eben nicht alles erklären, wenn die Ansprüche höher sind, als das, was man sinnvoll sagen kann.

Wir haben im letzten Kapitel einen relativ hohen Erklärungsanspruch an den Vorstellbarkeitstest gestellt. Wir haben verlangt, dass er auf der Ebene der rationalen Rekonstruktion als ein *logisch korrekter Schluss* rekonstruiert werden kann. Dabei hat die Vorstellung im Hintergrund gestanden, dass Gründe rationalen Fürwahrhaltens eng mit deduktiver Logik verbunden sind. Rationale Rekonstruktion bedeutet, dass ein Prozess, der zu Überzeugungsänderungen führt, nur dann als

---

<sup>368</sup> Diese Auffassung von Philosophie ist besonders deutlich von Nelson Goodman artikuliert worden. Vgl. Cohnitz und Rossberg [75].

rational eingestuft und als solcher verstanden werden kann, wenn er auf bekannte, ausgezeichnete Rationalitätsstandards zurückgeführt ist.

Wissenschaftstheorie in dieser Konzeption ist die des (frühen) kritischen Rationalismus<sup>369</sup>, wie auch die des logischen Positivismus, wie wir in Kapitel 2 sahen. Genauso, wie wir den Vorstellbarkeitstest zurückweisen würden, wenn sich prinzipiell zeigen ließe, dass er sich nicht in einer Weise rational rekonstruieren ließe, wie wir es soeben mit negativer Vorstellbarkeit versucht haben, hat Popper induktive Verfahren in den Wissenschaften zurückgewiesen (vgl. Kapitel 2). Wenn wir nur die Standardlogik als Maßstab für rationale Rekonstruierbarkeit akzeptieren und es neben dem Explikationsversuch über negative Vorstellbarkeit keinen anderen sinnvollen Ansatz gibt, wie man den phänomenologisch beschriebenen Vorstellbarkeitstest erklären und rechtfertigen könnte, sollte Vorstellbarkeit dasselbe Schicksal erfahren. Ob ein solcher prinzipieller Nachweis gelingt, kann hier nicht untersucht werden.<sup>370</sup>

#### 8.3.6.5.1 *Rationale Rekonstruktion: Mach und Kuhn*

Wie wir soeben sagten, entspricht unsere rationale Rekonstruktion negativer Vorstellbarkeit, bzw. unser Urteil, dass eine solche Rekonstruktion nicht möglich ist, einer wissenschaftstheoretischen Position, bei der im Vorhinein festgelegt ist, was als rationale Rekonstruktion und damit als Erklärung oder Rechtfertigung einer wissenschaftlichen Verfahrensweise akzeptiert werden kann. Wie wir in Kapitel 2 sahen, ist dies nicht die einzige Möglichkeit, Wissenschaftstheorie zu betreiben. Ernst Mach (zum Beispiel) hatte wissenschaftliche Verfahrensweisen, die auf der phänomenologischen Ebene erfolgreich schienen, als Datum genommen und versucht, Gesetze der „Denkökonomie“ zu finden, die dann diesen Daten entsprachen, anstatt von a priori Gesetzen der Denkökonomie auszugehen und dann alles aus dem Rechtfertigungszusammenhang auszusieben, was diesen nicht entspricht. Thomas S. Kuhn war – wie wir gesehen haben – Mach in dieser Richtung gefolgt. Auch Kuhn ging es zunächst um eine möglichst genaue Phänomenologie der wissenschaftlichen Verfahrensweisen, von denen dann in einem zweiten Schritt gezeigt wurde, dass sie den auf sie angelegten klassischen Rationalitätsschablonen nicht entsprechen. Wie Stegmüller es so passend formuliert hatte: Kuhn hat dadurch „nicht die Irrationalität im Verhalten von Wissenschaftlern aufgezeigt, sondern neue Dimensionen wissenschaftlicher Rationalität erblickt.“

Dies bedeutet für Stegmüller, dass man die Schablonen anpassen muss. Das bedeutet aber nicht, dass man aufgeben muss oder darf, nach formalen Systemen zu

<sup>369</sup> Ein prototypischer gegenwärtiger Vertreter wäre Alan Musgrave. Musgrave vertritt die Auffassung, dass nur deduktive Argumente und Schlüsse, die in der Standardlogik als gültige Schlüsse repräsentiert werden können, zu rationalen Überzeugungsänderungen führen, entsprechend macht es für ihn auch keinen Sinn, eine andere Logik als die Standardlogik zur Repräsentation unseres Schließens und Argumentierens zu entwickeln. Vgl. Musgrave [216].

<sup>370</sup> Eine denkbare Alternative wäre modelltheoretischer Natur und würde in einem engeren Sinne positive Vorstellbarkeit rational rekonstruieren. Barwise [11] scheint so etwas im Sinn gehabt zu haben. Vgl. auch Bremer und Cohnitz [34]. Eine Diskussion kann hier nicht geleistet werden. Wir kommen auf diese Idee im Schlusskapitel zurück.

suchen, die eine formale Rekonstruktion erlauben würden. Aus dem Scheitern formaler Mittel bestimmter Art kann man nicht auf das Scheitern formaler Mittel insgesamt schließen.<sup>371</sup>

Formale Mittel sind gerade unter solchen Umständen in besonderer Weise notwendig, weil gerade, wenn wir auf andere Konzeptionen von ‚Logik‘, ‚Schluss‘, ‚beweisbar‘ ausweichen müssen, um etwas rational zu rekonstruieren, die uns nicht so vertraut sind wie die der Standardlogik, es besonders wichtig ist, dass sie in präziser und klarer Weise expliziert werden. In anderen Worten: wenn man der Meinung ist, der Vorstellbarkeitstest wäre in seiner phänomenologischen Beschreibung hinreichend klar identifiziert und in der Tat zuverlässig (bzw. wenn es ausgemacht wäre, dass wir Wissen von nicht-trivialen Möglichkeiten besitzen und dieses Wissen nicht bloß aus unserem Wissen von *actualia* beziehen), so kann die Tatsache, dass die Standardlogik zur Rekonstruktion dieses Tests und zur Rechtfertigung unseres Wissens ungeeignet ist, keine Rechtfertigung dafür abgeben, eine rationale Rekonstruktion mit formalen Mitteln zu unterlassen. Im Gegenteil sollte es dann gerade die Aufgabe von Chalmers, Menzies und Hale sein, in einem klaren formalen System anzugeben, wie sie ‚Enthaltensein‘, ‚Ableiten‘ etc. denn sonst verstanden wissen wollen. Zunächst ist festzuhalten, dass eine solche alternative Rekonstruktion von *diesen* Autoren nicht erfolgt ist (und sie auch keinen Bezug auf irgendwelche anderen Rekonstruktionen nehmen, die eine alternative Interpretation ihres Sprachgebrauchs nahe legen würde).

Ob man dann ein solches alternatives formales System als Rechtfertigung und Erklärung akzeptiert, hängt letztendlich davon ab, welche Erklärungsstandards man ansetzt. Bleibt man bei der Auffassung, dass die Dynamik rationalen Fürwahrhaltens nur über die Standardlogik erklärt und gerechtfertigt werden kann, wird man einer solchen Alternative eben nicht folgen. In dem Maße, in dem der Vorstellbarkeitstest zu den Verfahren der Wissenschaften gehört, wird man dann eine Rationalitätskonzeption verteidigt haben, die an den Wissenschaften (zum Teil) vorbeigeht.<sup>372</sup>

Bevor wir das Gebiet der modalen Epistemologie wieder verlassen, soll in wenigen groben Zügen dargestellt werden, wie eine solche Schablonenerweiterung aussehen könnte, aber eine Diskussion ihrer Erklärungs- und Rechtfertigungskraft (aus den soeben genannten Gründen) unterlassen.

---

<sup>371</sup> Vgl. Cohnitz [72] und den dort diskutierten Einwand Chalmers', der in etwa besagt, dass die Gödel-Resultate gezeigt hätten, dass man in der Erkenntnistheorie mit formalen Mitteln nicht weiter kommt. Das gehört meines Erachtens in das ermüdend lange Kapitel falscher Schlussfolgerungen aus den Gödel-Resultaten.

<sup>372</sup> Auch das muss nicht problematisch sein. Vielleicht sind die Wissenschaften nicht der Standard der Rationalität, sondern die Wissenschaftstheorie als normatives System. Genauso kann man sich auch auf den (schwächeren) Standpunkt stellen, dass sich die Rationalitätsstandards der Wissenschaftstheorie mit den Verfahrensweisen der Wissenschaften größtenteils in einem reflektierten Gleichgewicht befinden, bei dem nur der (marginale) Vorstellbarkeitstest als irrationales Mittel zur Bestimmung dessen, was möglich ist, heraus fällt.

### 8.3.6.5.2 *Klassische Logik und interne Statik*

Wenn man die Standards verändern möchte, sollte man sich zunächst fragen, was an den „alten“ Standards denn verkehrt war. Wenn auch etwas unklar formuliert, lassen sich sowohl bei Chalmers als auch bei Hale Hinweise darauf finden, was am Standard der deduktiven Standard-Logik geändert werden müsste, um einen besseren Rationalitätsstandard abzugeben.

Alternatively, one can dispense with the notion of an ideal reasoner and simply invoke the notion of undefeatability by better reasoning. Given this notion, we can say that *S* is ideally conceivable when there is a possible subject for whom *S* is prima facie conceivable, with justification that is undefeatable by better reasoning. The idea is that when prima facie conceivability falls short of ideal conceivability, then either the claim that the relevant tests are passed will be unjustified, or the justification will be defeatable by better reasoning. For ideal conceivability, one needs justification that cannot be rationally defeated. (Chalmers [58], 148)

[A] procedure does not have to be effective or algorithmic in order for its implementation with negative result to provide what are admitted to be defeasible claims. (Hale [139], 16)

Beide Zitate legen nahe, dass die Konzentration auf die Standards der deduktiven Logik gewisse erlaubte Dynamiken rationalen Fürwahrhaltens unzulässiger Weise ausschließt. Worin bestehen diese?

Die Standardlogik ist in Bezug auf ihr metalogisches Vokabular, aber auch in Bezug auf die in ihr erlaubten Beweisschritte eine recht statische Angelegenheit. In der Standardlogik gibt es zunächst einmal nur zwei Arten von Dynamik, eine *externe* und eine *interne* Dynamik. Die externe Dynamik der Standardlogik besteht darin, dass die Menge an logischen Konsequenzen zunehmen kann, wenn die Menge der Prämissen erweitert wird. Eine Nicht-Standard Dynamik bestünde beispielsweise darin, dass nicht nur die Menge der Konsequenzen wächst oder wachsen kann, wenn neue Prämissen hinzugefügt werden, sondern Konsequenzen auch *revidiert* werden können, wenn neue Prämissen neue Informationen bringen. Diese zweite Art der externen Dynamik erfordert eine *nicht-monotone Folgerungsrelation*, die die Standardlogik nicht bereitstellt. Da wir einen bestimmten Idealfall betrachtet haben, in dem die Menge der Prämissen als fixiert betrachtet wurde, hat diese Dynamik keine besondere Rolle gespielt. Für den Fall, dass man ein vollständiges Bild unseres modalen Rasonierens geben wollte, müsste man aber sehr wohl beide Formen externer Dynamik modellieren. Neu erkannte Notwendigkeiten können dazu führen, dass Dinge die auf der alten Wissensbasis möglich schienen, nun als unmöglich ausgeschlossen werden müssen. Eine Folgerungsrelation, die unserem modalen Wissenszuwachs entsprechen soll, muss die nicht-monotone externe Dynamik irgendwie repräsentieren.

Eine zweite Form der Dynamik, die uns ebenfalls nicht gekümmert hat, ist die interne Dynamik logischen Schließens. In der Standardlogik kommt diese interne Dynamik zwar als Phänomen vor, wird aber metalogisch nicht repräsentiert.

Wenn wir Konsequenzen aus einer Prämissenmenge ziehen, dann können wir in einem formalen Beweis immer nur eine Regel nach der anderen anwenden und uns so Schritt für Schritt zu einer bestimmten Konsequenz durcharbeiten. Diese Konsequenz ist nicht unmittelbar ableitbar aus den Prämissen, sondern nur mittelbar ableitbar aus anderen Formeln, die wir aus den Prämissen ableiten durften und aus denen wir dann – gemäß den Regeln der Beweistheorie – nun auf eben jene Konsequenz schließen dürfen. Ableitbarkeit in der Metalogik ist aber eine Alles oder Nichts Angelegenheit. Eine Formel gilt als ableitbar aus einer Prämissenmenge, wenn es für sie einen Beweis gibt, gegeben diese Prämissenmenge und die Regeln der Beweistheorie. Diese Eigenschaft der Beweisbarkeit ist nicht relativ zu einem bestimmten Zeitpunkt im Prozess des Rasonierens (eine Formel, die nicht beweisbar ist, wird es nicht, solange keine neuen Prämissen hinzukommen).

Noch kann eine Formel jemals aufhören beweisbar zu sein (selbst wenn die Prämissenmenge erweitert wird). Diese zweite Form der internen Dynamik ist aber besonders interessant, wenn wir uns mit der Logik des modalen Rasonierens beschäftigen. Gerade beim modalen Rasonieren ist es – wie wir gesehen haben – so, dass ein Schluss, der zu einem bestimmten Zeitpunkt als gültig erscheint, zu späteren Zeitpunkten zurückgenommen werden muss. Wenn etwas zu einem Zeitpunkt im Beweis als möglich gilt, weil seine Negation nicht bewiesen ist, kann es sein, dass dieses Möglichkeitsurteil revidiert werden muss, wenn der Beweis der fraglichen Negation dann doch gelingt. Dies scheint die von Chalmers reklamierte Dynamik zu sein: Statt Ableitbarkeit durch einen idealen Denker bräuchte man so etwas wie Ableitbarkeit, die nicht durch weitere Überlegungen revidiert werden kann. Damit ließe sich negative Vorstellbarkeit so definieren, dass sie tatsächlich Möglichkeit „enthielte“. Darüber hinaus müsste man von einer solchen Logik mit interner Dynamik aber auch verlangen, dass sie es erlaubt, an formalen Kriterien einzuschätzen, ob dieser Zustand erreicht ist, oder wenigstens, ob man diesem Zustand durch das Rasonieren näher gekommen ist.

#### 8.3.6.5.3 Adaptive Logik als Ersatz

Interne Dynamiken sind das Spezialgebiet der Adaptiven Logik (AL). Zu den Charakteristika einer solchen Logik gehört eine so genannte „Upper Limit Logic“ (ULL), eine „Lower Limit Logic“ (LLL) und eine adaptive Strategie. Bekannt sind Adaptive Logiken hauptsächlich im Zusammenhang mit Parakonsistenz.<sup>373</sup> In der parakonsistenten Logik wird versucht, logisches Schließen auch dann „vernünftig“ zu ermöglichen, wenn die Prämissenmenge einen Widerspruch enthält. In der Standardlogik führt ein Widerspruch in der Prämissenmenge zur „Explosion“ – aus einem Widerspruch kann man alles ableiten. Will man Explosion verhindern, muss man die Standardlogik zunächst einmal in ihrer Beweistheorie soweit einschränken, dass aus einem Widerspruch nicht mehr alles abgeleitet werden kann. Zugleich muss diese Einschränkung aber so gewählt werden, dass es weiterhin möglich ist, überhaupt Konsequenzen aus der Prämissenmenge zu ziehen.

---

<sup>373</sup> Vgl. Bremer [33].

Die ULL in einer parakonsistenten AL ist typischerweise die Standardlogik (FOL). Die ULL erlaubt die uneingeschränkte Anwendung von Ableitungsregeln und damit die Ableitung so vieler Folgerungen wie möglich.

Die LLL in einer parakonsistenten AL ist hingegen eine parakonsistente Logik, die die Anwendung bestimmter Regeln der Standardlogik blockiert. Die Folgerungsmenge (für eine Prämissenmenge  $\Gamma$ ) der LLL ist kleiner als die der ULL:

$$\text{Con}_{\text{LLL}}(\Gamma) \subseteq \text{Con}_{\text{ULL}}(\Gamma)$$

Nun könnte eine parakonsistente AL einfach wie eine „Flip-Flop“-Logik funktionieren, also so lange als ULL, wie die jeweils gegebene Prämissenmenge  $\Delta$  konsistent ist und als LLL, wenn  $\Delta$  inkonsistent ist. Das wäre aber kaum interessant. Für Adaptive Logiken besteht der Hauptwitz gerade darin, dass sie eine Folgerungsmenge generieren können, die zwischen denen der LLL und der ULL liegt:

$$\text{Con}_{\text{LLL}}(\Delta) \subseteq \text{Con}_{\text{AL}}(\Delta) \subseteq \text{Con}_{\text{ULL}}(\Delta)$$

Die Regeln der ULL, die die LLL ausschließt, gehen ja nur dann fehl, wenn sie auf einen „wahren Widerspruch“<sup>374</sup> angewendet werden. In allen anderen Fällen sind sie aber absolut in Ordnung. In der AL können diese Regeln angewendet werden, solange nicht bekannt ist, dass man sie auf einen wahren Widerspruch anwendet. Stellt sich heraus, dass eine der Prämissen, auf die man die Regeln der ULL angewendet hat, einen Widerspruch enthält, werden alle Formeln, die man von diesem Widerspruch abgeleitet hat, sowie alle Formeln die indirekt von ihm abhängen, wieder zurückgenommen. Das ist die Rolle der internen Dynamik im Fall der parakonsistenten AL.

Die adaptive Strategie besteht nun darin, Bedingungen anzugeben, an denen man erkennt, welche der Prämissen, die mit den ULL-Regeln zu Ableitungen verwendet wurden, eigentlich unzulässig sind. Eine solche adaptive Strategie schlägt sich dann in einer dynamischen Beweistheorie nieder, die ihrerseits wiederum eine eigene dynamische Semantik zugeordnet bekommt.

Die parakonsistente AL ist in den Augen<sup>375</sup> von Diderik Batens *korrektiv*. D.h., dass der eigentliche Standard des logischen Folgerns, die ULL, unter besonderen Umständen korrigiert werden muss. Uns interessieren hier nicht hauptsächlich korrektive Adaptive Logiken, sondern so genannte „ampliative“ Adaptive Logiken. Eine ampliative Adaptive Logik hat (typischerweise) FOL als LLL und erweitert die Konsequenzen von FOL um solche, für die es in FOL keine positiven Test gibt. Genau dies ist in der Logik modalen Rasonierens der Fall. Es gibt – wie wir eben gesehen haben – keinen positiven Test für die nicht-triviale Möglichkeit einer bestimmten Formel (sei es logische Möglichkeit oder Kompatibilität mit einer be-

<sup>374</sup> Vgl. Bremer [33]

<sup>375</sup> Ob eine Logik korrektiv ist oder ampliativ, ist eine Frage der Perspektive. Eine Dialetheist würde eine adaptive parakonsistente Logik als ampliativ ansehen, da für ihn die LLL der Standard ist.

stimmten Prämissenmenge). Betrachten wir als Beispiel die Logik COM von Diderik Batens und Joke Meheus (vgl. Batens und Meheus [12]).

Da es keinen positiven Test für die Kompatibilität einer Formel  $A$  mit einer bestimmten Prämissenmenge  $\Gamma$  gibt, wird mit COM eine Logik mit dynamischer Beweistheorie angegeben. Wie auch im Zusammenhang mit  $A_C$  angenommen wurde, bedeutet die Kompatibilität von  $A$  mit einer Menge  $\Gamma$ , dass  $A$  in einem Modell von  $\Gamma$  wahr ist. Kompatibilität wird dabei ebenfalls wie bei uns mit  $\diamond A$  angegeben. Wenn  $\Gamma_{\text{FOL}} \models A$ , dann ist  $A$  in allen Modellen von  $\Gamma$  wahr, also  $\neg \diamond \neg A$ , d.h.  $\Box A$ . Batens und Meheus führen nun ein modales Gegenstück zu  $\Gamma$  ein,  $\Gamma^\square = \{\Box A \mid A \in \Gamma\}$ .  $\Gamma^\square \vdash_{\text{COM}} \diamond A$  drückt dann aus, dass  $A$  mit  $\Gamma$  kompatibel ist,  $\Gamma^\square \vdash_{\text{COM}} \neg \diamond A$ , dass sie inkompatibel sind.<sup>376</sup>

Wie wir uns erinnern, muss für eine AL eine Menge von „Abnormalitäten“ angegeben werden, damit unzulässige Schlüsse ausgesondert werden können. In parakonsistenten AL waren das Inkonsistenzen. Wird in einer parakonsistenten AL eine Prämissenmenge  $\Gamma$  verwendet, dann wird von  $\Gamma$  so lange angenommen, dass sie konsistent ist, bis das Gegenteil bewiesen wurde. In COM wird entsprechend vorausgesetzt, dass  $A$  mit  $\Gamma$  kompatibel ist (d.h., dass  $\diamond A$  aus  $\Gamma^\square$  ableitbar ist) bis das Gegenteil gezeigt wurde. Nun bestimmt die LLL die uneingeschränkten Regeln von COM, die ULL die „Abnormalitäten“ und die bedingten Regeln und schließlich die adaptive Strategie, wie die Prämissen, aus denen gefolgert wird, so „normal wie möglich“ zu interpretieren sind, was über eine *Markierungsregel* geleistet wird.

Beweise bestehen nun aus Anwendungen von bedingten und uneingeschränkten Regeln. Wenn eine Formel im Beweis durch eine bedingte Regel eingeführt wurde, wird zusätzlich eine Bedingung angegeben. Die Hinzufügung einer solchen Bedingung entfällt, wenn eine uneingeschränkte Regel verwendet wurde, allerdings sind immer alle Bedingungen anzugeben, wenn die neue Zeile von bedingten Zeilen abhängt. Nach der Markierungsregel wird bei jeder neuen Beweiszeile nachgesehen, ob alle Bedingungen, die in den vorhergehenden Zeilen des Beweises aufgeführt sind, auch erfüllt sind. Falls bestimmte Bedingungen nicht erfüllt sind, wird die jeweilige Zeile, wie auch alle Zeilen, die von ihr abhängen, aus dem Beweis genommen, indem sie markiert werden. Diejenigen Formeln, die zu einem Schritt im Beweis als abgeleitet gelten, sind diejenigen, die zu diesem Schritt (noch) keine Markierung haben. Es gibt damit sowohl „zu einem Schritt abgeleitete“ Formeln, wie auch „zu einem Schritt ableitbare“ Formeln.

Der Zusammenhang mit der Semantik wird nun über „endgültige Ableitbarkeit“ hergestellt.  $A$  ist genau dann *endgültig ableitbar* aus  $\Gamma$  in einer Zeile  $i$ , wenn  $i$  nicht markiert ist, und wenn immer, wenn  $i$  in einer Erweiterung des Beweises markiert ist, es eine weitere Erweiterung des Beweises gibt, in der  $i$  nicht markiert ist. Während endgültige Ableitbarkeit der Konzeption Chalmers' entgegenzukommen scheint, ist es natürlich zu einem gegebenen Schritt im Beweis nicht ohne

---

<sup>376</sup> Das führt natürlich (auch in der Aussagenlogik) zu gewissen Unterschieden zu unserem System, über deren Plausibilität man sich vermutlich lange streiten kann (für unsere Zwecke scheint das alles unerheblich).

weiteres klar, ob eine Zeile nicht im weiteren Verlauf nicht doch markiert wird. Aus diesem Grund wird der dynamischen Beweistheorie eine dynamische Semantik (Blocksemantik) zur Seite gestellt, die es auf metalogisch nachvollziehbare Weise erlaubt, im Beweisverlauf Rückschlüsse darauf zu ziehen, zu welchem Grad man die Informationen in den Prämissen im Beweis bereits verwendet hat, um damit zumindest abzuschätzen, wie sehr man den in diesem Schritt ableitbaren Formeln vertrauen darf, dass sie auch ableitbar bleiben. Endgültige Ableitbarkeit kann man in manchen Fällen natürlich auch erkennen, nämlich in all jenen, die sich im entscheidbaren Bereich von FOL befinden.

In die formalen Details von COM einzusteigen, wäre im Vergleich zum philosophischen Ertrag, den dies für unsere Zwecke hätte, unangemessen. Festzuhalten ist, dass eine Logik modalen Urteilens über nicht-triviale Möglichkeit dann angegeben werden kann, wenn man unter ‚Logik‘ ein formales System versteht, das Explikationen solcher Begriffe wie ‚Folgerung‘, ‚Beweis‘, ‚Logische Wahrheit‘ etc. bietet. Nimmt man nur dies als Erklärungsstandard, kann diese Adäquatheitsbedingung erfüllt werden (z.B. durch COM), nimmt man die Standardlogik und ihren statischen Beweisbegriff als Adäquatheitsbedingung für eine rationale Rekonstruktion, kann diese Erklärungsforderung nicht erfüllt werden.

COM als adaptive Logik kann außerdem noch in zwei wichtigen Hinsichten erweitert werden. Zum einen kann man COM um die oben angesprochenen externen Dynamiken erweitern, also die (mutmaßliche) Tatsache mit einbeziehen, dass wir manche notwendigen Wahrheiten nicht a priori kennen. Außerdem kann man auch eine Rangfolge unter den Prämissen einfügen, die dann die Zurücknahme (Markierung) einzelner Beweisschritte davon abhängig macht, wie viel Vertrauen einer bestimmten Prämisse entgegengebracht werden kann. Beide Erweiterungen wären plausibel, wollte man COM als angenäherte Modellierung unseres tatsächlichen modalen Rasonierens betrachten. Wir haben COM zunächst nur als eine Rekonstruktion modalen Rasonierens unter bestimmten Idealbedingungen (vollständige Information) betrachtet.

Damit verlassen wir das Gebiet der modalen Epistemologie. Wie wir gesehen haben, kann man unter bestimmten (nicht unerheblichen) Bedingungen eine Verteidigung des Vorstellbarkeitstests angeben, die sich (unter ebenfalls nicht unerheblichen Bedingungen) rational rekonstruieren lässt. Ob man die Verteidigung des Vorstellbarkeitstests überzeugend findet, hängt davon ab, ob man die phänomenologische Charakterisierung des fraglichen mentalen Aktes wieder erkennt und in Bezug auf a posteriori Notwendigkeiten meint, davon ausgehen zu können, dass wir in den meisten Fällen, in denen es sich um eine a posteriori Notwendigkeit handelt, a priori wissen, dass es sich um eine handelt.

Ob man die rationale Rekonstruktion überzeugend findet und als Erklärung wie als Rechtfertigung akzeptiert, hängt davon ab, ob man die Rationalitätsstandards der Standardlogik gegen die der Adaptiven Logik einzutauschen bereit ist. (Induktionsgegner werden dazu vermutlich nicht bereit sein, schließlich gibt es auch eine adaptive Logik der Induktion.)

Humes Einsicht, dass ein solches Ergebnis uns nicht davon abhalten sollte, aus pragmatischen Gründen weiterhin Urteile über Möglichkeiten abzugeben, auch

wenn uns prinzipiell unklar ist, wie sie fundiert sind, sollte zwar nicht aufgegeben werden, die Frage, woher wir Wissen von Möglichkeiten haben, sollte als erkenntnistheoretisches Problem allerdings ernster genommen werden, als dies bisher der Fall ist.<sup>377</sup>

Im folgenden Kapitel werden wir eine Funktion für das Gedankenexperiment in der Philosophie rekonstruieren, die uns aber auch dann ruhig schlafen lassen sollte, wenn wir nicht das Humesche Vertrauen in die pragmatische Richtigkeit unseres modalen Urteilens haben.

---

<sup>377</sup> Diese Frage ist eben auch wichtig für andere Bereiche der Wissenschaftstheorie. Wie stellen wir fest, ob eine (für wahr gehaltene) empirische Verallgemeinerung gesetzesartig ist? Die Antwort, dass wir sie dann und nur dann als gesetzesartig behandeln, wenn sie aus Fundamentalgesetzen abgeleitet werden kann, ist ziemlich unbefriedigend, gegeben, dass wir von vielen Verallgemeinerungen der Meinung sind, dass sie gesetzesartig sind, aber keine Vorstellung davon haben, wie sie aus Fundamentalgesetzen abgeleitet werden könnten. Auch in diesem Fall scheint es unabhängiges Möglichkeitswissen zu geben, das uns bei dieser Beurteilung hilft. Die Frage ist nur, wie dieses „Wissen“ selbst möglich ist.



## 9. Die Leistungsfähigkeit von Gedankenexperimenten in der Philosophie

In den letzten drei Teilen haben wir versucht, deutlich zu machen, dass je nachdem welche Art philosophischer These zur Debatte steht, Gedankenexperimente zur Überzeugungsänderung unterschiedliche Adäquatheitsbedingungen zu erfüllen haben. Da diese Hinweise bisher nur verstreut vorkamen, wollen wir sie nun systematisch sammeln und mit den erkenntnistheoretischen Einsichten verbinden, die wir in den letzten Kapiteln gewinnen konnten. Wir werden in diesem Kapitel auch eine weitere Rolle von Gedankenexperimenten zur Überzeugungsänderung diskutieren, die wir in unserer bisherigen Betrachtung übergangen haben: *Die Funktion kritischer Gedankenexperimente in normativen Kontexten.*

Gedankenexperimente gegen Realdefinitionen haben Aussagen als Targetthesen, die wahr oder falsch sein können (6.2.2). Ob sie wahr oder falsch sind, hängt davon ab, was möglicherweise bzw. notwendigerweise der Fall ist. Bei normativen Targetthesen (beispielsweise moralischen Prinzipien oder Begriffsexplikationen (6.2.3)) ist dies offenbar anders. Wenn der „naturalistische Fehlschluss“ uns schon darüber aufklärt, dass die Feststellung einer aktuellen Tatsache in der Regel<sup>378</sup> kein Gegenbeispiel gegen eine normative Aussage sein kann, wie kann dann die Feststellung einer bloß möglichen Tatsache ein Gegenbeispiel sein?

Bevor wir uns aber diesem Problem zuwenden, wollen wir – wie gesagt – unsere wichtigsten Ergebnisse rekapitulieren. Wir haben in dieser Untersuchung darauf geachtet, die methodologische Rolle von Gedankenexperimenten in der Philosophie unter der Annahme von Theorien zu verteidigen, die dem Gedankenexperimentieren *prima facie* feindlich gegenüber stehen. Schließlich war es die Aufgabenstellung „eine Charakterisierung und Beurteilung der methodologischen Rolle(n) von Gedankenexperimenten in der Philosophie zu leisten, die möglichst unabhängig von spezifischen Annahmen über Geltungs- und Erkenntnisansprüche nachvollziehbar ist.“

Die Annahmen, die wir herangezogen haben, entstammten hauptsächlich naturalistischen Ansätze in der Philosophie des Geistes und der Erkenntnistheorie, und externalistischen Ansätzen in der Sprachphilosophie.

Wir haben in Kapitel 7.3 zu zeigen versucht, dass, auch wenn man einen semantischen Externalismus annimmt und eine naturalistische Auffassung der Phi-

---

<sup>378</sup> ... wenn wir für einen Augenblick von Instanzen für „Brückenprinzipien“ absehen. Vermittels solcher Prinzipien wie ‚Sollen impliziert Können.‘ könnte es natürlich Tatsachenfeststellungen geben, die als unmittelbares Gegenbeispiel gegen normative Aussagen verwendet werden können. Um solche Fälle geht es bei Gedankenexperimenten in der Ethik aber offenbar nicht. Es wäre außerdem selbst dann noch seltsam, weshalb ein *bloß mögliches* nicht-Können ein tatsächliches nicht-Sollen implizieren sollte.

losophie vertritt, Gedankenexperimente immer noch erfolgreich verteidigt werden können. Dies macht oberflächlich den Eindruck als sei diese Untersuchung ein methodologischer Beitrag zum Naturalismus, sie ist aber nicht so intendiert. Naturalismus und semantischer Externalismus sind vielmehr als *prima facie* „worst case“ herangezogen worden: Falls man zeigen kann, dass Gedankenexperimente auch in naturalistischen und externalistisch konzipierten Methodologien Sinn machen, sollte ihr Status praktisch universal verteidigt sein. Anti-naturalistische und anti-externalistische Ansätze haben entweder mehr Vertrauen in das Apriori oder wollen „metaphysische“ Notwendigkeiten in unsere epistemische Reichweite bringen. In ihren Methodologien sollte das Gedankenexperiment daher einen unproblematischeren Stand haben.

Zugegeben, Kapitel 7.2 und 7.3 waren Rückzugsgefechte. Viele kluge Einwände gegen das Gedankenexperimentieren haben uns immer ein bisschen mehr in die Enge getrieben und die Methode des Gedankenexperiments ist bestenfalls bedingt verteidigt: Falls man mit Gedankenexperimenten nur bestimmte (eingeschränkte) Zwecke verfolgt, dann zieht die vorgetragene Kritik nicht, etc. Man könnte hier einwenden, dass es nicht interessant ist, dass Gedankenexperimente für die beschränkten Zwecke taugen, für die wir zu zeigen versuchten, dass sie nicht von der vorgebrachten Kritik in Bedrängnis gebracht werden. Haben wir also nur eine ziemlich uninteressante These verteidigt?

Nun, ich denke es kommt darauf an, welche Perspektive man dazu einnimmt. Zunächst hat diese „Medaille“ offensichtlich zwei Seiten: wenn wir gesagt haben, dass man *nur* bestimmte Erkenntnis- oder Geltungsansprüche mit dem Gedankenexperiment verbinden kann, dann heißt das eben auch, dass man darüber hinausgehende Ansprüche nach unserer Prüfung nicht damit verbinden kann.

Wenn man aber bestimmte Erkenntnisansprüche mit dem Gedankenexperimentieren nicht verteidigen kann bzw. bestimmte Geltungsansprüche nicht mit dem Gedankenexperimentieren rechtfertigen kann, muss man andere Methoden finden oder diese Geltungsansprüche aufgeben. Das ist sicherlich schon mal für all jene von Interesse, die solche Geltungsansprüche erheben und sonst keine neuen Methoden dem philosophischen Werkzeugkasten hinzugefügt haben, die weniger problematisch wären als das Gedankenexperimentieren.

Zum Zweiten ist zu sagen, dass die übrig gebliebenen Ansprüche so marginal nicht sind. Wir haben zu plausibilisieren versucht, dass Bedeutungsanalysen, selbst wenn man sie nur sehr moderat als Rekonstruktion eines Überzeugungsnetzes versteht, das für die Experten eines philosophischen Problemfeldes mit den Begriffen dieses Problemfeldes eng verbunden ist, eine notwendige Voraussetzung für eine ernsthafte systematische Behandlung dieses Problemgebietes ist. Dies gilt unabhängig davon, ob man der Meinung ist, dass eine solche systematische Auseinandersetzung dann mit den Mitteln der empirischen Psychologie vor sich gehen sollte oder mit den Mitteln formaler Modellbildung. Diesen Standpunkt werden wir in diesem Kapitel noch weiter ausarbeiten.

„Ernsthaft“ bedeutet hier, dass man intendiert, einen Lösungsvorschlag zu einem bereits existierenden wissenschaftlichen Forschungsgebiet zu leisten. Das braucht man freilich nicht zu tun. Natürlich kann man (wie Trout und Bishop in

Bishop und Trout [26]) bekannt geben, dass man von der philosophischen Erkenntnistheorie maßlos enttäuscht ist und erklären, dass man lieber etwas anderes machen möchte. Aber man wird es den Kollegen nicht übel nehmen dürfen, wenn sie nicht einsehen wollen, welche Relevanz dieses neue Projekt für *ihre* erkenntnistheoretischen Fragen hat. Wenn man aber behaupten will, dass das neue Projekt auch für die alten Fragen Konsequenzen hat (wie beispielsweise Bishop und Trout [26]), dann sollte man das auch nachweisen können<sup>379</sup>. Einen solchen Nachweis erbringt man zu einem Gutteil aber dadurch, dass man durch Bedeutungsanalyse zeigt, dass die alten Fragen in einem logischen Bezug zu den neuen Antworten stehen.

In diesem Kapitel werden wir diesem Gedankengang noch weiter folgen. Wir werden dabei wieder unsere systematischen Erläuterungen an einem Beispiel aus dem Bereich der Naturalisierung philosophischer Fragen vornehmen und die Beziehung zwischen der Philosophie des Geistes und der empirischen Psychologie unter die Lupe nehmen.<sup>380</sup> Dabei wird uns dieses Kapitel zur Zusammenfassung eines Standpunktes dienen, der Gedankenexperimenten gegen Realdefinitionen eine – nach allem bisher Gesagten – unproblematische Funktion zuerkennt und auch erläutert, welche Rolle Gedankenexperimente bei der Diskussion von Begriffsexplikationen spielen. Am Ende dieses Kapitels werden wir darauf zurückkommen müssen, welche philosophische Positionen tatsächlich mit dem Gedankenexperiment „nichts anfangen“ können, und darauf, wie man *damit* umgehen sollte.

---

<sup>379</sup> Natürlich kann man mit solchen Behauptungen auch durch Glück richtig liegen. Ein nicht uninteressantes historisches Beispiel ist die syntaktische Informationstheorie nach Shannon und Weaver, von der Weaver schon sehr früh und völlig ahnungslos behauptet hatte, sie sei von Relevanz für die Entwicklung einer semantischen Informationstheorie. Dass sie dies in der Tat ist – wenn auch nicht so, wie Weaver das dachte – ist erst von Dretske gezeigt worden. Vgl. Bremer und Cohnitz [34].

<sup>380</sup> Die „Naturalisierung“ besteht darin, dass die empirische Psychologie behauptet, bestimmte ehemals „philosophische“ Fragen (wie etwa die, was eine Überzeugung ist oder was den Gehalt einer Überzeugung bestimmt) mit empirischen Mitteln beantworten zu können.

## 9.1 PHILOSOPHIE VS. PSYCHOLOGIE

Innerhalb der Kognitionswissenschaft werden – so zumindest die Auffassung des Naturalismus – altehrwürdige philosophische Probleme einer naturwissenschaftlichen Beantwortung zugeführt. Welche Rolle spielen dabei noch Gedankenexperimente?

### 9.1.1 DAS GEDANKENEXPERIMENT ALS EINSICHT IN NATURZUSAMMENHÄNGE

Wenn wir uns an Machs, Kuhns und Sorensens Auffassung vom physikalischen Gedankenexperiment erinnern, war es diesen Philosophen gelungen, eine Rolle für das Gedankenexperiment zu bestimmen, die in nachvollziehbarer Weise mit naturwissenschaftlichen Forschungsmethoden verbunden war und auch unter relativ klaren methodologischen Beschränkungen stand. In Bezug auf empirische Phänomene, mit denen wir in unserem Alltag (bzw. im Alltag unserer Ur-Ur-Urahnen) häufig zu tun haben (bzw. unsere Ur-Ur-Urahnen häufig zu tun hatten), können wir (in gewissem Rahmen) zuverlässig intuitive Auskünfte geben. Dies kann sehr hilfreich sein, solange eine Wissenschaft sich noch in einem Frühstadium befindet, in dem nur wenige empirische Daten über den Gegenstandsbereich gesammelt vorliegen und davon ausgegangen werden kann, dass ein signifikanter Teil des Gegenstandsbereichs dieser Wissenschaft uns tatsächlich bekannt ist. In diesem Fall sind die Ergebnisse von Gedankenexperimenten nicht nur von heuristischem Interesse, sondern *rechtfertigen* auch die durch sie initiierten Überzeugungen. Es handelt sich dann um *prima facie gerechtfertigte Anfangshypothesen*. Im Fall von Galilei und Stevin haben wir solche Gedankenexperimente kennen gelernt. Diese Rolle kann auch für die Psychologie zugestanden werden:

Perhaps the most obvious application of intuitions to the empirical sciences in general, and those of philosophical intuitions to psychology, in particular, comes when we use our intuitions as a source of empirical hypotheses. We might think of intuitive judgments as particularly plausible initial hypotheses about the nature of the world. [...] Like all such hypotheses, our intuitions are accurate as a guide to the world to different degrees, depending on both the person who has those intuitions and the subject area to which they apply. [...] We are all experts, in a pragmatic sort of way, in everyday psychology, and for this reason our everyday psychological intuitions can be taken as plausible initial hypotheses about how the mind really works. Empirical psychology can then use these everyday intuitions as a starting point for its research. (Gopnik und Schwitzgebel [129], 78-79)

Die Intuition stattet uns also nicht bloß heuristisch mit Anfangshypothesen aus, wenn „Heuristik“ tatsächlich nur den Entdeckungszusammenhang und nicht den

Rechtfertigungszusammenhang einer Hypothese betrifft.<sup>381</sup> Sofern die Intuitionen der richtigen (relevanten) Personen in den richtigen Kontexten zu Hypothesen führen, sind diese Hypothesen aufgrund ihrer Herkunft auch mit rechtfertigender Anfangsplausibilität ausgestattet. Auch Kornblith hatte der Methode des Gedankenexperiments diese Funktion für die naturalisierte Erkenntnistheorie zugestanden:

[W]hy do philosophers spend so much time scrutinizing their intuitions, that is, looking inward, if, on my view, what they are really interested in is external phenomena? [I]f I am asked a question about rocks, for example, one way to answer the question is to ask myself what I believe what the answer is. Although I am asked a question about rocks, I answer it by enquiring into what I believe. This is a perfectly reasonable thing to do if I have good reason to think that my current beliefs are accurate, or if I do not have access to a better source of information. By looking inward I answer a question about an external phenomenon. This, to my mind, is what we do when we consult our intuitions. (Kornblith [167], 14-15)

Welche Adäquatheitsbedingungen hier gelten, ist durch Mach und Kuhn erläutert worden (Kapitel 2.1, 2.3). Unsere Intuitionen sind umso zuverlässiger, desto näher der fragliche Sachverhalt, der im Gedankenexperiment zur Debatte steht, an unseren vergangenen Erfahrungen ist. Zugleich ist aber auch klar, dass diese intuitiven Urteile über den Gegenstandsbereich mit der fortschreitenden Entwicklung der empirischen Wissenschaft von diesem Bereich in den Hintergrund treten müssen.

The intuitions articulated by philosophers, then, in so far they are treated as initial hypotheses about the way the world really is, can guide empirical research, especially in those domains in which human beings have particular expertise. [...] As sciences mature, however, they typically revise, alter, and sometimes entirely reject, these initial hypotheses. In Galileo's day, it was permissible for physicists in defending one view or another to appeal to what we would now call our "folk physical" intuitions about what would happen in various circumstances. In contemporary physics, such appeals would be ruled out. Similarly, as psychological science has matured, we have become more confident in leaving intuitions behind, when they conflict with well-supported empirical findings [...]. Our initial hypotheses may be, indeed often are, moderated or defeated by later evidence. So although we may justifiably take reflective psychological intuition as a good preliminary guide, we no longer take it as the final authority about the mind. (Gopnik und Schwitzgebel [129], 79)

Unsere physikalischen Intuitionen können uns heute nicht mehr helfen, wenn es darum geht, Theorienkonflikte zwischen rivalisierenden physikalischen Theorien

---

<sup>381</sup> Wenn es in der Heuristik darum geht, „gute“ Hypothesen zu finden, dann kann die Herkunft einer Hypothese sie schon vor anderen Hypothesen, die auf anderem Wege zustande gekommen sind, *prima facie* rechtfertigen. Heuristik wäre dann Bestandteil des Rechtfertigungszusammenhangs. Wenn man streng zwischen Entdeckungs- und Rechtfertigungszusammenhang trennt und die Heuristik dem Entdeckungszusammenhang zuordnet, wird man dies nicht zugeben wollen.

zu entscheiden. Selbst in Bezug auf den Mesokosmos sind sie auf enttäuschend stabile Weise unzuverlässig, da wir intuitiv einer Impetus-Theorie anzuhängen scheinen, deren Inadäquatheit uns eigentlich aus dem Physikunterricht bereits bekannt sein sollte. Ein ähnliches Schicksal mag unseren Intuitionen in Bezug auf den Gegenstandsbereich der Psychologie beschieden sein. Intuitionen, die sich auf empirische Sachverhalte beziehen, müssen empirischen Untersuchungen standhalten oder werden zurückgewiesen.

### 9.1.2 GEDANKENEXPERIMENTE ALS DATEN FÜR DIE PSYCHOLOGIE

Im Sinne des letzten Unterkapitels handelt es sich bei Gedankenexperimenten zur Einsicht in Naturzusammenhänge um Gedankenexperimente, die zur Überzeugungsänderung in Bezug auf empirische Hypothesen gedacht sind, oder in Bezug auf Realdefinitionen im Sinne von Sachanalysen.

Realdefinitionen im Sinne von Bedeutungsanalysen stehen in einer anderen Beziehung zum Gedankenexperiment. Zunächst kann man Bedeutungsanalysen als sprachwissenschaftliche Untersuchungen der Bedeutung bestimmter Ausdrücke in einer – mehr oder weniger weit gefassten – Sprechergemeinschaft auffassen. Unser Urteil in Bezug auf imaginäre Fälle im Gedankenexperiment wäre dann nicht – wie im Sinne Fodors aus Kapitel 5.1 – eine empirische Hypothese/Prognose über unser tatsächliches Sprachverhalten in kontrafaktischen Umständen, sondern ein simples Datum unseres tatsächlichen Sprachverhaltens: wir verwenden den Ausdruck X in Bezug auf kontrafaktische Umstände eben so. Sinn und Zweck von Gedankenexperimenten dieser Art wäre es, Daten bereitzustellen, die von der (jeweils relevanten) Theorie über die Bedeutung von X in dieser Sprechergemeinschaft zu erklären wären. Diese Daten könnten – wie bei anderen Datenerhebungsverfahren auch – in Bezug auf ihr Zustandekommen kritisiert werden. Soll die Sprechergemeinschaft größer sein als die der westlichen Philosophen und Philosophinnen, sollte man die Datenerhebung unter restringierteren Bedingungen durchführen. Man müsste empirische Untersuchungen im Stil von Machery et al. durchführen. Es sollte dann ausgeschlossen werden, dass die intuitiven Urteile durch theoretische Überzeugungen „verunreinigt“ sind, kultur- und einkommensabhängig sind, etc. Falls die Ergebnisse eines Gedankenexperiments dann einer Theorie widersprechen, deren Gegenstandsbereich diese Daten enthält, würde die Theorie falsifiziert. Dies gilt nicht nur für unsere intuitiven Urteile über Grammatikalität und die Bedeutung von Ausdrücken in kontrafaktischen Zusammenhängen, sondern auch in Bezug auf alle möglichen Überzeugungen, die wir sprachlich zum Ausdruck bringen und die von psychologischen Theorien zu erklären sind. So ist unsere intuitive Alltagsphysik zwar irrelevant für die moderne Physik, aber relevant für die Psychologie. Eine psychologische Theorie (wie die von Nancy Nersessian, vgl. Kapitel 3.1), die sowohl annimmt, dass es nicht einzelne Überzeugungen, sondern kognitive Mechanismen sind, die durch die Evolution geformt werden, wie auch, dass der kognitive Mechanismus des *Mental Modelling*

ein zuverlässiger Mechanismus zur Beurteilung physikalischer Sachverhalte der Alltagswelt ist, wird z.B. durch die intuitiven physikalischen Urteile von Physikabsolventen widerlegt. Gedankenexperimente zu physikalischen Zusammenhängen sind hierbei relevant, weil sie in den Gegenstandsbereich der Theorie fallen:

When we consider [intuitions] as evidence, intuitions have some of the quality of indefeasibility that evidence always has. Parallel lines may, or may not, meet, and we may, or may not, have first-person access to our mental states, but it is indubitably true that we believe that parallel lines will not meet and that we believe that we have first person access to our own mental states. A psychological theory has to explain why we have these intuitive beliefs even if, indeed especially if, the beliefs are quite false. The psychologist cannot simply reject the beliefs if she is treating them as evidence about the mind of the person who has them. (Gopnik und Schwitzgebel [129], 80)

Auch diese Position ist von einem naturalistischen Blickwinkel betrachtet unproblematisch. Welche Adäquatheitsbedingungen für das Gedankenexperiment gelten, hängt davon ab, wie der Gegenstandsbereich der fraglichen Theorie bestimmt ist. Wenn die Theorie Prognosen darüber macht, wie ein gut unterrichteter, geistig normaler Erwachsener bei mentalen Modellierungsaufgaben abschneiden sollte, braucht das abweichende Abschneiden von in ihren mentalen Fähigkeiten eingeschränkten Minderjährigen die Theorie nicht in Gefahr zu bringen. Dasselbe gilt für die Frage, ob man kulturrelevante Unterschiede u.Ä. in das Untersuchungsdesign aufnehmen muss, oder ob ein paar ruhige Minuten im eigenen Ohrensessel ausreichen. Hier gelten zunächst einfach die Standardbedingungen für psychologische Experimente.

### 9.1.3 GEDANKENEXPERIMENTE ZUR EXPLIKATION THEORETISCHER ZUSAMMENHÄNGE

In den beiden letzten Punkten haben wir Gedankenexperimente den beiden Typen von Realdefinitionen zugeordnet, die wir in Kapitel 7 erläutert haben. Wenn wir uns nun fragen, welche Rolle Gedankenexperimente bei der Explikation spielen, müssen wir zwei verschiedene Rollen unterscheiden, von denen man die eine als „interne“ kritische Rolle von Gedankenexperimenten bezeichnen könnte, die andere als „externe“ kritische Rolle. Die Übergänge sind in fast allen Fällen der Gegenwartsphilosophie fließend, an dem hier gewählten Beispiel Philosophie versus Psychologie kann der Unterschied aber erläutert werden.

Wenn wir einmal von sehr vielen komplizierenden Faktoren absehen, kann man Gedankenexperimente auch dazu verwenden, die interne Kohärenz und Struktur von Theorien zu untersuchen. Indem man imaginäre Fälle entwirft, die in den intendierten Anwendungsbereich der Theorie fallen und die Begrifflichkeiten der Theorie auf die intendierte Weise anwendet, kann man zu nicht-intendierten Ergebnissen kommen. Es kann sein, dass die Vertreter einer fraglichen Theorie

bestimmte Sachverhalte anders beschreiben möchten, als die begrifflichen und theoretischen Festlegungen, die sie in ihrer Theorie getroffen haben, dies zulassen. Das eine Extrem bestünde in der Aufdeckung interner Widersprüche in einer Theorie. Russells Antinomie wäre so ein Fall, bei dem an einem bestimmten Beispiel (der Russell-Menge) gezeigt wird, dass bei einer Anwendung der vorgenommenen theoretischen Festlegungen auf diesen Fall ein Widerspruch abgeleitet werden kann. Das andere Extrem bestünde darin, dass man an Beispielen aus dem intendierten Anwendungsbereich zeigt, dass eine konsistente Anwendung der Theorie auf diesen Fall es nicht erlaubt, den Fall so zu beschreiben, wie man es eigentlich wollte (etwa weil die begrifflichen Festlegung in der Theorie zu grobkörnig sind und die Unterscheidung, die man treffen wollte, gar nicht zulassen – man denke an die Beispiele aus Kapitel 2.3)). Die Adäquatheitsbedingungen für diesen Gebrauch sind von Popper und Kuhn erläutert worden: damit ein Gedankenexperiment in diesem Sinne funktioniert, muss es einen Fall präsentieren, der im intendierten Anwendungsbereich der Theorie liegt, und dieser Fall muss mit den fraglichen theoretischen Begrifflichkeiten im intendierten Sinne beschrieben werden.

Naturalisten sind oft der Meinung, dass mit den soeben aufgeführten Fällen (Gedankenexperimente zur Entwicklung von Anfangshypothesen, Gedankenexperimente zur Bereitstellung von Daten für die Psychologie, Gedankenexperimente zur internen Kritik) alle hauptsächlich relevanten Anwendungen des Gedankenexperiments abgedeckt sind, und dass diejenigen Fälle, die uns philosophisch interessanter erscheinen, nur darauf beruhen können, dass man diese Funktionen durcheinander wirft:

Within the philosophy of mind, there has been a tradition of either accidentally confusing different uses of intuition, or deliberately treating them as identical, and consequently drawing inappropriate conclusions. An unspoken assumption of much argumentation in the philosophy of mind has been that to articulate our folk psychological intuitions, our ordinary concepts of belief, truth, meaning, and so forth, is itself sufficient to give a theoretical account of what belief, truth, meaning, and so forth, actually are. We believe that this assumption rests on an inadequate understanding of the nature of intuition and its appropriate applications, and that it results in errors [...]. (Gopnik und Schwitzgebel [129], 84)

Alison Gopnik und Eric Schwitzgebel vermuten hier, dass Philosophen mit Intuitionen darüber, was sie glauben, was Bedeutung, Wissen, Überzeugung, etc. sind, psychologische Theorien zu kritisieren versuchen. Dies sei aber eine unzulässige Verwendung des Gedankenexperiments. Da die Psychologie aus der Phase herausgetreten ist, in der es noch von Interesse war, zu erforschen, was man denn so über den Gegenstandsbereich für Meinungen hat, sind philosophische Intuitionen heute bestenfalls als empirisches Datum zu nehmen, das ggf. von einer Theorie erklärt werden müsste. Das muss aber nicht notwendigerweise durch dieselbe Theorie geleistet werden, die die Philosophen mit ihren Intuitionen angreifen.<sup>382</sup> Wenn

---

<sup>382</sup> Ähnliche Argumente findet man bei Cummins [81] und Chomsky [64], Chomsky [65].

Psychologen der Meinung sind, dass sie mit einem anderen Wissensbegriff oder einem anderen als dem „supra-individuellen“, externalistischen Bedeutungsbegriff besser klar kommen, kann es keinen Grund geben, diese Verwendung aus philosophischer Sicht zu kritisieren:

The fact that the ordinary concept of meaning within our daily experience has [the character of a supra-individual concept] provides only rudimentary evidence about what notion of meaning will work best for psychologists and others who wish to talk about meaning in a scientifically informed way—just as our ordinary intuitions about physics provide only a rudimentary starting point for the creation of physical laws. These intuitions may have played a crucial role for Galileo, but most of them did not survive in the end. (Gopnik und Schwitzgebel [129], 85)

Haben die beiden nicht Recht? Ist es uns nicht manchmal peinlich, darauf hingewiesen zu werden, mit welcher Inbrunst sich Kant für bestimmte synthetische Urteile a priori eingesetzt hat, nur um dann durch die Physikgeschichte widerlegt zu werden? Wollen wir das etwa auch riskieren, indem wir das junge Pflänzlein einer interdisziplinären Kognitionswissenschaft dadurch riskieren, dass wir den fleißigen Psychologen mit ausgedachten Experimenten „nachweisen“, sie redeten gar nicht über ‚Wissen‘ und ‚Bedeutung‘?

Was, wenn die Kinder aus Piagets Untersuchung, die wir in Kapitel 2.3.2 kennen gelernt haben, so auf ihrem defizitären Geschwindigkeitsbegriff beharrten, wie Philosophen auf ihrem Begriff von ‚Überzeugung‘ oder ‚Bedeutung‘. Wir würden sie sicherlich zu einem Umdenken überreden wollen. Wir würden ihnen sagen wollen, dass ihr Geschwindigkeitsbegriff für eine empirische Untersuchung der Welt eben inadäquat ist und sie sich besser den der empirischen Wissenschaft aneignen sollten. Dass dieser Begriff „ihrem“ Begriff nicht entspricht und „ihre“ Intuitionen verletzt, würde uns dabei kaum als ein gutes Argument erscheinen (vgl. auch Gopnik und Schwitzgebel [129] mit einem ähnlichen Argument). Es hat in wissenschaftlichen Zusammenhängen oft keinen Wert, sich an den Alltagsgebrauch eines Begriffs zu klammern, wenn er in einer anderen Bedeutung fruchtbarer verwendet werden kann.

Wenn Ausdrücke, die in der Alltagssprache bereits vorkommen, in der Wissenschaftssprache mit einer neuen Bedeutung versehen werden, hat dies offensichtlich etwas mit *Begriffsexplikationen* zu tun. Aus Kapitel 6.3 wissen wir, dass die Güte von Begriffsexplikationen *zum Teil* davon abhängt, wie eng das Explicatum am Explicandum ist. Carnap hatte sich in Bezug auf das Verhältnis von Explikation und herkömmlicher Begriffsverwendung recht vorsichtig ausgedrückt, da er sich der Tatsache bewusst war, dass theoretische Fruchtbarkeit letztlich viel interessanter ist als die Nähe zur Umgangssprache. Die umgangssprachliche Verwendung – so die Überlegung Carnaps – war ja ungenau und revisionsbedürftig. Die „neue“ Bedeutung *sollte* ja gerade abweichen, um diese Ungenauigkeit zu vermeiden.

Wie eng bzw. wie weit die Bedeutung des Explicatums vom Explicandum abweicht, ist eine Frage der Bedeutungsanalyse und deskriptiv. Sie wird dadurch beantwortet, dass man feststellt, wie ein bestimmter Begriff *de facto* verwendet wird. Dass man dabei feststellt, dass die Bedeutung des Explicatums abweicht, oder

stark abweicht, ist gemessen an den Adäquatheitsbedingungen alleine unerheblich. Wie wir gesagt haben, sind Explikationen ein *normatives* Unterfangen. Was wir bisher außer Acht gelassen haben, sind die normativen Beschränkungen, die auf die Festlegung des Explicandums wirken. Carnap hatte ganz richtig festgestellt, dass der zu explizierende Begriff unklar und vage und problematisch ist, man also bei der Formulierung der Aufgabenstellung für die Explikation nicht recht sagen kann, was da eigentlich expliziert werden soll:

A problem of explication is characteristically different from ordinary scientific (logical or empirical) problems, where both the datum and the solution are, under favorable conditions, formulated in exact terms. [...] In a problem of explication the datum, viz., the explicandum, is not given in exact terms; if it were, no explication would be necessary. Since the datum is inexact, the problem itself is not stated in exact terms; and yet we are asked to give an exact solution. This is one of the puzzling peculiarities of explication. (Carnap [50], 4)

Die Aufgabenstellung für eine Explikation kann also nicht genau formuliert werden. Ist es aber nicht ohnehin unerheblich, wie man die Aufgabenstellung genau definiert? Immerhin gibt es doch Adäquatheitsbedingungen (Nähe zur Umgangssprache, Fruchtbarkeit, Einfachheit), die im Nachhinein eine Beurteilung des Explicatums zulassen. Man sollte meinen, es spiele keine Rolle, wie das Problem der Explikation genau zu formulieren ist.

On the contrary, since even in the best case we cannot reach full exactness, we must, in order to prevent the discussion of the problem from becoming entirely futile, do all we can to make at least practically clear what is meant as the explicandum. What X means by a certain term in contexts of a certain kind is at least practically clear to Y if Y is able to predict correctly X's interpretation for most of the simple, ordinary cases of the use of the term in those contexts. It seems to me that, in raising problems of analysis or explication, philosophers very frequently violate this requirement. They ask questions like: 'What is causality?', 'What is life?', 'What is mind?', 'What is justice?', etc. Then they often immediately start to look for an answer without first examining the tacit assumption that the terms of the question are at least practically clear enough to serve as the basis for an investigation, for an analysis or explication. (Carnap [50], 4)

Auch bei Begriffsexplikationen ist „alles“ dafür zu tun, dass wir uns wechselseitig darüber „zumindest praktisch verständlich“ machen, wie wir das Explicandum meinen wollen. Ohne eine solche Verständigung wären philosophische Explikationsversuche „absolut zwecklos“. Carnap gibt uns auch einen Hinweis darauf, wie eine solche Verständigung erzielt werden kann:

Even though the terms in question are unsystematic, inexact terms, there are means for reaching a relatively good mutual understanding as to their intended meaning. An indication of the meaning with the help of examples for its intended use and other examples for uses not now intended can help the understanding. An informal explanation in general terms may be added. All explanations of this kind serve only to make clear what is meant as the explicandum; they do not yet supply an expli-

tion; say a definition of the explicatum; they belong still to the formulation of the problem, not yet to the construction of an answer. [...] By explanations of this kind the reader may obtain step by step a clearer picture of what is intended to be included and what is intended to be excluded; thus he may reach an understanding of the meaning intended which is far from perfect theoretically but may be sufficient for the practical purposes of a discussion of possible explanations. (Carnap [50], 5)

Zur wechselseitigen Verständigung über das Explicandum werden also Fälle genannt, in denen man das Explicatum benutzen will, und Fälle, auf die es nicht zutreffen soll. Man gibt also *Kontexte* an, in denen der fragliche Ausdruck in einer wahren Beschreibung vorkommt, sowie Fälle, in denen der Ausdruck nicht Bestandteil einer wahren Beschreibung sein darf. Die Aufgabe der Explikation ist es dann, ein Explicatum zu finden, das die Wahrheitswerte dieser Beschreibungen unverändert lässt, wenn man in ihnen das Explicandum durch das Explicatum ersetzt. In jeder anderen Hinsicht mag das Explicatum abweichen. Die vorher festgelegten Kontexte bestimmen, was man (mindestens) expliziert haben will. Die Anforderung, dass diese Kontexte ihre Wahrheitsbewertungen behalten, wenn man in ihnen das Explicandum durch das Explicatum ersetzt, gehört dann ebenfalls zu den Adäquatheitsbedingungen für ein Explicatum.

Als Beispiele für solche Kontexte nennt Carnap isolierte Sätze, in denen der Explicandum-Ausdruck vorkommt. Wir sehen aber jetzt, dass in der Philosophie nicht nur isolierte Sätze zur Verständigung über das Explicandum herangezogen werden sondern durchaus reichere Kontexte: Gedankenexperimente sind ebenfalls solche Kontexte. Die Intuitionen, die wir in Bezug auf Gedankenexperimente haben und in der philosophischen Diskussion präsentieren, sind vornehmlich solche, von denen wir der Meinung sind, dass sie als Adäquatheitsbedingungen für Explikationen dienen *sollen*, und wir gehen implizit davon aus, dass diese Auffassung in der wissenschaftlichen Gemeinschaft geteilt wird. Idealerweise erwarten wir, dass jede zukünftige Begriffsexplikation die intuitive Beurteilung des Gedankenexperiment-Falls erhalten wird.

**Definition 9.1-1 (Adäquatheitskontext):** Ein Gedankenexperiment  $G$  (im Sinne von  $I$ , Kap. 3.2.5) ist ein Adäquatheitskontext für die Explikation eines Explicandum-Ausdrucks  $B$  aus  $G$  gdw. jede adäquate Explikation  $E$  von  $B$  den Wahrheitswert aller Aussagen in  $G$  erhält, wenn  $E$  in  $G$  für  $B$  ersetzt wird.

Was dies in der Praxis dann bedeutet, kann man an vielen Beispielen sehen.

So führt – beispielsweise – Dretske in *Knowledge and the Flow of Information* (Dretske [92]) eine Definition des Wissensbegriffs ein, um dann an klassischen Gedankenexperimenten gegen frühere Explikationen nachzuweisen (wie – beispielsweise – den Gettier-Fällen), dass seine Explikation in obigem Sinne die konventionelle Bewertung dieser Fälle reproduziert. David Chalmers leistet dasselbe in ‚The Components of Content‘ (Chalmers [57]) für verschiedene Gedankenexperimente aus dem Bereich der Semantik propositionaler Einstellungen, Robert

Nozick argumentiert in *Philosophical Explanations* (Nozick [243]) für die Adäquatheit seiner *closest-continuer theory*, indem er verschiedene Gedankenexperimente aus der Debatte um transtemporale personale Identität durchgeht und zeigt, dass seine Theorie in der Anwendung auf Gedankenexperimente aus der Tradition zu den erwünschten Resultaten kommt. Diese Liste könnte man beliebig verlängern. Interessant ist, dass diese Gedankenexperimente als Adäquatheitskontexte *allgemein* akzeptiert werden.

Ähnliche Rollen kann man auch für die Gedankenexperimente in der Praktischen Philosophie annehmen. Es handelt sich um Fallbeschreibungen, die durch die gesuchten ethischen Prinzipien auf die Weise zu beurteilen sind, wie wir uns über diese Fälle verständigt haben. Gedankenexperimente in der Ethik setzen keinen Kognitivismus voraus, so dass man sie nur dann sinnvoll verwenden könnte, wenn man annimmt, dass es so etwas gibt wie objektive, unabhängige ethische Richtigkeit, die wir mit unseren intuitiven Urteilen aufzuspüren imstande sind. Natürlich kann man auch die Gedankenexperimente in der praktischen Philosophie als Adäquatheitskontexte auffassen, die zukünftige normative Regelungen so zu bewerten haben, wie wir es für diese Kontexte festgelegt haben.

Man einigt sich also bei Gedankenexperimenten als Adäquatheitskontexten darauf, dass jede zukünftige Explikation etwas zu diesem Kontext sagen muss, idealer Weise muss jede zukünftige Explikation bezüglich dieses Kontextes zu den verabredeten Beurteilungen kommen.

Letzteres kann – und jetzt schlagen wir die Brücke wieder zurück zur Psychologie – in Fällen, in denen man gezwungen ist von der verabredeten Beurteilung abzuweichen, aber auch erfüllt werden, indem man den Adäquatheitskontext auf andere Weise *wegerklärt*. Im Fall von Frank Jacksons Gedankenexperiment mit Mary bestanden solche Strategien darin, dass man versuchte, eine Theorie zu finden, derzufolge es sich nicht um einen eigentlichen Adäquatheitskontext für den fraglichen Gegenstandsbereich der Theorie handelt, etwa weil es in diesem Kontext um *knowing how*, nicht um *knowing that* geht, oder weil es in diesem Kontext nur um einen anderen Aspekt einer bekannten Tatsache geht, also eigentlich um einen Kontext für Theorien kognitiven Gehalts, aber nicht für Theorien metaphysischer Zusammenhänge zwischen Körper und Geist. Konservativität in Bezug auf den Adäquatheitskontext wird dann dadurch hergestellt, dass man die Theorie, die *prima facie* an dem Kontext scheitert, um eine unabhängige aber kompatible Theorie erweitert, die die ursprüngliche Beurteilung wieder herstellt. Für diese Zusatztheorie gelten selbstverständlich ebenfalls Adäquatheitskontexte, deren Erfüllung unabhängig sichergestellt werden muss.

An die Stelle philosophischer Zusatztheorien können aber auch psychologische Theorien treten, wie – beispielsweise – bei Alvin Goldman, der eine relevante-Alternativen Explikation des Wissensbegriffs verteidigt und die verabredete Beurteilung von Gedankenexperimenten gegen diese Theorie dadurch rekonstruiert, dass er eine psychologische Theorie angibt, die für die fraglichen Kontexte *prognostiziert*, dass es zu von der (relevante-Alternativen) Theorie abweichenden Urteilen kommt:

In general, it looks as if uniformities about knowledge could be predicted from the psychology of context effects together with the relevant-alternatives account of 'know'. This illustrates how empirical psychology might lend credibility (or incredibility) to specific philosophical analyses. (Goldman [126], 147)

Ähnliches gilt für John Perrys Argumentation in ‚The Importance of Being Identical‘ (Perry [256]). Wie alle Mentalisten muss auch Perry in seiner Explikation diachroner Personenidentität mit Butlers berühmten Herausforderung fertig werden: Warum haben wir so ein besonderes Interesse an unserer eigenen Zukunft und unserem eigenen zukünftigen Wohlergehen, wenn unser Überleben nur darin besteht, dass in Zukunft eine Person existieren wird, die mit uns jetzt in bestimmten abstrakten Beziehungen steht? Wie kann unser intrinsisches Interesse von dieser Beziehung bloß *abgeleitet* sein? Die mit dieser Frage verbundenen Adäquatheitskontexte werden durch Perrys Theorie verletzt, seine Lösung besteht darin, die Interessen, die in solchen Kontexten zum Ausdruck kommen, durch ein soziobiologisches Modell zu erklären, demzufolge diese Interessen unabhängig von intrinsischen Fragen personaler Identität durch den kontingenten Prozess der Evolution mit diesen verbunden wurden (vgl. Perry [256], Cohnitz [70]).

Wenn wir diese Überlegungen auf den Streit zwischen Psychologen und Philosophen zurückübertragen, macht das Verhalten der Philosophen durchaus Sinn. Philosophen versuchen darauf einzuwirken, welche Frage sie von den Psychologen geklärt und ggf. empirisch untersucht wissen wollen. Indem festgestellt wird, dass psychologische Theorien von ‚Bedeutung‘, ‚Wissen‘, etc. „Intuitionen“ verletzen, soll nicht eigentlich behauptet werden, dass man intuitiv die Dinge besser beurteilen kann als derjenige, der empirisch nachgesehen hat, sondern dass die vorgelegte Antwort bzw. das anvisierte Forschungsprogramm an der eigentlich interessierenden Fragestellung vorbeigeht. Das Explicandum ist eben anders intendiert als die Psychologen meinen.

## 9.2 ARBEITSTEILUNG IN DER PHILOSOPHIE

Die am Anfang von Kapitel 5.3 zitierte Äußerung Moritz Schlicks, in der die Hoffnung zum Ausdruck kam, die „neue“ Philosophie werde weniger als hermetisches Großsystem als vielmehr als arbeitsteiliges Gemeinschaftsunternehmen betrieben werden, hat natürlich einen wahren Kern. Die heutige Philosophie *ist* arbeitsteilig. Wir kennen systematische Teilbereiche, in denen unterschiedliche Theorien miteinander konkurrieren, die auf eben jene Teilbereiche zugeschnitten sind. Im vergangenen Zeitalter der Großsysteme konnte sich die interne Kritik von Gedankenexperimenten immer nur auf ihre theoretische Rolle beschränken: es konnte bestenfalls gezeigt werden, dass (gegeben die Intentionen ihres jeweiligen Autors) die Theorie inkonsistent war, man konnte aber nicht sinnvoll mit Adäquatheitskontexten an sie herantreten, die von ihrem jeweiligen Autor nicht eindeutig abgesegnet waren. Jedes hermetische System kann dadurch gegen Kritik immunisiert werden, dass man die problematischen Adäquatheitskontexte von der

Theorienbewertung einfach ausschließt. Damit die moderne Form philosophischer Arbeitsteilung möglich ist, einigt man sich über die Diskussion von Gedankenexperimenten auf die relevanten Adäquatheitskontexte, innerhalb derer zunächst Begriffe expliziert werden sollen, die dann ggf. das Grundgerüst empirischer Theorien bilden. Diese Adäquatheitskontexte sind ausschlaggebend dafür, welche Richtung die gemeinsamen Forschungsanstrengungen in Zukunft nehmen sollen, entsprechend heiß sind sie auch diskutiert. Diese Diskussionen dürfen gerade nicht so missverstanden werden wie in Kapitel 5.3, nämlich als Ausdruck der unüberbrückbaren Unvereinbarkeit verschiedener Positionen der Gegenwartsphilosophie. Über diese „Streitigkeiten“ einigen sich die Wissenschaftler auf das gemeinsame Verständnis des jeweiligen Forschungsprogramms.

Natürlich bemisst sich die Güte eines Gedankenexperiments zunächst daran, ob es mittelbar zu den intendierten Überzeugungsänderungen führt. D.h., ob es mittelbar dazu geeignet ist, die inhaltliche Richtung eines Forschungsprogramms zu beeinflussen. Dies ist sicherlich die Hauptintention desjenigen, der ein Gedankenexperiment vorträgt: Er möchte zum Ausdruck bringen, dass der gegenwärtige Stand der Explikation oder die Richtung, in der nach einer gesucht wird, Kontexte außer Acht lässt, von denen wir alle wollen sollen, dass sie durch die Explikation bedacht werden.<sup>383</sup>

Langfristig ist ein Gedankenexperiment gut, wenn es in den Kanon der dauerhaften Adäquatheitskontexte aufgenommen wird. Gedankenexperimente, die von der Forschergemeinde akzeptiert werden, wandeln sich zu Standards, die die Grundlage der weiteren Arbeit bilden. Das ändert sich nicht dadurch, dass ein Adäquatheitskontext durch eine Zusatztheorie  $T_2$  als ein Kontext erwiesen werden kann, der nicht zum Gegenstandsbereich von  $T_1$ , der eigentlich attackierten Theorie gehört. Der Adäquatheitskontext bleibt auch dann „virulent“, schließlich kann sich durch die weitere theoretische Entwicklung herausstellen, dass  $T_1$  und  $T_2$  inkompatibel sind, oder  $T_2$  erweist sich aus unabhängigen Gründen als falsch oder tatsächlich nicht auf den fraglichen Kontext zutreffend, etc. Wichtig ist, dass ein solcher Adäquatheitskontext dauerhaft als solcher akzeptiert wird.

Hier eröffnet sich eine interessante weitere Verbindung zur Rolle von Gedankenexperimenten in den Naturwissenschaften. Wie wir in Kapitel 3.2 gesehen haben, etablieren Galilei und Stevin durch die Spezialisierung auf einen unproblematischen Fall allgemeine Naturprinzipien, die zum Zeitpunkt des Gedankenexperiments noch nicht experimentell bestätigt werden konnten. Diese Naturprinzipien (Ausschluss des Perpetuum Mobile, beliebige Teilbarkeit des physikalischen Raums in physikalische Teile) blieben von da ab Adäquatheitsbedingungen für zukünftige Theorien<sup>384</sup>, die dann später auch eine empirische Fundierung be-

<sup>383</sup> Die historische Entwicklung solcher Adäquatheitskontexte entspricht den Phasen der Theorienentwicklung, wie sie von Lakatos [177], Lakatos [178] in Bezug auf die Mathematik dargestellt wurden. Es kann durchaus sein, dass Adäquatheitskontexte zunächst als „Monster“ empfunden werden und erst allmählich zu einem Adäquatheitskontext werden, der allgemein akzeptiert ist. Ein „Monster“ wäre ein Gedankenexperiment, das zunächst als „zu bizarr“ zurückgewiesen wird. Vgl. auch Feferman [102] für die Geschichte mathematischer „Monster“.

<sup>384</sup> Vgl. Kühne [174], Kühne [175], Kühne [176].

kamen. Diejenigen allgemeinen Prinzipien, die hinter den philosophischen Adäquatheitskontexten stehen, sind wesentlich weiter von der Frage empirischer Bestätigung entfernt als die, die in der Physik eine Rolle spielen. In beiden Fällen geht es darum, dass man Erklärungen bestimmter Art will, diese „bestimmte Art“ aber nicht besser angeben kann als dadurch, dass man sich auf eine Menge von Adäquatheitskontexten einigt.

Natürlich bedeutet das nicht, dass uns die Welt diese Erklärungswünsche auch erfüllen lässt. Wie in unserer Darstellung des EPR-Falles kann es sein, dass wir Adäquatheitsbedingungen fordern, die keine wahre Theorie des fraglichen Gegenstandsbereichs je erfüllen können wird, weil wir uns einfach über die Natur der empirischen Phänomene täuschen. Solange man das aber nicht gezeigt hat, darf man darauf insistieren, das erklärt zu bekommen, was man erklärungsbedürftig findet.

Soweit die Charakterisierung von Gedankenexperimenten und ihrer Funktion in Bezug auf Begriffsexplikationen. Für ihren Erfolg kommt es darauf an, dass sie durch die Philosophengemeinschaft akzeptiert werden. Dafür gelten verschiedene Bedingungen, die kurz genannt werden sollen:

Wenn wir uns das Häggqvist-Schema in Erinnerung rufen, geht es zunächst darum, ein Szenario (C) zu finden, das als mögliches Szenario betrachtet wird. Wie wir gesehen haben, ist es theorieabhängig, in welchem Sinne genau von „Möglichkeit“ gesprochen wird. Worauf es ankommt, ist, dass die Beschreibung des Szenarios keine Inkonsistenz enthält. Wir haben schon erläutert, welche formalen Schwierigkeiten es verhindern, diese Konsistenzbedingung auf algorithmische Weise zu prüfen. Aus diesem Grunde empfiehlt es sich, Szenarien zu wählen, die möglichst nahe an den vermuteten Umständen in der aktuellen Welt orientiert sind, um das Problem dadurch „modelltheoretisch“ zu lösen. Schließlich ist die aktuelle Welt in jeder relevanten Hinsicht eine mögliche Welt und kann daher zumindest als partielles Modell des Szenarios gelten.

Der methodologische Rat, bei der Konstruktion eines Gedankenexperiments ein Szenario zu wählen, das möglichst ähnlich zu den Umständen in der aktuellen Welt ist, ist häufig vorgebracht worden. In der Regel wird dieser Rat über ein Relevanzkriterium motiviert: Szenarien, die von der aktuellen Welt zu weit entfernt sind, sind nicht relevant für die Beurteilung philosophischer Fragen.

Solche Kriterien sind in der Regel überzogen, wie wir in Kapitel 7.1 gesehen haben. Der hier formulierte methodologische Rat ist auch nicht aus einem solchen Kriterium motiviert. Dieser Rat ergibt sich vielmehr aus den erkenntnistheoretischen Überlegungen, die in Kapitel 8 angestellt wurden.

Die Bedingung der Konsistenz des Adäquatheitskontextes (wie auch die nächste) stellen dabei sicher, dass die Übereinstimmung im Wahrheitswert bei Austausch von Explicandum und Explicatum nicht trivial ist. Wäre der Kontext für sich inkonsistent oder wäre er vom Explicatum logisch abhängig, würde eine Er-

setzung trivialer Weise nie zu abweichenden Wahrheitsbewertungen des Gesamtkontextes führen.<sup>385</sup>

Die Beschreibung des Szenarios sollte aus demselben Grund auf keinen Fall die Falschheit der anzugreifenden Theorie oder die Wahrheit einer rivalisierenden Alternative voraussetzen. Szenarien, die inkonsistent sind oder den Verdacht einer *petitio*<sup>386</sup> nahe legen, werden nicht akzeptiert

Das Szenario sollte in jedem Fall eindeutig in den Gegenstandsbereich der kritisierten Theorie fallen und von dieser möglichst eindeutig beurteilt werden. Genauso sollten diejenigen Umstände im Szenario, die uns zu einem intuitiv abweichenden Urteil führen, so herausstechend sein, dass das Scheitern der Theorie an diesem Fall offensichtlich ist. Bei dieser abweichenden Beurteilung muss es sich um eine Beurteilung handeln, die möglichst weit geteilt ist, mindestens innerhalb der relevanten Expertengruppe. (Dies erreicht man häufig dadurch, dass man Kontexte wählt, die denen stark ähneln, die die Experten in diesem Bereich bereits akzeptiert haben. Aus diesem Grund kann die Geschichte bestimmter Teile der Philosophie als Ausdifferenzierung bestimmter Gedankenexperimente rekonstruiert werden.)

Beim Übergang vom intuitiven Urteil zur Widerlegung der attackierten Theorie kommen die in Kapitel 4 herausgearbeiteten Hintergrundannahmen ins Spiel. Die Verbindung zwischen unserem intuitiven Urteil in Bezug auf irgendeinen Aspekt des imaginären Falls und der kritisierten Theorie muss so unproblematisch wie möglich gewählt werden. Philosophische Theorien begegnen uns mit zunehmender Reife des Forschungsgebiets in ausdifferenzierter und reflektierter Gestalt. Es ist unwahrscheinlich, dass sich im unmittelbaren Anwendungsgebiet der Theorie Fälle finden lassen, deren intuitive Beurteilung ohne „Hilfshypothese“ als Widerlegung aufgefasst werden kann.<sup>387</sup> Falls die Verbindung zwischen der intuitiven Beurteilung eines bestimmten Falls und der zu kritisierenden Theorie indirekter Natur ist, kommt es hauptsächlich darauf an, einen Fall zu finden, für den die Zwischenhypothese offensichtlich wahr ist. Bei zweifelhaften Zwischenhypothesen findet – wie wir gesehen haben – ansonsten eine Problemverschiebung statt.

Diese Problemverschiebungen sind nicht grundsätzlich negativ zu bewerten. Egal, welchen Ausgang sie nehmen, sie klären in jedem Fall die begrifflichen Zusammenhänge zwischen verschiedenen Teilbereichen der Philosophie. Da die systematischen Teilgebiete der Gegenwartsphilosophie auf Dauer Gefahr laufen würden, zu stark auseinander zu driften, sind solche Phasen, in denen die Diskussion scheinbar auf Nebengleise ausweicht, wichtig für den inhaltlichen Zusammenhalt der Philosophie.

---

<sup>385</sup> Welche logische Struktur Adäquatheitskontexte genau haben, ist zunächst unerheblich. Für unsere Zwecke ist es am Sinnvollsten, sich diese Kontexte im Sinne der im Häggqvist-Schema (Kapitel 3.2.3) aufgeführten kontrafaktischen Konditionale aus Prämisse (3.) vorzustellen. Diesen Prämissen gegenüber sollen die Targetthesen Wahrheitswert-konservativ sein.

<sup>386</sup> Leider kann man *petitio principii* kaum formal sauber explizieren.

<sup>387</sup> Die Gettier-Fälle bilden hier vielleicht eine besondere Ausnahme.

### 9.3 ALTERNATIVE KONZEPTIONEN

Mit dieser letzten Charakterisierung der Rolle des Gedankenexperiments in der Philosophie haben wir eine Funktion für das Gedankenexperiment gefunden, die auch für jene philosophische Positionen nachvollziehbar ist, die nicht der Auffassung sind, der Philosophie ginge es um die Aufdeckung metaphysischer Notwendigkeiten oder hauptsächlich um Fragen der Begriffsanalyse. Die Rolle des Gedankenexperiments zur Explikation findet sich ebenso in Quines Konzeption der philosophischen Erläuterung durch Paraphrase (*Word and Object*, Quine [268]). So stellt auch Jackson fest:

But now what we are doing [in conceptual analysis] is very like what Quine calls paraphrasing. As he puts it, 'The objective would not be synonymy, but just approximate fulfillment of likely purposes of the original sentences ...'. [W]hat guides us is very like what guides the Quinean who refuses to talk of synonymy, but seeks paraphrases that do the jobs that need doing. (Jackson [158], 45)

In anderen Worten: obwohl der modale Rationalismus Jacksons (Kapitel 5.4) mit seiner Methodologie weitergehende Ansprüche verfolgt als Explikation und Paraphrase, sind die bei der Explikation und die im modalen Rationalismus verwendeten Methoden, wie auch die dabei in Anschlag gebrachten methodologischen Standards dieselben. Das Gedankenexperiment in der Philosophie spielt auch unter naturalistischen und externalistischen Prämissen eine sinnvolle und wichtige Rolle bei der philosophischen Theorieentwicklung, die unter denselben methodologischen Beschränkungen steht wie weitergehende Ansätze, die dem Apriori mehr zutrauen.

Natürlich gibt es daneben auch mögliche Positionen, die keine Paraphrase oder Explikation anstreben oder sich weigern, dies mit der Hilfe von Gedankenexperimenten zu tun. Was ist zu solchen Positionen zu sagen?

Natürlich steht es jedem frei, sich aus dem Methodenarsenal der Philosophie soweit zu bedienen, wie es nicht zu pragmatischen Widersprüchen führt (wenn man etwa Methoden verwendet, die nach eigener Auffassung *in dieser Verwendung* unzulässig sind). Wenn man dann auf Methoden *verzichtet*, die nachgewiesenermaßen nicht zu solchen Widersprüchen führen, weil man sie aus anderen Gründen ablehnt, ist dies natürlich legitim. Wenn das eigentliche Ziel solcher Ansätze aber explikativer Art ist, dann sind sie nicht davor geschützt, dass man Gedankenexperimente in kritischer Absicht an diese Ansätze heranträgt. Kritische Gedankenexperimente zurückzuweisen mit dem Hinweis darauf, dass man sie nicht mag, ist *nicht* gerechtfertigt. Man kann Gedankenexperimente zurückweisen, weil sie aus methodologischen Gründen schlecht sind (woran man das feststellt und wie man sie zurückweist, steht in diesem Kapitel), man kann sie aber nicht nur

deswegen zurückweisen, weil sie Gedankenexperimente sind.<sup>388</sup> Das wäre methodologisch genauso falsch, wie empirische Daten zu ignorieren, weil sie gegen die eigene Theorie zu sprechen scheinen. Natürlich gibt es Umstände, in denen man eine Theorie gegen die empirischen Daten in Schutz nehmen muss, und auch Umstände in denen man an einer Theorie festhält, obwohl es Gedankenexperimente gibt, die ihre explikative Adäquatheit in Frage stellen. Solche Umstände bestehen etwa dann, wenn keine befriedigende alternative Theorie existiert, die die akzeptierten Adäquatheitskontexte ansonsten im selben Grad erfüllt. Man kann und wird dann versuchen, die problematischen Adäquatheitskontexte wegzuerklären. Gelingt dies nicht kann es dennoch gerechtfertigt sein, die Theorie nicht aufzugeben, solange kein adäquater Nachfolger gefunden ist. Snowdon – etwa – sah sich in genau einer solchen Situation mit seiner Theorie personaler Identität:

[T]here is no counterintuition that can be generated, and there are no grounds of a general sort for being suspicious of such intuitions, no independent evidence that they do lead us astray. But this is just a case where we find ourselves with what, in the light of all of the evidence, has to be recognized as a deviant, although recalcitrant, intuition. This is a case where what is genuinely possible for us is to be decided of what the best theory of us is – and we must swallow the consequences of that. (Snowdon [301], 126)

Anders liegt die Sache bei Positionen, die keine Explikation zu intendieren scheinen, die die Methode des Gedankenexperiments womöglich in pragmatische Widersprüche brächte. Ein solcher Fall liegt bei der oben rekonstruierten Diskussion zwischen Psychologen und Philosophen vor, in der die Psychologen nicht nachvollziehen können, wieso die Philosophen meinen, mit ihren ausgedachten Experimenten die empirischen Theorien in Frage stellen zu können. Aus der Sicht der Psychologie, sofern sie der Auffassung ist, sie habe bestimmte Begriffe für ihre Theoriebildung gewählt, weil sie für die Erklärungszwecke besonders fruchtbar schienen, ist es durchaus legitim, die Ansprüche der Philosophen zu ignorieren. Diese Forderungen kann man nur dann nicht zurückweisen, wenn man den Anspruch erhebt, an einem gemeinsamen Problem zu arbeiten. Ob es nicht bisweilen heuristisch klüger wäre, den Hinweis der Philosophen ernst zu nehmen, dass man mit dem psychologischen Bedeutungs- oder Wissensbegriff vielleicht an Phänomenen vorbeigeht, die wir intuitiv für erklärungsbedürftig halten, kann man – wie so Vieles – a priori nicht beurteilen.

---

<sup>388</sup> ... oder Gedankenexperimente sind, die der eigenen These widersprechen. Diesen Eindruck macht etwa Johnston [160] und Dennett [85], die die Methode des Gedankenexperimentierens insbesondere dann in Frage stellen, wenn es um Gedankenexperimente geht, denen die eigene Theorie nicht gerecht wird.

## 10. Literaturverzeichnis

- [1] Apostle, Hippocrates G.: *Aristotle's Physics*. Bloomington: Indiana University Press, 1969.
- [2] Aristoteles: *Vom Himmel. Von Der Seele. Von der Dichtkunst*. Zürich: Artemis Verlag, 1983.
- [3] Arthur, Richard: ‚On thought experiments as a priori science‘. *International Studies in the Philosophy of Science*, 13 (1999), 215-229.
- [4] Aspect, Alain: ‚Proposed Experiment to Test the Nonseparability of Quantum Mechanics‘. *Physical Review*, D14 (1976), 1944-1951.
- [5] Aspect, Alan und Grangier, Philippe: ‚Fifty Years Later: When Gedanken Experiments Become Real Experiments‘. In *Niels Bohr: Physics and the World. Proceedings of the Niels Bohr Centennial Symposium, Boston, MA, USA November 12-14, 1985, American Academy of Arts and Sciences, Cambridge, Massachusetts*. Feshbach, Herman, Matsui, Tetsuo und Oleson, Alexandra (Hrsg.) Chur: Harwood Academic Publishers, 1988, 267-283.
- [6] Atkinson, David und Peijnenburg, Jeanne 2004: ‚Galileo and Prior Philosophy‘. In (Manuscript) Groningen.
- [7] Audi, Robert: *Epistemology. A Contemporary Introduction to the Theory of Knowledge*. London: Routledge, 2000.
- [8] Aydede, Murat und Güzeldere, Güven: ‚Consciousness, Conceivability Arguments, and Perspectivalism: The Dialectics of the Debate‘. *Communication & Cognition*, 34 (2001), 97-120.
- [9] Balog, Katalin: ‚Conceivability, Possibility and the Mind-Body Problem‘. *Philosophical Review*, 108 (1999), 497-528.
- [10] Bartelborth, Thomas: *Begründungsstrategien. Ein Weg durch die analytische Erkenntnistheorie*. Berlin: Akademie Verlag, 1996.
- [11] Barwise, Jon: ‚Information and Impossibilities‘. *Notre Dame Journal of Formal Logic*, 38 (1997), 488-515.
- [12] Batens, Diderik und Meheus, Joke: ‚The Adaptive Logic of Compatibility‘. *Studia Logica*, 66 (2000), 327-348.
- [13] Batterman, Robert W.: *The Devil in the Details. Asymptotic Reasoning in Explanation, Reduction, and Emergence*. New York: Oxford University Press, 2002.
- [14] Bealer, George: ‚Intuition and the Autonomy of Philosophy‘. In *Rethinking Intuition. The Psychology of Intuition and Its Role in Philosophical Inquiry*. DePaul, M. R. und Ramsey, W. (Hrsg.) Lanham: Rowman & Littlefield, 1998, 201-239.
- [15] Bealer, George: ‚A Theory of the A Priori‘. *Philosophical Perspectives*, 13 (1999), 29-55.
- [16] Bealer, George: ‚Modal Epistemology and the Rationalist Renaissance‘. In *Conceivability and Possibility*. Gendler, Tamar Szabó und Hawthorne, John (Hrsg.) Oxford: Clarendon Press, 2002, 71-125.

- [17] Beck, Simon: ‚Should We Tolerate People Who Split?‘ *Southern Journal of Philosophy*, XXX (1992), 1-45.
- [18] Beckermann, Ansgar: *Analytische Einführung in die Philosophie des Geistes*. Berlin: de Gruyter, 2001.
- [19] Bell, John S.: ‚On the Einstein Podolsky Rosen paradox‘. *Physics*, I (1964), 195-200.
- [20] Bell, John S.: ‚On the problem of hidden variables in quantum mechanics‘. *Reviews of Modern Physics*, 38 (1966), 447-452.
- [21] Bennett, Jonathan: ‚Descartes's Theory of Modality‘. *Philosophical Review*, 103 (1994), 639-667.
- [22] Bigelow, John, Ellis, Brian und Lierse, Carolin: ‚The World as One of a Kind: Natural Necessity and Laws of Nature‘. *British Journal for the Philosophy of Science*, 43 (1992), 371-388.
- [23] Bird, Alexander: *Thomas Kuhn*. Princeton: Princeton University Press, 2000.
- [24] Birnbacher, Dieter: *Analytische Einführung in die Ethik*. Berlin: de Gruyter, 2003.
- [25] Bishop, Michael A.: ‚Why Thought Experiments are not Arguments‘. *Philosophy of Science*, 66 (1999), 534-541.
- [26] Bishop, Michael A. und Trout, J. D.: *Epistemology and the Psychology of Human Judgment*. New York: Oxford University Press, 2005.
- [27] Blackburn, Simon: *Essays in Quasi-Realism*. New York: Oxford University Press, 1993.
- [28] Bloom, Alfred: *The Linguistic Shaping of Thought. A Study in the Impact of Language on Thinking in China and the West*. Hillsdale: Lawrence Erlbaum Associates, 1981.
- [29] Bohr, Niels: ‚Can quantum-mechanical description of physical reality be considered complete?‘ *Physical Review*, 48 (1935), 696-702.
- [30] Bohr, Niels: ‚Discussion with Einstein on Epistemological Problems in Atomic Physics‘. In *Albert Einstein: Philosopher-Scientist*. Schilpp, Paul Arthur (Hrsg.) New York: Tudor Publishing, 1951, 200-241.
- [31] Bohr, Niels: *Atomphysik und menschliche Erkenntnis*. Braunschweig: Vieweg, 1958.
- [32] Borsbroom, Denny, Mellenbergh, Gideon J. und Van Heerden, Jaap: ‚Functional Thought Experiments‘. *Synthese*, 130 (2002), 379-387.
- [33] Bremer, Manuel: *An Introduction to Paraconsistent Logics*. Frankfurt am Main: Peter Lang, 2005.
- [34] Bremer, Manuel und Cohnitz, Daniel: *Information and Information Flow. An Introduction*. Frankfurt am Main: Ontos-Verlag, 2004.
- [35] Brendel, Elke: ‚Gedankenexperimente als Motor der Wissenschaftsdynamik‘. In *Die Zukunft des Wissens. XVIII. Deutscher Kongress für Philosophie Konstanz 1999*. Mittelstraß, Jürgen (Hrsg.) Konstanz: UVK, 1999, 785-792.
- [36] Broad, Charlie D.: *Ethics and the History of Philosophy*. London: Routledge, 1952.
- [37] Brooks, D. H. M.: ‚The Method of Thought Experiments‘. *Metaphilosophy*, 25 (1994).
- [38] Brown, James Robert: *The Laboratory of the Mind. Thought Experiments in the Natural Sciences*. London: Routledge, 1991.
- [39] Brown, James Robert: ‚Thought Experiments: A Platonic Account‘. In *Thought Experiments in Science and Philosophy*. Horowitz, Tamara und Massey, Gerald J. (Hrsg.) Savage: Rowman & Littlefield, 1991, 119-128.

- [40] Brown, James Robert: *Philosophy of Mathematics. An Introduction to the World of Proofs and Pictures*. London: Routledge, 1999.
- [41] Brown, James Robert: ‚Peeking into Plato’s Heaven’. *Philosophy of Science*, 71 (2004), 1126-1138.
- [42] Brown, James Robert: ‚Why Thought Experiments Transcend Empiricism’. In *Contemporary Debates in Philosophy of Science*. Hitchcock, Christopher (Hrsg.) Malden, Mass.: Blackwell, 2004, 23-43.
- [43] Brunner, J. R.: *Lehrbuch der Physik für die oberen Klassen der Mittelschulen und zum Selbstunterricht*. Zürich: A.-G. Gebr. Leemann & Co., 1936.
- [44] Bühler, Axel: *Die Logik kognitiver Sätze. Über logische Grundlagen der Argumentation in den Geistes- und Sozialwissenschaften*. Berlin: Duncker & Humblot, 1983.
- [45] Bunzl, Martin: ‚Bunzl on Sorensen’s Thought Experiments’. *Informal Logic*, 17 (1995), 388-393.
- [46] Bunzl, Martin: ‚The Logic of Thought Experiments’. *Synthese*, 106 (1996), 227-240.
- [47] Byrne, Alex 2000: ‚Chalmers’ Two-Dimensional Argument Against Physicalism’. (Manuskript, Harvard).
- [48] Caramazza, A., McCloskey, M. und Green, B.: ‚Naive Beliefs in ‚Sophisticated’ Subjects: Misconceptions About Trajectory of Objects’. *Cognition*, 9 (1981), 117-123.
- [49] Carnap, Rudolf: ‚Modalities and Quantification’. *Journal of Symbolic Logic*, 11 (1946), 33-64.
- [50] Carnap, Rudolf: *Logical Foundations of Probability*. London: Routledge, 1967.
- [51] Carnap, Rudolf: *Logische Syntax der Sprache*. Wien: Springer-Verlag, 1968.
- [52] Carnap, Rudolf: *Two Essays on Entropy*. Berkeley: University of California Press, 1977.
- [53] Carnap, Rudolf: *Meaning and Necessity. A Study in Semantics and Modal Logic*. Chicago: University of Chicago Press, 1988.
- [54] Carnap, Rudolf: *Der logische Aufbau der Welt*. Hamburg: Felix Meiner Verlag, 1998.
- [55] Chalmers, David J.: *The Conscious Mind. In Search of a Fundamental Theory*. New York: Oxford University Press, 1996.
- [56] Chalmers, David J.: ‚Materialism and the Metaphysics of Modality’. *Philosophy and Phenomenological Research*, LIX (1999), 473-496.
- [57] Chalmers, David J.: ‚The Components of Content’. In *Philosophy of Mind: Classical and Contemporary Readings*. Chalmers, David J. (Hrsg.) New York: Oxford University Press, 2002, 608-633.
- [58] Chalmers, David J.: ‚Does Conceivability Entail Possibility?’ In *Conceivability and Possibility*. Gendler, Tamar Szabó und Hawthorne, John (Hrsg.) Oxford: Clarendon Press, 2002, 145-200.
- [59] Chalmers, David J.: ‚On Sense and Intension’. *Philosophical Perspectives*, 16 (2002), 135-182.
- [60] Chalmers, David J.: ‚Epistemic Two-Dimensionalism’. *Philosophical Studies*, 118 (2004), 153-226.
- [61] Chalmers, David J.: ‚The Foundations of Two-Dimensional Semantics’. In *Two-Dimensional Semantics: Foundations and Applications*. Garcia-Carpentero, M. und Macia, J. (Hrsg.) Oxford: Oxford University Press, 2004.
- [62] Chalmers, David J.: ‚Imagination, Indexicality, and Intensions’. *Philosophy and Phenomenological Research*, 68 (2004), 182-190.

- [63] Chalmers, David J. und Jackson, Frank: ‚Conceptual Analysis and Reductive Explanation‘. *Philosophical Review*, 110 (2001), 315-361.
- [64] Chomsky, Noam: ‚Language and Nature‘. *Mind*, 104 (1995), 1-59.
- [65] Chomsky, Noam: *New Horizons in the Study of Language and Mind*. New York: Cambridge University Press, 2002.
- [66] Clauser, John F., Horne, Michael A., Shimony, Abner und Holt, Richard A.: ‚Proposed Experiment to Test Local Hidden-Variable Theories‘. *Physical Review Letters*, 23 (1969), 880-884.
- [67] Clement, John: ‚Student’s Preconceptions in Elementary Mechanics‘. *American Journal of Physics*, 50 (1982), 66-71.
- [68] Cohnitz, Daniel: ‚Explanations are Like Salted Peanuts. Why you can’t cut the route toward further reduction‘. In *Argument und Analyse: Proceedings of GAP4*. Beckermann, Ansgar und Nimtz, Christian (Hrsg.) Paderborn: Mentis, 2002, 22-36.
- [69] Cohnitz, Daniel: ‚Modal Skepticism. Philosophical Thought Experiments and Modal Epistemology‘. In *The Vienna Circle and Logical Empiricism. Re-Evaluation and Future Perspectives*. Stadler, Friedrich (Hrsg.) Dordrecht: Kluwer Academic Publishers, 2003, 281-296.
- [70] Cohnitz, Daniel: ‚Personal Identity and the Methodology of Imaginary Cases‘. In *On Human Persons*. Petrus, Klaus (Hrsg.) Frankfurt am Main: Ontos-Verlag, 2003, 145-181.
- [71] Cohnitz, Daniel: ‚Two-Dimensionalism and the Metaphysical Possibility of Zombies‘. In *Foundations of the Formal Sciences II: Applications of Mathematical Logic in Philosophy and Linguistics, Papers of a Conference held in Bonn, November 10-13, 2000*. Löwe, Benedikt, Malzkorn, Wolfgang und Räscher, Thoralf (Hrsg.) Dordrecht: Kluwer Academic Publishers, 2003, 55-62.
- [72] Cohnitz, Daniel: ‚Why Consistentism Won’t Work‘. In *Modal Epistemology. 23 September 2004*. Weber, Erik und De Mey, Tim (Hrsg.) Brüssel: Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten, 2004, 33-51.
- [73] Cohnitz, Daniel: ‚Poor Thought Experiments? A Comment on Peijnenburg and Atkinson‘. *Journal for General Philosophy of Science*, (Im Erscheinen).
- [74] Cohnitz, Daniel: ‚War Hans Christian Ørsted der ‚Vater‘ des Gedankenexperiments? Ein Beitrag zur Begriffsgeschichte und zur Mathematikgeschichte des frühen 19. Jahrhunderts‘. *Kant-Studien*, (Im Erscheinen).
- [75] Cohnitz, Daniel und Rossberg, Marcus: *Nelson Goodman*. Chesham: Acumen, Im Erscheinen.
- [76] Cohnitz, Daniel und Smith, Barry: ‚Assessing Ontologies: The Question of Human Origins and Its Ethical Significance‘. In *Persons: An Interdisciplinary Approach*. Runggaldier, E. und Kanzian, C. (Hrsg.) Wien: öbv&hpt, 2003, 243-259.
- [77] Coleman, Stephen: ‚Thought Experiments and Personal Identity‘. *Philosophical Studies*, 98 (2000), 53-69.
- [78] Cooper, Rachel: ‚Thought Experiments‘. *Metaphilosophy*, 36 (2005), 328-347.
- [79] Craik, K.: *The Nature of Explanation*. Cambridge: Cambridge University Press, 1943.
- [80] Crick, Nathan: ‚Conquering Our Imagination: Thought Experiments and Enthymemes in Scientific Argument‘. *Philosophy and Rhetoric*, 37 (2004), 21-41.

- [81] Cummins, Robert: ‚Methodological Reflections on Belief‘. In *Mind and Common Sense*. Bogdan, Radu (Hrsg.) Cambridge: Cambridge University Press, 1991, 53-70.
- [82] Dancy, Jonathan: ‚The Role of Imaginary Cases in Ethics‘. *Pacific Philosophical Quarterly*, 66 (1985), 141-153.
- [83] Dancy, Jonathan: *Moral Reasons*. Oxford: Blackwell, 1993.
- [84] De Mey, Tim und Weber, Erik: ‚Explanation and Thought Experiments in History‘. *History and Theory*, 42 (2003), 28-38.
- [85] Dennett, Daniel C.: *Consciousness Explained*. Penguin, 1993.
- [86] Descartes, René: *Philosophische Schriften in einem Band*. Hamburg: Felix Meiner Verlag, 1996.
- [87] Dickie, George: ‚The Institutional Theory of Art‘. In *Theories of Art Today*. Carroll, Noël (Hrsg.) Madison: University of Wisconsin Press, 2000, 93-108.
- [88] Dietrich, Frank: ‚The Computer: A Tool for Thought-Experiments‘. *Leonardo*, 20 (1987), 315-325.
- [89] Disalle, Robert: ‚Newton’s Philosophical Analysis of Space and Time‘. In *The Cambridge Companion to Newton*. Cohen, I. Bernard und Smith, George E. (Hrsg.) New York: Cambridge University Press, 2002, 33-56.
- [90] Drake, Stillman: ‚Galileo’s Discovery of the Law of Free Fall‘. *Scientific American*, 228 (1973), 84-92.
- [91] Drake, Stillman: *Galileo at Work. His Scientific Biography*. Chicago: University of Chicago Press, 1978.
- [92] Dretske, Fred I.: *Knowledge and the Flow of Information*. Stanford: CSLI, 1999.
- [93] Duhem, Pierre: *Le Système du Monde, tom. I*. Paris: 1914.
- [94] Duhem, Pierre: *Ziel und Struktur der physikalischen Theorien*. Hamburg: Felix Meiner, 1978.
- [95] Düsberg, Klaus-Jürgen: ‚Probleme einer Interpretation der Quantenmechanik. Zur Relevanz der Quantenmechanik für die Philosophie‘. In *Analyse, Rekonstruktion, Kritik*. Lechner, Jochen (Hrsg.) Frankfurt am Main: Peter Lang, 1998.
- [96] Earman, John und Norton, John D.: ‚Exorcist XIV: The Wrath of Maxwell’s Demon: From Maxwell to Szilard‘. *Studies in the History and Philosophy of Modern Physics*, 29 (1998), 435-471.
- [97] Earman, John und Norton, John D.: ‚Exorcist XIV: The Wrath of Maxwell’s Demon: From Szilard to Landauer and Beyond‘. *Studies in the History and Philosophy of Modern Physics*, 30 (1999), 1-40.
- [98] Edidin, Aron: ‚A Priori Knowledge for Fallibilists‘. *Philosophical Studies*, 46 (1984), 189-197.
- [99] Einstein, Albert: ‚Die Grundlage der allgemeinen Relativitätstheorie‘. In *Das Relativitätsprinzip. Eine Sammlung von Abhandlungen*. Einstein, Albert, Lorenz, H. A. und Minkowski, H. (Hrsg.) Leipzig: B.G. Teubner, 1920, 81-124.
- [100] Einstein, Albert, Podolsky, B. und Rosen, N.: ‚Can quantum-mechanical description of physical reality be considered complete?‘ *Physical Review*, 47 (1935), 777-780.
- [101] Fallis, Don: ‚The Source of Chaitin’s Incorrectness‘. *Philosophia Mathematica*, 4 (1996), 261-269.
- [102] Feferman, Solomon: ‚Mathematical Intuition vs. Mathematical Monsters‘. *Synthese*, 125 (2000), 317-332.

- [103] Feldman, Richard: ‚Feldman on Sorensen’s Thought Experiments’. *Informal Logic*, 17 (1995), 394-398.
- [104] Festinger, L.: *A Theory of Cognitive Dissonance*. Palo Alto: Stanford University Press, 1957.
- [105] Feys, Robert: ‚Carnap on Modalities’. In *The Philosophy of Rudolf Carnap*. Schilpp, Paul Arthur (Hrsg.) La Salle: Open Court, 1963, 283-297.
- [106] Fodor, Jerry A.: ‚On Knowing What We Would Say’. *The Philosophical Review*, LXXIII (1964), 198-212.
- [107] Fodor, Jerry A., Garrett, Merrill F., Walker, Edward C. T. und Parkes, Cornelia H.: ‚Against Definitions’. In *Concepts. Core Readings*. Margolis, Eric und Laurence, Stephen (Hrsg.) Cambridge, Mass.: MIT Press, 1999, 491-512.
- [108] Fodor, Jerry A. und Lepore, Ernest: *Holism. A shopper’s guide*. Oxford: Blackwell, 1992.
- [109] Fodor, Jerry A. und LePore, Ernest: ‚Why Meaning (Probably) Isn’t Conceptual Role’. In *Mental Representation. A Reader*. Stich, Stephen P. und Warfield, Ted A. (Hrsg.) Oxford: Blackwell, 1994, 142-156.
- [110] Føllesdal, Dagfinn: *Referential Opacity and Modal Logic*. Oslo: Universitetsforlaget, 1966.
- [111] French, Peter A.: ‚Kinds and Persons’. *Philosophy and Phenomenological Research*, XLIV (1983), 241-254.
- [112] Fuhrmann, André: ‚Das Mögliche und das Vorstellbare’. *Logos*, N.F. 7 (2002), 343-360.
- [113] Gähde, Ulrich: ‚Gedankenexperimente in Erkenntnistheorie und Physik: strukturelle Parallelen’. In *Rationalität, Realismus, Revision. Vorträge des 3. internationalen Kongresses der Gesellschaft für Analytische Philosophie vom 15. bis zum 18. September 1997 in München*. Nida-Rümelin, Julian (Hrsg.) Berlin: de Gruyter, 2000, 457-464.
- [114] Gähde, Ulrich: ‚Zur Funktion ethischer Gedankenexperimente’. In *Wirtschaftsethische Perspektiven V. Methodische Ansätze, Probleme der Steuer- und Verteilungsgerechtigkeit, Ordnungsfragen*. Gaertner, Wulf (Hrsg.) Berlin: Duncker & Humblot, 2000, 183-206.
- [115] Gale, Richard: ‚On Some Pernicious Thought Experiments’. In *Thought Experiments in Science and Philosophy*. Horowitz, Tamara und Massey, Gerald J. (Hrsg.) Savage: Rowman & Littlefield, 1991, 297-305.
- [116] Galilei, Galileo: *Unterredungen und mathematische Demonstrationen über zwei neue Wissenszweige, die Mechanik und die Fallgesetze betreffend, erster und zweiter Tag (1638)*. Leipzig: Wilhelm Engelmann, 1890.
- [117] Galilei, Galileo: *Unterredungen und mathematische Demonstrationen über zwei neue Wissenszweige, die Mechanik und die Fallgesetze betreffend, fünfter und sechster Tag (1638)*. Leipzig: Wilhelm Engelmann, 1891.
- [118] Galilei, Galileo: *Dialog über die beiden hauptsächlichsten Weltsysteme: das Ptolemäische und das Kopernikanische*. Stuttgart: B. G. Teubner, 1982.
- [119] Gendler, Tamar Szabó: ‚Exceptional Persons: On the Limits of Imaginary Cases’. *Journal of Consciousness Studies*, 5 (1998), 592-610.
- [120] Gendler, Tamar Szabó: ‚Galileo and the Indispensability of Scientific Thought Experiment’. *British Journal for the Philosophy of Science*, 49 (1998), 397-424.

- [121] Gendler, Tamar Szabó: *Thought Experiments. On the Power and Limits of Imaginary Cases*. New York: Garland Publishing, 2000.
- [122] Gesang, Bernward: *Kritik des Partikularismus*. Paderborn: Mentis, 2000.
- [123] Gettier, Edmund L.: ‚Is Justified True Belief Knowledge?‘ *Analysis*, 23.6 (1963), 121-123.
- [124] Gilbert, Daniel T. und Wilson, Timothy D.: ‚Miswanting: Some problems in forecasting of future affective states‘. In *Feeling and Thinking: The Role of Affect in Social Cognition*. Forgas, J. (Hrsg.) Cambridge: Cambridge University Press, 2000, 178-197.
- [125] Goeres, Ralf: ‚Sensualistischer Phänomenalismus und Denkökonomie. Zur Wissenschaftskonzeption Ernst Machs‘. *Journal for General Philosophy of Science*, 35 (2004), 41-70.
- [126] Goldman, Alvin I.: *Liaisons. Philosophy Meets the Cognitive and Social Sciences*. Cambridge, Mass.: MIT Press, 1992.
- [127] Goldman, Alvin I. und Pust, Joel: ‚Philosophical Theory and Intuitional Evidence‘. In *Rethinking Intuition. The Psychology of Intuition and Its Role in Philosophical Inquiry*. DePaul, M. R. und Ramsey, W. (Hrsg.) Lanham: Rowman & Littlefield, 1998, 179-197.
- [128] Gooding, David C.: ‚What is Experimental about Thought Experiments?‘ *Philosophy of Science*, 2 (1992), 280-290.
- [129] Gopnik, Alison und Schwitzgebel, Eric: ‚Whose concepts are they, anyway? The role of philosophical intuition in empirical psychology‘. In *Rethinking Intuition. The Psychology of Intuition and Its Role in Philosophical Inquiry*. DePaul, M. R. und Ramsey, W. (Hrsg.) Lanham: Rowman & Littlefield, 1998, 75-91.
- [130] Gottlob, Georg: ‚Remarks on a Carnapian Extension of S5‘. In *Alfred Tarski and the Vienna Circle. Austro-Polish Connections in Logical Empiricism*. Wole ski, Jan und Köhler, Eckehart (Hrsg.) Dordrecht: Kluwer Academic Publishers, 1998, 243-259.
- [131] Gottlob, Georg: ‚Review of a Carnapian extension of S5‘. In *Logic at Work. Essays Dedicated to the Memory of Helena Rasiowa*. Orłowska, Ewa (Hrsg.) Heidelberg: Physica-Verlag, 1999, 55-78.
- [132] Green, Brian: *The Elegant Universe. Superstrings, Hidden Dimensions, and the Quest for the Ultimate Theory*. New York: Vintage Books, 1999.
- [133] Güzeldere, Güven: ‚Varieties of Zombiehood‘. *Journal of Consciousness Studies*, 2 (1995), 326-333.
- [134] Haas-Spohn, Ulrike: *Versteckte Indexikalität und subjektive Bedeutung*. Berlin: Akademie-Verlag, 1995.
- [135] Hacking, Ian: *Logic of Statistical Inference*. London: Cambridge University Press, 1965.
- [136] Hacking, Ian: ‚Do Thought Experiments Have a Life of Their Own? Comments on James Brown, Nancy Nersessian and David Gooding‘. *Philosophy of Science*, 2 (1992), 302-308.
- [137] Häggqvist, Sören: *Thought Experiments in Philosophy*. Stockholm: Almqvist & Wiksell International, 1996.
- [138] Haidt, J., Koller, S. und Dias, M.: ‚Affect, Culture and Morality, or Is It Wrong to Eat Your Dog?‘ *Journal of Personality and Social Psychology*, 65 (1993), 613-628.

- [139] Hale, Bob: ‚Knowledge of Possibility and Knowledge of Necessity’. *Proceedings of the Aristotelian Society*, CIII (2002), 1-20.
- [140] Halloun, I. A. und Hestenes, D.: ‚Common Sense Concepts About Motion’. *American Journal of Physics*, 53 (1985), 1056-1065.
- [141] Helbig, Gerhard und Buscha, Joachim: *Deutsche Grammatik. Ein Handbuch für den Ausländerunterricht*. Leipzig: Langenscheidt, 1994.
- [142] Held, Carsten: *Die Bohr-Einstein Debatte. Quantenmechanik und physikalische Wirklichkeit*. Paderborn: Schöningh, 1998.
- [143] Hempel, Carl Gustav: *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press, 1965.
- [144] Hempel, Carl Gustav: ‚Fundamentals of Concept Formation in Empirical Science’. In *Foundations of the Unity of Science*. 2. Neurath, Otto, Carnap, Rudolf und Morris, Charles (Hrsg.) Chicago: University of Chicago Press, 1970, 652-745.
- [145] Hempel, Carl Gustav: *Grundzüge der Begriffsbildung in den empirischen Wissenschaften*. Düsseldorf: Bertelsmann Universitätsverlag, 1974.
- [146] Hendry, Herbert E. und Pokriefka, M. L.: ‚Carnapian Extensions of S5’. *Journal of Philosophical Logic*, 14 (1985), 111-128.
- [147] Hetherington, Stephen Cade: ‚Conceivability and Modal Knowledge’. In *Thought Experiments in Science and Philosophy*. Horowitz, Tamara und Massey, Gerald J. (Hrsg.) Savage: Rowman & Littlefield, 1991, 317-325.
- [148] Hintikka, Jaakko: ‚Who is about to kill analytic philosophy?’ In *The Story of Analytic Philosophy. Plot and Heroes*. Biletzki, Anat und Matar, Anat (Hrsg.) New York: Routledge, 1998, 253-269.
- [149] Hintikka, Jaakko: ‚The Emperor’s New Intuitions’. *Journal of Philosophy*, 96 (1999), 127-147.
- [150] Hitchens, Christopher: *The Elgin Marbles*. London: Chatto and Windus, 1987.
- [151] Hoeschen, Oliver: *Verstehen und Rationalität: Donald Davidsons Rationalitätsthese und die kognitive Psychologie*. Paderborn: Mentis, 2002.
- [152] Horowitz, Tamara: ‚Philosophical Intuitions and Psychological Theory’. *Ethics*, 108 (1998), 367-385.
- [153] Hughes, G. E. und Cresswell, M. J.: *A New Introduction to Modal Logic*. London: Routledge, 1996.
- [154] Hume, David: *A Treatise of Human Nature*. Oxford: Clarendon Press, 1888.
- [155] Hume, David: *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*. Oxford: Clarendon Press, 2000.
- [156] Humphreys, Paul: ‚Seven Theses on Thought Experiments’. In *Philosophical Problems of the Internal and External Worlds: Essays on the Philosophy of Adolf Grünbaum*. Earman, J., Janis, A., Massey, Gerald J. und Rescher, Nicholas (Hrsg.) Pittsburgh: University of Pittsburgh Press, 1993, 205-227.
- [157] Jackson, Frank: ‚Epiphenomenal Qualia’. *Philosophical Quarterly*, 32 (1982), 127-136.
- [158] Jackson, Frank: *From Metaphysics to Ethics. A Defence of Conceptual Analysis*. Oxford: Clarendon Press, 2000.
- [159] Johnson-Laird, P. N.: *Mental Models*. Cambridge: Cambridge University Press, 1983.
- [160] Johnston, Mark: ‚Human Beings’. *Journal of Philosophy*, LXXXIV (1987), 59-83.

- [161] Kannuck, Kathryn 2000: ‚Personal Identity: The Significance of The Fission Thought Experiment‘. Manuskript, University of Florida p. v+92.
- [162] Kannuck, Kathryn 2003: ‚Gendler on Why We Can't Trust Thought Experiments on Personal Identity‘. Manuskript, University of Florida.
- [163] Kaplan, David: ‚Demonstratives‘. In *Themes from Kaplan*. Almog, J., Perry, J. und Wettstein, H. (Hrsg.) Oxford: Oxford University Press, 1989, 481-563.
- [164] Kitcher, Patricia: ‚Natural Kinds and Unnatural Persons‘. *Philosophy*, 54 (1979), 541-547.
- [165] Knott, C. G.: *Life and Scientific Work of Peter Guthrie Tait*. London: Cambridge University Press, 1911.
- [166] Kornblith, Hilary (Hrsg.): *Naturalizing Epistemology*. Harvard: MIT Press, 1994.
- [167] Kornblith, Hilary: *Knowledge and its Place in Nature*. Oxford: Clarendon Press, 2002.
- [168] Koyré, Alexandre: *Galileo Studies*. Sussex: The Harvester Press, 1978.
- [169] Koyré, Alexandre: *Galilei. Die Anfänge der neuzeitlichen Wissenschaft*. Berlin: Verlag Klaus Wagenbach, 1988.
- [170] Krimsky, Sheldon: ‚The Use and Misuse of Critical Gedankenexperiment‘. *Journal for General Philosophy of Science*, IV (1973), 323-334.
- [171] Kripke, Saul A.: ‚A Completeness Theorem in Modal Logic‘. *Journal of Symbolic Logic*, 24 (1959), 1-14.
- [172] Kripke, Saul A.: *Name und Notwendigkeit*. Frankfurt am Main: Suhrkamp, 1981.
- [173] Kuhn, Thomas S.: *Die Entstehung des Neuen. Studien zur Struktur der Wissenschaftsgeschichte*. Frankfurt am Main: Suhrkamp, 1978.
- [174] Kühne, Ulrich: *Gedankenexperiment und Erklärung*. 5. Bremen: Universität Bremen, Studiengang Philosophie, 1997.
- [175] Kühne, Ulrich 2001: ‚Die Methode des Gedankenexperiments. Untersuchung zur Rationalität naturwissenschaftlicher Theoriereformen‘. Manuskript, Bremen.
- [176] Kühne, Ulrich: *Die Methode des Gedankenexperiments*. Frankfurt am Main: Suhrkamp, 2005.
- [177] Lakatos, Imre: *Beweise und Widerlegungen: die Logik mathematischer Entdeckungen*. Braunschweig: Vieweg, 1974.
- [178] Lakatos, Imre: *Proofs and Refutations. The Logic of Mathematical Discovery*. New York: Cambridge University Press, 1995.
- [179] Laymon, Ronald: ‚Newton's Bucket Experiment‘. *Journal of the History of Philosophy*, XVI (1978), 399-413.
- [180] Lazarsfeld, P. F.: ‚Latent Structure Analysis‘. In *Psychology: a study of a science*. III. Koch, S. (Hrsg.) New York: MacGraw-Hill, 1959.
- [181] Leff, Harvey S. und Rex, Andrew F. (Hrsg.): *Maxwell's Demon. Entropy, Information, Computing*. Bristol: Adam Hilger, 1990.
- [182] Leibniz, G. W.: *Neue Abhandlungen über den menschlichen Verstand*. Ditzingen: Reclam, 1993.
- [183] Leibniz, G. W.: *Philosophische Schriften*. 1. Frankfurt am Main: Suhrkamp, 1996.
- [184] Levine, Joseph: *Purple Haze. The Puzzle of Consciousness*. New York: Oxford University Press, 2001.
- [185] Lewis, David: *Counterfactuals*. Malden: Blackwell, 1973.

- [186] Lewis, David: *Die Identität von Körper und Geist*. Frankfurt am Main: Klostermann, 1983.
- [187] Lewis, David: *On the Plurality of Worlds*. Malden: Blackwell, 1986.
- [188] Locke, John: ‚Of Identity and Diversity‘. In *Personal Identity*. Perry, John (Hrsg.) Berkeley: University of California Press, 1975, 33-52.
- [189] Loomis, Eric: ‚Necessity, The A priori, And The Standard Meter‘. *Synthese*, 121 (1999), 291-307.
- [190] Lord, Frederic M. und Novick, Melvin R.: *Statistical Theories of Mental Test Scores*. Reading, Mass.: Addison-Wesley, 1974.
- [191] Mach, Ernst: *Erkenntnis und Irrtum*. Leipzig: 1905.
- [192] Mach, Ernst: *Populär-wissenschaftliche Vorlesungen*. Leipzig: Johann Ambrosius Barth, 1910.
- [193] Mach, Ernst: *Die Mechanik in ihrer Entwicklung. Historisch-kritisch dargestellt*. Berlin: Akademie-Verlag, 1988.
- [194] Mach, Ernst: *Analyse der Empfindungen*. Darmstadt: Wissenschaftliche Buchgesellschaft, 1991.
- [195] Machery, Edouard: ‚Concepts Are Not a Natural Kind‘. *Philosophy of Science*, (Im Erscheinen).
- [196] Machery, Edouard, Mallon, Ron, Nichols, Shaun und Stich, Stephen P.: ‚Semantics, cross-cultural style‘. *Cognition*, 92 (2004), B1-B12.
- [197] MacLachlan, James: ‚A Test of an "Imaginary" Experiment of Galileo's‘. *Isis*, 64 (1973), 374-379.
- [198] Maffie, James: ‚"Just-so" stories about "inner cognitive Africa": some doubts about Sorensen's evolutionary epistemology of thought experiments‘. *Biology and Philosophy*, 12 (1997), 207-224.
- [199] Makinsons, David: ‚How Meaningful are Modal Operators?‘ *Australasian Journal of Philosophy*, 44 (1966), 331-337.
- [200] Martin, Raymond: ‚Real Values: Why the Wilkes-Donagan Prohibition is Mistaken‘. *Metaphilosophy*, 24 (1993), 400-406.
- [201] Mayer, Verena: ‚Was zeigen Gedankenexperimente?‘ *Philosophisches Jahrbuch*, 106 (1999), 357-378.
- [202] McAllister, James W.: ‚The Evidential Significance of Thought Experiment in Science‘. *Studies in the History and Philosophy of Science*, 27 (1996), 233-250.
- [203] McAllister, James W.: ‚Thought Experiments and the Belief in Phenomena‘. *Philosophy of Science*, 71 (2004), 1164-1175.
- [204] McDermott, L.: ‚An overview on conceptual understanding in physics‘. *Physics Today*, 37 (1984), 636-649.
- [205] McFarland, C., Ross, M. und DeCourville, N.: ‚Women's theories of menstruation and biases in recall of menstrual symptoms‘. *Journal of Personality and Social Psychology*, 57 (1989), 522-531.
- [206] McGinn, Colin: *Die Grenzen vernünftigen Fragens*. Stuttgart: Klett-Cotta, 1996.
- [207] Meinong, Alexis: *Über die Stellung der Gegenstandstheorie im System der Wissenschaften*. Leipzig: R. Voigtländers Verlag, 1907.
- [208] Menzies, Peter: ‚Possibility and Conceivability: A Response-Dependent Account of their Connections‘. *European Review of Philosophy*, 3 (1998), 255-277.

- [209] Mill, John Stuart: *A System of Logic*. New York: Haper & Brothers, 1874.
- [210] Mišcevic, Nenad: ‚Mental Models and thought experiments’. *International Studies in the Philosophy of Science*, 6 (1992), 215-226.
- [211] Misselhorn, Catrin: *Wirkliche Möglichkeiten - Mögliche Wirklichkeiten*. Paderborn: Mentis, 2005.
- [212] Moore, George Edward: *Philosophical Papers*. New York: Macmillan, 1966.
- [213] Mormann, Thomas: *Rudolf Carnap*. München: Beck, 2000.
- [214] Müller, Olaf: *Wirklichkeit ohne Illusionsn*. 1: Hilary Putnam und der Abschied vom Skeptizismus oder Warum die Welt keine Computersimulation sein kann. Paderborn: Mentis, 2003.
- [215] Murphy, Gregory und Medin, Douglas: ‚The Role of Theories in Conceptual Coherence’. In *Concepts. Core Readings*. Margolis, Eric und Laurence, Stephen (Hrsg.) Cambridge, Mass.: MIT Press, 1999, 425-458.
- [216] Musgrave, Alan: *Essays on Realism and Rationalism*. Amsterdam: Rodopi, 1999.
- [217] Myers, Mason C.: ‚Thought Experiments and Secret Stors of Information’. *International Philosophical Quarterly*, VIII (1968), 180-192.
- [218] Myers, Mason C.: ‚Moore’s Paradox of Analysis’. *Metaphilosophy*, 2 (1971), 295-308.
- [219] Myers, Mason C.: ‚Analytical Thought Experiments’. *Metaphilosophy*, 17 (1986), 109-118.
- [220] Nagel, Thomas: ‚Brain Bisection and the Unity of Consciousness’. *Synthese*, 22 (1971), 396-413.
- [221] Nagel, Thomas: ‚Conceiving the Impossible and the Mind-Body Problem’. *Philosophy*, 73 (1998), 337-352.
- [222] Needham, Paul: ‚What is Water?’ *Analysis*, 60.1 (2000), 31-21.
- [223] Nersessian, Nancy J.: ‚How Do Scientists Think? Capturing the Dynamics of Conceptual Change in Science’. In *Cognitive Models of Science*. Giere, R. N. (Hrsg.) Minneapolis: University of Minnesota Press, 1989, 3-45.
- [224] Nersessian, Nancy J.: ‚In the theoretician’s laboratory: Thought experimenting as mental modeling’. *Philosophy of Science*, 2 (1992), 291-301.
- [225] Nersessian, Nancy J.: ‚Model-based reasoning in conceptual change’. In *Model-Based Reasoning in Scientific Discovery*. Magnani, L., Nersessian, N. J. und Thagard, P. (Hrsg.) Dordrecht: Kluwer Academic Publishers, 1999, 5-22.
- [226] Nersessian, Nancy J. und Resnick, Lauren B. 2004: ‚Comparing Historical and Intuitive Explanations of Motion: Does "Naive Physics" Have a Structure?’ Manuskript, Princeton/Pittsburgh.
- [227] Neumann, Carl: *Über die Principien der Galilei-Newtonschen Theorie*. Leipzig: 1870.
- [228] Newen, Albert 2002: ‚Theorien des Selbstbewußtseins. Ihr neuzeitlicher Ursprung und die Entwicklung in der modernen analytischen Philosophie’. Manuskript, Bonn: Universität Bonn.
- [229] Newton, Isaac: *Mathematische Prinzipien der Naturlehre*. Darmstadt: Wissenschaftliche Buchgesellschaft, 1963.
- [230] Nichols, Shaun, Stich, Stephen P. und Weinberg, Jonathan M.: ‚Metaskepticism: Meditations in Ethno-Epistemology’. In *The Sceptics*. Luper, S. (Hrsg.) Aldershot: Ashgate Publishing, 2003, 227-247.

- [231] Nimtz, Christian: ‚Analytic Truths- Still Harmless After All These Years?‘ *Grazer Philosophische Studien*, 66 (2003), 91-188.
- [232] Nimtz, Christian: ‚Two-Dimensionalism and Natural Kind Terms‘. *Synthese*, 138 (2004), 125-148.
- [233] Nimtz, Christian: ‚Semantics (Two-Dimensional)‘. In *The Encyclopedic Reference of Neuroscience*. Binder, Marc D., Hirokawa, Nobutaka, Windhorst, Uwe und Hirsch, Martin C. (Hrsg.) Berlin: Springer, 2005.
- [234] Nisbett, Richard E.: *The Geography of Thought: How Asians and Westerners Think Differently...and Why*. New York: Free Press, 2003.
- [235] Nisbett, Richard E., Peng, K., Choi, I. und Norenzayan, A.: ‚Culture and Systems of Thought: Holistic Versus Analytic Cognition‘. *Psychological Review*, 108 (2001), 291-310.
- [236] Noonan, Harold: ‚Critical Study: Wiggins‘ Second Thoughts on Identity‘. *Philosophical Quarterly*, 31 (1981), 260-268.
- [237] Noonan, Harold: *Personal Identity*. London: Routledge, 1991.
- [238] Norton, John D.: ‚Thought Experiments in Einstein’s Work‘. In *Thought Experiments in Science and Philosophy*. Horowitz, Tamara und Massey, Gerald J. (Hrsg.) Savage: Rowman & Littlefield, 1991, 129-148.
- [239] Norton, John D.: ‚Are Thought Experiments Just What You Thought?‘ *Canadian Journal of Philosophy*, 26 (1996), 333-366.
- [240] Norton, John D.: ‚On Thought Experiments: Is There More to the Argument?‘ *Philosophy of Science*, 71 (2004), 1139-1151.
- [241] Norton, John D.: ‚Why Thought Experiments do not Transcend Empiricism‘. In *Contemporary Debates in Philosophy of Science*. Hitchcock, Christopher (Hrsg.) Malden, Mass.: Blackwell, 2004, 44-66.
- [242] Norton, John D.: ‚Eaters of the Lotus: Landaur’s Principle and the Return of Maxwell’s Demon‘. *Studies in the History and Philosophy of Modern Physics*, (Im Erscheinen).
- [243] Nozick, Robert: *Philosophical Explanations*. Cambridge, Mass.: Harvard University Press, 1981.
- [244] Nozick, Robert: *Invariances: The Structure of the Objective World*. Cambridge, Mass.: Belknap Press, 2001.
- [245] O’Leary-Hawthorne, John: ‚The Epistemology of Possible Worlds: A Guided Tour‘. *Philosophical Studies*, 84 (1996), 183-202.
- [246] Ørsted, Hans Christian: ‚Über Geist und Studium der allgemeinen Naturlehre‘. *Gebellens Journal für Chemie und Physik*, 63 (1822), 458-488.
- [247] Pagin, Peter: ‚Meaning Holism‘. In *Handbook of Philosophy of Language*. Lepore, Ernest und Smith, B. (Hrsg.) New York: Oxford University Press, Im Erscheinen.
- [248] Pap, Arthur: *Elements of Analytic Philosophy*. New York: Hafner Publishing Company, 1949.
- [249] Papineau, David: *Thinking About Consciousness*. Oxford: Clarendon Press, 2002.
- [250] Parfit, Derek: *Reasons and Persons*. Oxford: Clarendon Press, 1987.
- [251] Pauen, Michael: *Grundprobleme der Philosophie des Geistes*. Frankfurt am Main: Fischer, 2002.
- [252] Peijnenburg, Jeanne und Atkinson, David: ‚When are thought experiments poor ones?‘ *Journal for General Philosophy of Science*, 34 (2003), 305-322.

- [253] Perry, John: ‚Personal Identity, Memory, and the Problem of Circularity‘. In *Personal Identity*. Perry, John (Hrsg.) Berkeley: University of California Press, 1975, 135-155.
- [254] Perry, John: ‚Kann das Ich sich teilen?‘ In *Personale Identität*. Quante, Michael (Hrsg.) Paderborn: Schöningh, 1999, 121-154.
- [255] Perry, John: *Knowledge, Possibility, and Consciousness*. Cambridge, Mass.: MIT Press, 2001.
- [256] Perry, John: *Identity, Personal Identity, and the Self*. Indianapolis: Hackett, 2002.
- [257] Piaget, J.: *Les notions de mouvement et de vitesse chez l'enfant*. Paris: 1946.
- [258] Pitt, D.: ‚In defense of definitions‘. *Philosophical Psychology*, 12 (1999), 139-156.
- [259] Polger, Thomas W.: ‚Zombies Explained‘. In *Dennett's Philosophy. A Comprehensive Assessment*. Ross, Don, Brook, Andrew und Thompson, David (Hrsg.) Cambridge, Mass.: MIT Press, 2000, 259-286.
- [260] Popper, Karl R.: *Logik der Forschung*. Tübingen: J.C.B. Mohr (Paul Siebeck), 1994.
- [261] Prinz, Jesse: *Furnishing the Mind. Concepts and their Perceptual Basis*. Cambridge, Mass.: MIT Press, 2002.
- [262] Pulte, Helmut: ‚Gedankenexperimente in der neueren Wissenschaftstheorie: Über ihre Funktion und ihren Gebrauch in Hinblick auf erfahrungswissenschaftliche Theorien‘. *Journal for General Philosophy of Science*, (Im Erscheinen).
- [263] Putnam, Hilary: *Die Bedeutung von "Bedeutung"*. Frankfurt am Main: Klostermann, 1990.
- [264] Putnam, Hilary: *Vernunft, Wahrheit und Geschichte*. Frankfurt am Main: Suhrkamp, 1990.
- [265] Putnam, Hilary: *Realism with a Human Face*. Cambridge, Mass.: Harvard University Press, 1992.
- [266] Putnam, Hilary: ‚Is Semantics Possible?‘ In *Concepts. Core Readings*. Margolis, Eric und Laurence, Stephen (Hrsg.) Cambridge, Mass.: MIT Press, 1999, 177-187.
- [267] Quine, W. V.: ‚Milton K. Munitz (ed.): Identity and Individuation‘. *Journal of Philosophy*, LXIX (1972), 488-497.
- [268] Quine, W. V.: *Wort und Gegenstand (Word and Object)*. Stuttgart: Reclam, 1993.
- [269] Quine, W. V.: ‚Two Dogmas of Empiricism‘. In *Concepts. Core Readings*. Cambridge, Mass.: MIT Press, 1999, 153-169.
- [270] Raatikainen, Panu: ‚On Interpreting Chaitin's Incompleteness Theorem‘. *Journal of Philosophical Logic*, 27 (1998), 569-586.
- [271] Raatikainen, Panu: ‚Algorithmic Information Theory and Undecidability‘. *Synthese*, 123 (2000), 217-225.
- [272] Railton, Peter: ‚Probability, Explanation, and Information‘. *Synthese*, 48 (1981), 223-256.
- [273] Ramsey, William: ‚Prototypes and Conceptual Analysis‘. In *Rethinking Intuition. The Psychology of Intuition and Its Role in Philosophical Inquiry*. DePaul, M. R. und Ramsey, W. (Hrsg.) Lanham: Rowman & Littlefield, 1998, 161-177.
- [274] Rawls, John: *A Theory of Justice*. Cambridge, Mass.: Harvard University Press, 1971.
- [275] Rehder, Wulf: ‚Versuche zu einer Theorie von Gedankenexperimenten‘. *Grazer Philosophische Studien*, 11 (1980), 105-123.
- [276] Reichenbach, Hans: *The Rise of Scientific Philosophy*. Berkeley: University of California Press, 1951.

- [277] Reid, Thomas: *Philosophical Works*. I/II. Hildesheim: Georg Olms, 1983.
- [278] Rescher, Nicholas: ‚Thought Experiments in Presocratic Philosophy‘. In *Thought Experiments in Science and Philosophy*. Horowitz, Tamara und Massey, Gerald J. (Hrsg.) Savage: Rowman & Littlefield, 1991, 31-41.
- [279] Reynolds, Steven L.: ‚Imagining Oneself to be Another‘. *Noûs*, 23 (1989), 615-633.
- [280] Riecke, Eduard: *Lehrbuch der Physik*. Leipzig: Veit & Comp., 1912.
- [281] Roberts, Francis: ‚Thought Experiments and Social Transformation‘. *Journal for the Theory of Social Behavior*, 23 (1993), 399-421.
- [282] Rohrbaugh, Guy und deRosset, Louis: ‚A New Route to the Necessity of Origin‘. *Mind*, 113 (2004), 705-725.
- [283] Russell, Bertrand: ‚On Denoting‘. *Mind*, 14 (1905), 484-485.
- [284] Russell, Bertrand: ‚Alexius Meinong: Über die Stellung der Gegenstandstheorie im System der Wissenschaften‘. *Mind*, n.s. 16 (1906), 436-439.
- [285] Russell, Bertrand: *Human Knowledge. Its Scope and Limits*. London: George Allen and Unwin, 1948.
- [286] Sambursky, S.: *Das physikalische Weltbild der Antike*. Zürich: 1965.
- [287] Schlick, Moritz: ‚Die Wende in der Philosophie‘. *Erkenntnis*, 1 (1930), 4-11.
- [288] Schrenk, Markus: ‚Galileo versus Aristotle on Free Falling Bodies‘. *Philosophiegeschichte und logische Analyse / Logical Analysis and History of Philosophy*, 7 (2004), 81-89.
- [289] Schurz, Gerhard 1993: ‚Koexistenzweisen rivalisierender Paradigma. Eine begriffsklärende und problemtypologisierende Studie‘. In *IPS-Preprints 1993/2*. (Manuscript) Salzburg.
- [290] Schurz, Gerhard: *Wissenschaftliche Erklärung. Ansätze zu einer logisch-pragmatischen Wissenschaftstheorie*. Graz: dbv-Verlag, 1995.
- [291] Schurz, Gerhard: ‚Tarski and Carnap on Logical Truth - or: What is Genuine Logic?‘ In *Alfred Tarski and the Vienna Circle. Austro-Polish Connections in Logical Empiricism*. Wole ski, Jan und Köhler, Eckehart (Hrsg.) Dordrecht: Kluwer Academic Publishers, 1999, 77-94.
- [292] Schurz, Gerhard: ‚Rudolf Carnap’s Modal Logic‘. In *Zwischen traditioneller und moderner Logik*. Stelzner, Werner und Stöckler, Manfred (Hrsg.) Paderborn: Mentis, 2001, 365-380.
- [293] Searle, John R.: *Minds, Brains and Science. The 1984 Reith Lectures*. London: BBC Publications, 1984.
- [294] Searle, John R.: ‚Geist, Gehirn, Programm‘. In *Künstliche Intelligenz. Philosophische Probleme*. Zimmerli, Walter Ch. und Wolf, Stefan (Hrsg.) Stuttgart: Reclam, 1994, 232-265.
- [295] Seddon, George: ‚Logical Possibility‘. *Mind*, LXXXI (1972), 481-494.
- [296] Shoemaker, Sydney: *Self-Knowledge and Self-Identity*. Ithaca: Cornell University Press, 1963.
- [297] Sidelle, Alan: *Necessity, Essence, and Individuation. A Defense of Conventionalism*. Ithaca: Cornell University Press, 1989.
- [298] Sider, Theodore: *Four-Dimensionalism. An Ontology of Persistence and Time*. New York: Oxford University Press, 2001.
- [299] Singer, Peter: *Praktische Ethik*. Stuttgart: Reclam, 1994.

- [300] Smart, J. J. C. und Williams, Bernard: *Utilitarianism: for and against*. London: Cambridge University Press, 1973.
- [301] Snowdon, P. F.: ‚Personal Identity and Brain Transplants‘. In *Human Beings*. Cockburn, David (Hrsg.) Cambridge: Cambridge University Press, 1991, 109-126.
- [302] Soames, Scott: *Reference and Description: The Case Against Two-Dimensionalism*. Princeton: Princeton University Press, 2004.
- [303] Sober, Elliott: ‚The Evolution of Rationality‘. *Synthese*, 46 (1981), 95-120.
- [304] Sorensen, Roy A.: *Thought Experiments*. New York: Oxford University Press, 1992.
- [305] Sorensen, Roy A.: ‚Thought Experiments and the Epistemology of Laws‘. *Canadian Journal of Philosophy*, 22 (1992), 15-44.
- [306] Sorensen, Roy A.: ‚Précis of Thought Experiments‘. *Informal Logic*, 17 (1995), 385-387.
- [307] Sorensen, Roy A.: ‚Sorensen’s Reply to Bunzl and Feldman‘. *Informal Logic*, 17 (1995), 399-405.
- [308] Sorensen, Roy A.: ‚Sören Häggqvist: Thought Experiments in Philosophy‘. *Theoria*, LXIV (1998), 108-118.
- [309] Sorensen, Roy A.: *Vagueness and Contradiction*. New York: Oxford University Press, 2001.
- [310] Stadler, Friedrich: *Studien zum Wiener Kreis. Ursprung Entwicklung und Wirkung des Logischen Empirismus im Kontext*. Frankfurt am Main: Suhrkamp, 1997.
- [311] Stalnaker, Robert: ‚Assertion‘. In *Syntax and Semantics Vol. 9: Pragmatics*. Cole, P. (Hrsg.) New York: 1978, 315-332.
- [312] Stalnaker, Robert und Baldwin, Thomas: ‚On Considering a Possible World as Actual‘. *Proceedings of the Aristotelian Society*, suppl. vol. LXXV (2001), 141-174.
- [313] Stalnaker, Robert C.: *Ways a World Might Be. Metaphysical and Anti-Metaphysical Essays*. Oxford: Clarendon Press, 2003.
- [314] Stäudner, Frank 1998: ‚Virtuelle Erfahrung. Eine Untersuchung über den Erkenntniswert von Gedankenexperimenten und Computersimulationen in den Naturwissenschaften‘. In *Philosophische Fakultät*. Jena: Friedrich-Schiller-Universität.
- [315] Stegmüller, Wolfgang: *Das Problem der Induktion: Humes Herausforderung und moderne Antworten - Der sogenannte Zirkel des Verstehens*. Darmstadt: Wissenschaftliche Buchgesellschaft, 1975.
- [316] Stevin, Simon: *The Principal Works of Simon Stevin*. Amsterdam: C. V. Swets & Zeitlinger, 1955.
- [317] Stich, Stephen P.: ‚Naturalizing Epistemology: Quine, Simon and the Prospects for Pragmatism‘. In *Philosophy and Cognitive Science*. Hookway, C. und Peterson, D. (Hrsg.) Cambridge: Cambridge University Press, 1993, 1-17.
- [318] Stich, Stephen P.: ‚Reflective Equilibrium, Analytic Epistemology and the Problem of Cognitive Diversity‘. In *Rethinking Intuition. The Psychology of Intuition and Its Role in Philosophical Inquiry*. DePaul, M. R. und Ramsey, W. (Hrsg.) Lanham: Rowman & Littlefield, 1998, 95-112.
- [319] Stich, Stephen P. und Nisbett, Richard E.: ‚Justification and the Psychology of Human Reasoning‘. In *Nelson Goodman’s New Riddle of Induction*. Elgin, Catherine Z. (Hrsg.) New York: Garland, 1997, 274-288.

- [320] Stich, Stephen P. und Weinberg, Jonathan M.: 'Jackson's Empirical Assumptions'. *Philosophy and Phenomenological Research*, LXII (2001), 637-643.
- [321] Stöltzner, Michael: 'The Dynamics of Thought Experiments. A Comment on David Atkinson'. In *Observation and Experiment in the Natural and Social Sciences*. Galavotti, M. C. (Hrsg.) Dordrecht: Kluwer Academic Publishers, 2003, 243-258.
- [322] Strawson, Peter F.: *Einzelnding und logisches Subjekt (Individuals)*. Stuttgart: Reclam, 1972.
- [323] Swinburne, Richard G.: 'Personal Identity'. *Proceedings of the Aristotelian Society*, 74 (1973-74), 231-247.
- [324] Thomson, Judith Jarvis: 'A Defense of Abortion'. *Philosophy & Public Affairs*, 1.1 (1971), 47-66.
- [325] Tidman, Paul 1990: 'The Epistemology of Modality'. In *Department of Philosophy*. Notre Dame, Indiana: University of Notre Dame.
- [326] Tidman, Paul: 'Conceivability as a Test for Possibility'. *American Philosophical Quarterly*, 31 (1994), 297-308.
- [327] Tidman, Paul: 'Logic and Modal Intuitions'. *The Monist*, 77 (1994), 389-398.
- [328] Tidman, Paul: 'The Justification of A Priori Intuitions'. *Philosophy and Phenomenological Research*, LVI (1996), 161-171.
- [329] Tittle, Peg: *What If: Collected Thought Experiments in Philosophy*. New York: Pearson/Longman, 2005.
- [330] Turing, A. M.: 'Computing machinery and intelligence'. *Mind*, 59 (1950), 433-460.
- [331] Tversky, Amos und Kahneman, Daniel: 'Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment'. *Psychological Review*, 90 (1983), 293-315.
- [332] Van Cleve, J.: 'Conceivability and the Cartesian Argument for Dualism'. *Pacific Philosophical Quarterly*, 64 (1983), 35-45.
- [333] Van Inwagen, Peter: 'Modal Epistemology'. *Philosophical Studies*, 92 (1998), 67-84.
- [334] Van Lambalgen, Michiel: 'Algorithmic Information Theory'. *Journal of Symbolic Logic*, 54 (1989), 1389-1400.
- [335] Viennot, L.: 'Spontaneous Reasoning in Elementary Dynamics'. *European Journal of Science Education*, 1 (1979), 205-221.
- [336] Ward, David: 'Imaginary Scenarios, Black Boxes and Philosophical Method'. *Erkenntnis*, 43 (1995), 181-198.
- [337] Weber, Max: *Schriften zur Wissenschaftslehre*. Stuttgart: Reclam, 1991.
- [338] Weinberg, Jonathan M., Nichols, Shaun und Stich, Stephen P.: 'Normativity and Epistemic Intuitions'. *Philosophical Topics*, 29 (2001), 429-460.
- [339] Weitz, Morris: 'The Role of Theory in Aesthetics'. *Journal of Aesthetics and Art Criticism*, XV (1956), 27-35.
- [340] Wheeler, John Archibald und Zurek, Wojciech Hubert (Hrsg.): *Quantum Theory and Measurement*. Princeton: Princeton University Press, 1983.
- [341] Wiggins, David: *Identity and Spatio-Temporal Continuity*. Oxford: Blackwell, 1967.
- [342] Wiggins, David: 'Locke, Butler and the Stream of Consciousness: and Men as a Natural Kind'. *Philosophy*, 51 (1976), 131-158.
- [343] Wiggins, David: *Sameness and Substance*. Oxford: Blackwell, 1980.
- [344] Wiggins, David: *Sameness and Substance Renewed*. New York: Cambridge University Press, 2001.

- [345] Wilkes, Kathleen V.: *Real People*. Oxford: Clarendon Press, 1999.
- [346] Williams, Bernard: ‚The Self and the Future‘. In *Personal Identity*. Perry, John (Hrsg.) Berkeley: University of California Press, 1975, 179-198.
- [347] Williams, Bernard: *Probleme des Selbst*. Stuttgart: Reclam, 1978.
- [348] Williamson, Timothy: ‚Armchair Philosophy, Metaphysical Modality and Counterfactual Thinking‘. *Proceedings of the Aristotelian Society*, 105 (2005), 1-23.
- [349] Wilson, Timothy D. und Gilbert, Daniel T.: ‚Affective Forecasting‘. *Advances in Experimental Social Psychology*, 35 (2003), 345-411.
- [350] Witt-Hansen, Johannes: ‚H. C. Ørsted, Immanuel Kant, and the Thought Experiment‘. *Danish Yearbook of Philosophy*, 13 (1976), 48-65.
- [351] Wittgenstein, Ludwig: *Werkausgabe*. 8. Frankfurt am Main: Suhrkamp, 1992.
- [352] Wittgenstein, Ludwig: *Werkausgabe*. 1. Frankfurt am Main: Suhrkamp, 1997.
- [353] Wolff, Michael: *Fallgesetz und Massebegriff*. Berlin: de Gruyter, 1971.
- [354] Woolhouse, R. S.: ‚From Conceivability to Possibility‘. *Ratio*, 14 (1972), 144-154.
- [355] Wright, Crispin: *Truth and Objectivity*. Cambridge, Mass.: Harvard University Press, 1994.
- [356] Wright, Crispin: ‚On Putnam's Proof that We are not Brains in a Vat‘. In *Reading Putnam*. Clark, Peter und Hale, Bob (Hrsg.) Cambridge, Mass.: Blackwell, 1995, 216-241.
- [357] Wright, Crispin: *Wahrheit und Objektivität*. Frankfurt am Main: Suhrkamp, 2001.
- [358] Yablo, Stephen: ‚Review of A. Sidelle: Necessity, Essence, and Individuation. A Defense of Conventionalism‘. *The Philosophical Review*, 101 (1992), 878-881.
- [359] Yablo, Stephen: ‚Is Conceivability a Guide to Possibility?‘ *Philosophy and Phenomenological Research*, LIII (1993), 1-42.
- [360] Yablo, Stephen: ‚Textbook Kripkeanism and the Open Texture of Concepts‘. *Pacific Philosophical Quarterly*, 81 (2000), 98-122.
- [361] Yablo, Stephen: ‚Coulda, Woulda, Shouda‘. In *Conceivability and Possibility*. Gendler, Tamar Szabó und Hawthorne, John (Hrsg.) Oxford: Clarendon Press, 2002, 441-492.