Hans Rott

Reduktion und Revision

Aspekte des nichtmonotonen Theorienwandels

PETER LANG
Frankfurt am Main · Bern · New York · Paris

CIP-Titelaufnahme der Deutschen Bibliothek

Rott, Hans

Reduktion und Revision: Aspekte des nichtmonotonen Theorienwandels / Hans Rott. - Frankfurt am Main ; Bern ; New York ; Paris : Lang, 1991
(Europäische Hochschulschriften : Reihe 20, Philosophie ; Bd. 290)
ISBN 3-631-42080-3

NE: Europäische Hochschulschriften / 20

D 19
ISSN 0721-3417
ISBN 3-631-42080-3

Printed in Germany 1 2 3 5 6 7

Our scholastic headpieces and logicians shew no such superiority above the mere vulgar in their reason and ability, as to give us any inclination to imitate them in delivering a long system of rules and precepts to direct our judgment, in philosophy. All the rules of this nature are very easy in their invention, but extremely difficult in their application; and even experimental philosophy, which seems the most natural and simple of any, requires the utmost stretch of human judgment. There is no phænomenon in nature, but what is compounded and modify'd by so many different circumstances, that in order to arrive at the decisive point, we must carefully separate whatever is superfluous, and enquire by new experiments, if every particular circumstance of the first experiment was essential to it. These new experiments are liable to a discussion of the same kind; so that the utmost constancy is requir'd to make us persevere in our enquiry, and the utmost sagacity to choose the right way among so many that present themselves. ... I am much afraid, lest the small success I meet with in my enquiries will make this observation bear the air of an apology rather than of boasting.

David Hume
A Treatise of Human Nature, Book I, Part III, Section XV

Vorwort

Bei dem vorliegenden Buch handelt es sich um eine leicht überarbeitete Fassung meiner Dissertation vom Oktober 1988, die am Seminar für Philosophie, Logik und Wissenschaftstheorie der Ludwig-Maximilians-Universität München entstanden ist. Frühere, meist kürzere Fassungen einiger Teile erschienen oder erscheinen an anderer Stelle. Dies trifft zu auf die Abschnitte 1.1–1.7 [als 'Reduction: Some Criteria and Criticisms of the Structuralist Concept' in *Erkenntnis* **27** (1987), S. 231–256], auf Kapitel 4 [als 'Conditionals and Theory Change: Revisions, Expansions, and Additions' in *Synthese* **81** (1989), S. 91–113], auf die Abschnitte 6.1–6.7 [als 'Ifs, Though, and Because' in *Erkenntnis* **25** (1986), S. 345–370], auf die Abschnitte 7.1–7.5 [als 'On Relations Between Successive Theories' in *Abstracts of the 8th International Congress of Logic, Methodology and Philosophy of Science*, Band 4.2, Moskau: Nauka 1987, S. 123–126], und auf die Teilkapitel 8.1 [als 'Approximation Versus Idealization: The Kepler-Newton Case' in *Idealization-II: Forms and Applications*, hrsg. v. J. Brzeziński, F. Coniglione, T. A. F. Kuipers und L. Nowak (= *Poznań Studies in the Philosophy of the Sciences and the Humanities* **17**), Amsterdam: Rodopi 1990, S. 101–124] und 8.2 [als 'Approximation versus Idealisierung: Das Verhältnis zwischen idealem und van der Waalsschem Gasgesetz' in *Philosophie der Naturwissenschaften — Akten des 13. Internationalen Wittgenstein Symposiums (1988)*, hrsg. v. Paul Weingartner und Gerhard Schurz, Wien: Hölder-Pichler-Tempsky 1989, S. 224-228]. Eine elegantere, aber weniger motivierende Darstellung der wesentlichen Ergebnisse von Kapitel 3 wird unter dem Titel 'Two Methods of Constructing Contractions and Revisions of Knowledge Systems' im *Journal of Philosophical Logic* erscheinen.

Ich möchte mich bei meinem Doktorvater, Herrn Professor Stegmüller, herzlich dafür bedanken, daß er mir die Freiheit ließ, die Dissertation nach meinem Gutdünken zu gestalten. Besonderer Dank gebührt auch dem Kor-

referenten Wolfgang Spohn, der mir über viele Jahre hinweg mit freundschaftlichem und höchst kompetentem Rat zur Seite stand. Ich danke Peter Gärdenfors, der mir stets großzügig unveröffentlichtes Material und wichtige Anregungen zukommen ließ. Für wertvolle Kommentare und Diskussionen zu einzelnen Teilen dieses Buchs bin ich Wolfgang Balzer, André Fuhrmann, Theo Kuipers, Isaac Levi, David Makinson, Felix Mühlhölzer und David Pearce zu großem Dank verpflichtet. Dankbar bin ich auch Hans Kamp und dem Stuttgarter Institut für Maschinelle Sprachverarbeitung für die Möglichkeit, ein druckreifes Manuskript herzustellen. Nicht vergessen möchte ich, die finanzielle Unterstützung meiner Dissertation durch die Studienstiftung des deutschen Volkes zu erwähnen.

Den meisten Dank schulde ich endlich meiner Familie: meinen Eltern, meiner Frau Ingrid und unserer kleinen Jana. Sie haben mir unermeßlich viel gegeben. Wenn ich der Ansicht wäre, daß Dissertationen jemandem gewidmet werden sollten, wüßte ich nicht, wem von ihnen ich dieses Buch widmete. Weil ich dieser Ansicht aber nicht bin, weiß ich es.

Stuttgart, im September 1990 H. R.

Inhalt

Vorwort 7

Inhalt 9

Einleitung 13

Kapitel 1. Der strukturalistische Begriff der striken Reduktion 19

1.1 Die Idee der strukturalistischen Reduktion 19
1.2 Definierbarkeit und Ableitbarkeit, intendierte Anwendungen und Modelle: Reduktion nach Adams und Sneed 23
1.3 Spezialisierungen und Anomalien: der Vorschlag von D. Mayr . 26
1.4 Wahrheit und Übersetzung : der Vorschlag von D. Pearce. 30
1.5 Empirische Behauptungen und erlaubte Modelle: der Vorschlag von A. Kamlah 34
1.6 Erklärungen: der Vorschlag von T. Mormann 40
1.7 Zwischenbilanz 42
1.8 Typen von Reduktion I: Pearce und Rantala über theoretische und erklärende Reduktion 47
1.9 Typen von Reduktion II: Scheibe über (bereichseinschränkend-)deduktive und empirische Erklärung von Theorien 57
1.10 Ein Miniaturbeispiel: Bereichseinschränkungen und kontrafaktische Konditionalsätze 65

Kapitel 2. Der klasssische Begriff der Reduktion 75

2.1 Ableitbarkeit und Definierbarkeit: das Originalkonzept von Nagel und Hempel 75
2.2 Relativierung auf Beobachtungsdaten: die Kritik von Kemeny und Oppenheim 78
2.3 Inkonsistenz und Inkommensurabilität, die beiden Arten der Nichtmonotonie: die fundamentale Kritik von Feyerabend 80
2.4 Inkonsistenz, Approximation und Kontrafaktizität: Sklar, Schaffner, Fine, Glymour, Eberle und Nickles 93
2.4.1 Bereichseinschränkung, Approximation und Kontrafaktizität ... 95
2.4.2 Reduktion, Idealisierung und Erklärung 106

Kapitel 3. Kontrafaktische Annahmen: Zum Modell der Theorienrevision nach Gärdenfors 113

3.1 Theorienrevisionen und -kontraktionen: die Gärdenfors-Postulate ... 113
3.2 Die Relation der theoretischen Wichtigkeit ... 120
3.3 Der erste Weg zu eindeutigen Kontraktionen und Revisionen: Durchschnitte maximalkonsistenter Teilmengen .. 125
3.4 Der zweite Weg zu eindeutigen Kontraktionen und Revisionen: eine direkte Konstruktion ... 131
3.5 Der Zusammenhang zwischen dem ersten und dem zweiten Weg ... 134
3.6 Ein Ansatz zu einer Dynamik der theoretischen Wichtigkeit .. 139
3.7 Anhang: Beweise ... 141

Kapitel 4. Konditionalsätze und Theorienwandel: Revisionen, Expansionen und Additionen 157

4.1 Theorienrevisionsmodelle ... 157
4.2 Konditionalsätze und Gärdenfors' Trivialisierungstheorem 160
4.3 Expansionen und Additionen ... 165
4.4 Autoepistemische Allwissenheit ... 168
4.5 Möglicherweise-Konditionalsätze ... 172
4.6 Warum Additionen nicht parasitär von Expansionen abhängen ... 174
4.6.1 Der Realitätsgehalt modalisierter Sätze ... 175
4.6.2 Eine Abhängigkeit nichtmodalisierter von modalisierten Sätzen ... 177
4.7 Konditionalsätze und Nichtmonotonie ... 178

Kapitel 5. Nichtmonotone und autoepistemische Logik 183

5.1 Schlüsse aus der Unwissenheit ... 183
5.2 Nichtmonotone Logik I ... 187
5.3 Nichtmonotone Logik II: Nichtmonotone Modallogik ... 190
5.4 Autoepistemische Logik ... 194
5.5 Mögliche-Welten-Semantik für autoepistemische Logik ... 198
5.6 Kritische Beispiele ... 200
5.6.1 Die autoepistemische Logik liefert zu wenig Schlußfolgerungen 201
5.6.2 Die autoepistemische Logik liefert zu viele Schlußfolgerungen . 203

5.7 Autoepistemische Logik ohne Fundiertheit 206
5.8 Fortschritte der autoepistemischen Logik 207

Kapitel 6. Konditionalsätze und Erklärungen 211

6.1 Wenn-dann und weil 211
6.2 Konditionalsätze und Erklärungen: der Ausgangspunkt der Analyse 215
6.3 Die Normalanalyse und der Starke Ramsey-Test 217
6.4 Eine alternative Analyse von weil 220
6.5 Universelle Konditionale 223
6.6. Natürlichsprachliche Konditionale: Thesen 225
6.7 Natürlichsprachliche Konditionale: Akzeptabilitätsbedingungen 226
6.8 Vergleich mit der Theorie von McCall 234

Kapitel 7. Intertheoretische Erklärungen, gute und überlegene Nachfolgertheorien 245

7.1 Kontinuität und Widerspruch 245
7.2 Definitionen 248
7.3 Literaturverweise 251
7.4 Analyse 253
7.5 Eine intuitive Beschränkung für minimale Revisionen 257
7.6 Idealisierung und Approximation 260

Kapitel 8. Approximation versus Idealisierung: zwei Fallbeispiele 263

8.1 Die Keplerschen Gesetze und Newtons Gravitationstheorie 265
8.1.1 Die Inkonsistenz zwischen den Keplerschen Gesetzen und Newtons Gravitationstheorie 265
8.1.2 Das Kepler-Newton-Beispiel als ein Fall von Approximation .. 270
8.1.3 Das Kepler-Newton-Beispiel als ein Fall von Idealisierung 280
8.1.3.1 Erste Idealisierung: Einkörpersysteme 284
8.1.3.2 Zweite Idealisierung: Zweikörpersysteme 286
8.1.4 Idealisierung und Approximation im Kepler-Newton-Fall 292
8.2 Das ideale und das van der Waalssche Gasgesetz 295
8.2.1 Die Inkonsistenz der idealen und der van der Waalsschen Gastheorie 295
8.2.2 Das Beispiel des idealen und des van der Waalsschen Gasgesetzes als ein Fall von Approximation 299

8.2.3 Das Beispiel des idealen und des van der Waalsschen Gasgesetzes als ein Fall von Idealisierung 304
8.2.4 **KGP — NTG** und **IGG — VDW**: Vergleich zweier intertheoretischer Idealisierungen 309
8.2.5 Quantitative Ähnlichkeiten und qualitative Unterschiede 313

Kapitel 9. Über Idealisierung in der Wissenschaft **319**

9.1 Idealisierungen in der Wissenschaftstheorie 319
9.2 Zwei Arten von kontrafaktischen Konditionalsätzen und Essentialismus in Naturgesetzen 324
9.3 Idealisierungen und Theoriendynamik 329
9.4 Wie rechtfertigt man Idealisierungen? 334
9.5 Was ist eine wissenschaftliche Theorie? 338

Literaturverzeichnis **343**

Einleitung

Je länger man forscht, desto mehr lernt man dazu. Je neuer eine Theorie ist, desto mehr Wissen steckt in ihr. Soll heißen: spätere Theorien fügen zum Wissen der früheren Theorien lediglich etwas hinzu. Das ist das naive Bild der Wissenschaft, nach welchem die Wissenschaft kumulativ, additiv, linear oder, wie ich sagen werde, *monoton* fortschreitet. Obgleich wohl niemand, der sich ernsthaft mit der Sache beschäftigt hatte, so naiv war, sich dieses Bild wirklich zu eigen zu machen, war es doch schockierend, als zu Beginn der 60er Jahre einige Wissenschaftsphilosophen zu zeigen versuchten, daß das Bild nicht stimmen *könne*. Vor allem Kuhn und Feyerabend wurden berühmt für die These, daß einander ablösende Theorien inkommensurabel seien. Solche Theorien, heißt das, erlauben überhaupt keinen direkten inhaltlichen Vergleich, sie sind begrifflich unvereinbar und deshalb nicht ineinander übersetzbar. Demnach kann es kein echtes Mehr im Theorienwandel geben, Theorienwandel ist wesentlich nichtmonoton. Diese *Nichtmonotonie im ersten Sinn* bedeutete eine ungeheure Herausforderung, schien sie doch jegliche Art von Rationalität und Fortschritt beim Übergang von einer Theorie zu ihrer Nachfolgertheorie in Frage zu stellen. Nichts Geringeres als die Existenzberechtigung der Wissenschaftstheorie stand auf dem Spiel, die ja gerade versuchte, Wissenschaft als rational und als fortschrittlich zu rekonstruieren und eine Logik des Theorienwandels herauszufinden.

Die Wissenschaftstheorie hat diese Herausforderung angenommen. Eine wichtige Erwiderung wurde von Professor Stegmüller gegeben. Bei der Formulierung seiner *universellen Reduktionsthese*

> If it was T_2 which dislodged T_1, then we speak of *progress* only if T_1 is reducible to T_2 ... Therewith the concept of reduction became the cornerstone to explain scientific rationality.

(Stegmüller 1979, S. 71)[1] greift er durchaus auf überliefertes Gedankengut zurück.[2] Der Grund, warum die Kuhn-Feyerabendschen Argumente Stegmüllers Vorschlag nicht treffen, ist ein neuer Theorienbegriff, der eine wissenschaftliche Theorie nicht wie üblich als Menge von Aussagen („Statement view"), sondern als mathematische Stuktur ansieht („Strukturalismus" oder „Non-statement view"). Entscheidend ist die Idee, daß eine intertheoretische Relation — oder genauer: eine Nachfolgerrelation zwischen Theorien — zur Explikation von Fortschritt und Rationalität in der Wissenschaft dienen soll. Wir werden uns in Kapitel 1 dieses Buchs „die" strukturalistische Antwort auf die These der Nichtmonotonie im ersten Sinn genauer betrachten. Es wird sich herausstellen, daß kein einheitliches *strukturalistisches Konzept der (strikten) Reduktion* auszumachen ist, sondern daß viele verschiedene, in verschiedenen Kombinationen miteinander unverträgliche Kriterien für Reduktionen konkurrieren. Auch der Versuch, den Reduktionsbegriff in zwei logische Typen von Reduktion aufzuspalten, wird als wenig überzeugend zurückgewiesen werden.

Auf der Suche nach einem passenden Reduktionsbegriff gehen wir in Kapitel 2 zurück zu den Wurzeln und sehen uns noch einmal an, was aus der *klassischen Konzept der Reduktion* geworden ist. Es hatte Reduzierbarkeit mit Definierbarkeit-plus-Deduzierbarkeit gleichgesetzt und wurde Opfer der Feyerabendschen Attacken: Aufeinander folgende Theorien wären irreduzibel, weil sie erstens miteinander inkonsistent und zweitens inkommensurabel seien. Eine genauere Betrachtung der Argumente des frühen Feyerabend wird erweisen, daß sich Inkommensurabilität in seinem Sinne hauptsächlich oder sogar ausschließlich auf Inkonsistenz gründet. Das impliziert aber eine *zweite Art von Nichtmonotonie*, denn wenn man zuerst die Theorie T_1 und später die mit T_1 logisch widersprüchliche Theorie T_2 vertritt, dann muß man in T_2 etwas aus T_1 als falsch eingesehen und gestrichen haben. In T_2 ist also nicht eigentlich *mehr* Information als in T_1 enthalten, sondern T_2 korrigiert T_1. Ende der 60er/Anfang der 70er Jahre hatte man verschiedenerlei Vorstellungen, wie man mit dem Inkonsistenzeinwand zu Rande kommen könnte. Ideen der approximativen Reduzierbarkeit wurden seitdem mehr oder weniger systematisch verfolgt. Doch es wurde auch die Ansicht vorgebracht, daß T_1 unter bestimmten kontrafaktischen Annahmen A aus T_2 ableitbar ist, und auch die ganz ähnliche Idee,

[1] Vgl. auch Stegmüller (1979, S. 35–37, 68f, 78) und Stegmüller (1985, S. 14, 249f, 255, 283).

[2] Vgl. den Überblick über die These „Fortschritt geschieht durch Reduktion" in Suppe (1974, S. 53–56, 170–175).

daß der kontrafaktische Konditionalsatz Wenn A der Fall wäre, dann wäre T_1 korrekt „von T_2 aus gesehen" zutreffend ist. Da die formalen Mittel zur Bewältigung dieser Vorschläge seinerzeit noch nicht zur Verfügung standen, sind sie aber im Sande verlaufen.

Heutzutage kann man diese Ideen jedoch wieder in Schwung bringen. Dazu muß man sich freilich einige Einschränkungen auferlegen. Für die logische Analyse der Beziehung zwischen einander ablösenden Theorien T_1 und T_2 werden wir von den Problemen der Nichtübersetzbarkeit und der Inkommensurabiltät ganz absehen und uns nur mehr der Inkonsistenz, also der Nichtmonotonie im zweiten Sinn zuwenden.[3] Wir verwenden den idealisierten Theorienbegriff des Logikers, der Theorien als deduktiv abgeschlossene Satzmengen betrachtet. In Kapitel 3 untersuchen wir dann ein von Peter Gärdenfors stammendes Modell für den Umgang mit *kontrafaktischen Annahmen*. Besondere Aufmerksamkeit schenken wir der Relation der theoretischen Wichtigkeit, die es ermöglicht, minimale Revisionen von Theorien aufgrund kontrafaktischer (oder besser: kontratheoretischer) Annahmen mehr oder weniger konstruktiv darzustellen. Das wichtigste Ergebnis dieses eher technischen Kapitels ist es, daß Revisionen gemäß der theoretischen Wichtigkeit sozusagen äquivalent sind mit Revisionen, die man über Durchschnitte maximalkonsistenter Teilmengen erhält.

In Kapitel 4 werden wir sehen, daß *kontrafaktische Konditionalsätze* über den sog. Ramsey-Test ganz eng an Theorienrevisionen durch kontrafaktische Annahmen angebunden werden können. Gärdenfors hat aber vor kurzem ein Trivialisierungstheorem für diesen Ansatz geliefert. Wir werden einen einfachen Beweis dieses Theorems betrachten und uns überlegen, worin die Ursache für dieses mißliche Ergebnis zu sehen ist. Es wird sich als die natürlichste Lösung zeigen, daß es in Sprachen, die Konditionalsätze enthalten, vermutlich kein monotones Dazulernen mehr gibt und daß dort wahrscheinlich jede Art von Neuinformation mit alten Überzeugungen in Konflikt gerät. In einem ganz präzisen Sinn scheint sowohl die „Addition" von Hypothesen oder Neuinformationen als auch die Logik, nach der man in solchen Sprachen aus vorgegebenen Prämissen Schlüsse gezogen werden, nichtmonoton zu sein.

Damit ist die Bedeutung der sogenannten *nichtmonotonen und autoepistemischen Logiken*, bei denen aus mehr Axiomen nicht unbedingt mehr Theoreme abgeleitet werden können, für den gegenwärtigen Zusammen-

[3] Zum Inkommensurabilitätsproblem vgl. Pearce (1987) und Schroeder-Heister und Schaefer (1989), die die Meinung vertreten, daß Reduzierbarkeit eine Art von Übersetzbarkeit, also auch Kommensurabilität impliziert.

hang evident. Solche Logiken bieten eine Modellvorstellung dafür an, wie es passieren kann, daß eine Theorie T_1, in der man „weniger" weiß, mit einer Nachfolgertheorie T_2, die mehr Information enthält, inkonsistent ist. In Kapitel 5 werden wir die hier entwickelten formalen Systeme, die alle aus der KI-Forschung stammen, vorstellen und diskutieren. Trotz der ganz offensichtlichen Relevanz für unser Thema wird die Besprechung des nichtmonotonen Schließens letztlich eher Exkurscharakter haben. Denn die Grundlagen und die logischen Eigenschaften der einschlägigen Systeme sind äußerst verzwickt und zumindest derzeit noch viel zu unklar, als daß man mit ihnen zu wissenschaftstheoretisch tragfähigen Begriffen und Ergebnissen kommen könnte.

Ich werde also über einen formal einfachen, aber philosophisch bedenklichen Trick (keine Übernahme von „modalisierten" Sätzen in „Expansionen") die Nichtmonotonie von Konditionalsätzen ausblenden und in Kapitel 6 zur *Analyse verschiedener Arten von natürlichsprachlichen Konditionalsätzen* schreiten. Indem man der vereinfachenden These nachgeht, daß weil-Sätze mit Erklärungen identisch sind, gerät insbesondere der Zusammenhang von (indikativischem und konjunktivischem) wenn-dann und weil in den Brennpunkt des Interesses. Der Einbau einer Relevanzbedingung in den Ramsey-Test wird der Schlüssel zu einer vereinheitlichten Interpretation einer ganzen Anzahl von Konditionalsätzen (im weiteren Sinn) sein.

Im zentralen Kapitel 7 wird diese Interpretation dann zur Analyse von *intertheoretischen Erklärungen* verwendet, die als ungefähr synonym mit Reduktionsrelationen verstanden werden. Intertheoretische Erklärungen im hier eingeführten Sinn sollen verständlich machen, wie eine Theorie T_2 ihrer Vorgängertheorie T_1 widersprechen und dennoch eine weitreichende Kontinuität im Übergang von T_1 auf T_2 — was für den Fortschrittsgedanken Voraussetzung ist — bestehen kann. Der entscheidende Punkt ist, daß gute und überlegene Nachfolgertheorien T_2 faktisch genaugenommen oft das Scheitern ihrer Vorgänger T_1 erklären, während sie die Theorie T_1 selbst „nur" kontrafaktisch oder als Idealsierung erklären. Wenn gewisse idealisierte Bedingungen erfüllt wären, dann wäre T_1 korrekt, gesteht der T_2-Theoretiker in diesem Falle zu, oder, was nach Kapitel 6 dasselbe heißt: Die minimale Abänderung von T_2, die nötig ist, um die idealisierten Bedingungen in T_2 mit aufzunehmen, gestattet die Ableitung von T_1. Als Alternative zu Stegmüllers universeller Reduktionsthese kann man die *Revisionsthese* formulieren, daß bei fortschrittlichem Theorienwandel die Vorgängertheorie T_1 stets per geeigneter Revision der Nachfolgertheorie

T_2 erreichbar ist.[4]

Zwei einfache und vielzitierte Fallbeispiele sind Gegenstand von Kapitel 8: In 8.1 wird das Verhältnis der *Keplerschen Gesetze* zur *Newtonschen Gravitationstheorie* untersucht, in 8.2 wende ich mich dem *idealen* und dem *van der Waalsschen Gasgesetz* zu. Hierbei skizziere ich, wie das Modell aus Kapitel 7 zur Anwendung kommen kann.[5] Besonderes Gewicht messe ich der Tatsache bei, daß der Ansatz der kontrafaktischen oder idealisierenden Erklärung eine zur approximativen Erklärung wirklich alternative Sichtweise darstellt und auch andere, interessante Ergebnisse liefert. Allerdings ist in Kapitel 8 noch keine detailliert ausgeführte Anwendung der Relation der theoretischen Wichtigkeit inbegriffen.

Nicht von den Beispielen, sondern von der Literatur ausgehend, werde ich in Kapitel 9 einige abstrakte Überlegungen zum Stellenwert von *Idealisierungen* in der Wissenschaft anstellen. Ich betrachte, welch vielfältige Rolle kontrafaktische Konditionalsätze in Theorien spielen, und schlage am Ende vor, daß eine wissenschaftliche Theorie im Statement view als eine Satzmenge *zusammen mit einer Struktur der theoretischen Wichtigkeit* verstanden werden sollte.

Das vorliegende Buch läßt sich in drei größere Teile untergliedern. Kapitel 1–2 diskutieren den Reduktionsbegriff und sind somit zur Wissenschaftstheorie zu zählen. Eher zur philosophischen Logik gehört der um den Revisionsbegriff zentrierte Mittelteil mit den Kapiteln 3–6. Schließlich wird in den Kapiteln 7–9 versucht, die im Mittelteil erarbeiteten Begriffe als für die Wissenschaftstheorie fruchtbar zu erweisen. Der ursprüngliche Plan, ein fertiges Revisionsmodell wissenschaftstheoretisch anzuwenden, konnte nur zu einem kleinen Teil realisiert werden, vor allem weil der Revisionsbegriff doch noch sehr viele und, wie ich meine, sehr interessante Probleme aufgeworfen hat. Folglich ist das Buch nicht so homogen, wie man es sich wünschen würde. Sie beleuchtet recht verschiedene *Aspekte* des nichtmonotonen Theorienwandels und hätte ihr Ziel erreicht, wenn sie die Kluft zwischen Wissenschaftstheorie und philosophischer Logik um ein paar Zentimeter kleiner erscheinen ließe.

---

[4] Ein vergleichbarer Ansatz wurde unabhängig von Rantala (1987; 1988) entwickelt.

[5] Obwohl die Beispiele einfach sind und längst nicht alle Fragen geklärt werden, ist Kapitel 8 relativ lang. En-passant-Beispielsdiskussionen sind in der Wissenschaftstheorie zwar sehr verbreitet, aber meines Erachtens nicht viel wert.

Kapitel 1

Der strukturalistische Begriff der strikten Reduktion

1.1 Die Idee der strukturalistischen Reduktion

Die von Sneed und Stegmüller ausgelöste strukturalistische Wende der Wissenschaftstheorie vor allem in Westdeutschland war unter anderem insofern sehr wirkungsvoll, als der Begriff der Reduktion hier seit Mitte der 70er Jahre beinahe ausschließlich im strukturalistischen Rahmen diskutiert worden ist. Erinnern wir uns noch einmal daran, daß der Strukturalismus als eine Reaktion auf die Herausforderungen von Kuhn und Feyerabend angesehen werden kann, die die einst stillschweigend vorausgesetzte begriffliche Vergleichbarkeit oder die „Kommensurabilität" von einander verdrängenden wissenschaftlichen Theorien stark in Frage stellten. Wissenschaftlicher Wandel kann natürlich nicht „monoton" sein, wenn nicht einmal auf der Beobachtungsebene ein informativer Vergleich zwischen Nachfolgertheorien möglich ist. Das Problem von Bedeutungsverschiebungen wissenschaftlicher Ausdrücke wurde im Strukturalismus dadurch umgangen, daß man einfach davon absah, die Sprache von wissenschaftlichen Theorien explizit zu beschreiben und sich allein auf Stukturen im mathematischen Sinne

konzentrierte. Der Reduktionsbegriff in diesem Non-statement view soll helfen, letztendlich doch Kontinuität und Fortschritt im Theorienwandel erkennbar werden zu lassen.

Man sollte sich aber davor hüten, die Kluft zum traditionellen „Statement view" der Wissenschaftstheorie größer zu machen, als sie tatsächlich ist. Die gegenseitige Übersetzbarkeit von Statement und Non-statement view geht viel weiter als manchmal vermutet worden war.[1] Insofern sind die folgenden Überlegungen weitestgehend übertragbar und also auch für den von Interesse, der sich der strukturalistischen Sichtweise nicht so sehr verpflichtet fühlt. Speziell für die Reduktionsdiskussion ist außerdem zu festzuhalten, daß Adams, der Pionier des strukturalistischen Reduktionsbegriffs, ausdrücklich die Adäquatheitsbedingungen der für den Statement view geradezu paradigmatischen Nagel-Hempelschen (D-)Reduktion zu erfüllen trachtet.[2] Es sei gleich zu Beginn betont, daß wir uns in diesem Kapitel auf den Begriff der *strikten* Reduktion beschränken werden und erst später (in den Kapiteln 8 und 9) auf Approximationen zu sprechen kommen werden.

Das strukturalistische Bild der Dynamik wissenschaftlicher Theorien, speziell der Reduktionsbeziehung, wurde in ziemlich regelmäßigen Abständen kritisiert. In meinen Augen sind die Beiträge von Mayr (1976), Tuomela (1978), Niiniluoto (1980), Pearce (1982), Hoering (1984), Kamlah (1985) und Mormann (1984) besonders zu beachten. Die Darstellungen der Vorreiter des Strukturalismus sowie die ihrer Kritiker haben allerdings den Nachteil, daß praktisch jeder Aufsatz seine eigene, neue Notation einführt und kleinere Abänderungen an zentralen Definitionen vornimmt. Dies und die relativ große Komplexität des strukturalistischen Apparates führt dazu, daß man Gefahr läuft, die Kontinuität der Diskussion aus den Augen zu verlieren. Im folgenden werde ich mich den Fragen zuwenden, wie „der" strukturalistische Reduktionsbegriff aussieht und wie die verstreuten Kriterien und Kritiken dieses Begriffs im Zusammenhang zu bewerten sind. Hierzu verwende ich eine einheitliche Notation und vereinfachte Definitionen, welche im wesentlichen den Zweck verfolgen, das raffinierte Instrumentarium der Sneedschen Behandlung theoretischer Terme herauszulassen. Auch wenn die Ergänzung der folgenden Definitionen durch dieses

[1] Vgl. zum Beispiel Tuomela (1978, S. 222f), Niiniluoto (1980, S. 25), Pearce (1981, S. 24f, 29f; 1982, S. 330). Auch bei Stegmüller (1979, S. 48, 87f) kann man ähnliche Bemerkungen finden.

[2] Zum Verhältnis der Reduktionsbegriffe nach Suppes (1957) und Adams (1959) siehe Abschnitt 1.7; zur Nagel-Hempelschen „D-Reduktion" siehe Kapitel 2.

Instrumentarium wohl keine prinzipiellen Probleme aufwürfe, so würde es doch ein unmittelbares Verstehen der Zusammenhänge erschweren. Ich werde, wie zum Beispiel Kamlah, auf den Adamsschen Vorschlag zurückgehen und unter Vernachlässigung von theoretischen Termen und Constraints eine Theorie T als ein geordnetes Paar $\langle M,I\rangle$ auffassen, wobei M die Klasse der Modelle von T und I die Klasse der intendierten Anwendungen von T ist. M und I sind immer als nichtleere Klassen von Strukturen desselben Ähnlichkeitstyps τ gedacht, und die Klasse M_p der potentiellen Modelle von T sei einfach mit der Klasse aller Stukturen vom Typ τ (modulo einer geeigneten Logik Cn) identifiziert.[3]

Seien nun $T_1=\langle M_1,I_1\rangle$ und $T_2=\langle M_2,I_2\rangle$ zwei Theorien. Die Quintessenz der strukturalistischen Definition der Reduzierbarkeit von T_1 auf T_2 ist die Existenz einer sogenannten *Reduktionsrelation*, die man am bequemsten als eine Funktion[4] der Form

(S) $\quad \mathbf{F}\colon M^{\circ}_{p2}\rightarrow M_{p1}$, wobei $M^{\circ}_{p2}\subseteq M_{p2}$.

wiedergibt. Im allgemeinen wird gefordert, daß **F** *surjektiv* ist und daß M°_{p2} eine *echte* Teilmenge von M_{p2} ist;[5] ich werde im folgenden kenntlich machen, wo diese Forderungen verwendet werden. Wir können Bilder unter **F**

$$\mathbf{F}[X_2] := \{x_1\in M_{p1}\colon \exists x_2\in M^{\circ}_{p2}\cap X_2(x_1=\mathbf{F}(x_2))\}$$

für *jedes* $X_2\subseteq M_{p2}$ (nicht unbedingt $X_2\subseteq M^{\circ}_{p2}$) und ebenso Urbilder unter **F**

$$\mathbf{F}^{-1}[X_1] := \{x_2\in M^{\circ}_{p2}\colon \mathbf{F}(x_2)\in X_1\}$$

für jedes $X_1\subseteq M_{p1}$ und $\mathbf{F}^{-1}(x_1):=\mathbf{F}^{-1}[\{x_1\}]$ für jedes $x_1\in M_{p1}$ definieren.

[3] Dies sind nicht die urspünglichen strukturalistischen Ideen, sondern Verbesserungen, die zuerst von Veikko Rantala vorgeschlagen wurden. Vgl. Niiniluoto (1980, S. 9–11), Pearce (1982, S. 312f) und den programmatischen Aufsatz von Pearce und Rantala (1983a), wo man insbesondere die Begriffe „Ähnlichkeitstyp" und „allgemeine Logik" erklärt finden kann.

[4] In letzter Zeit wurde betont, daß die Funktionalität der Reduktionsrelation wohl als idealisierende, vereinfachende Annahme betrachtet werden muß (siehe Balzer, Moulines und Sneed 1987, S. 271f, 276). Ich gehe davon aus, daß diese Idealisierung im folgenden nichts schadet.

[5] Für die Motivation dieser Forderungen siehe zum Beispiel Stegmüller (1985, S. 145). Bei Adams (1959) fehlen noch beide Bedingungen. Bei Sneed (1971, S. 221) wird die Surjektivität von **F** nicht verlangt, erst in Sneed (1976, S. 122, 136f) gibt es diese Forderung. Kamlah (1985, S. 133) postuliert $M^{\circ}_{p2}=M_{p2}$, aber nicht die Surjektivität von **F**. In den Anwendungen von Pearce und Rantala sind beide Bedingungen zu finden, aber stets ohne den Index „p", vgl. z.B. Pearce und Rantala (1984b, S. 171f) und Abschnitt 1.8 unten. Vergleiche auch die Definitionen DVI-5 und DVI-6 in Balzer, Moulines und Sneed (1987, S. 277).

Klassen solcher Art können sehr leicht zur Formulierung suggestiver Bedingungen benutzt werden.

Bevor wir ins Detail gehen, möchte ich einen grundlegenden Einwand gegen die strukturalistische Idee vorwegnehmen. Wenn Adams und Sneed von der Existenz einer Funktion **F** sprechen, so meinen sie dies nicht im rein mathematischen Sinn. Eine solche Funktion **F** soll nämlich auch den intuitiven Anspruch erfüllen, daß **F** jedem potentiellen Modell x_2 der reduzierenden Theorie T_2 — sofern x_2 aus M^o_{p2} ist — ein potentielles Modell x_1 der reduzierten Theorie T_1 zuordnet, welches, sozusagen von außen betrachtet, „mit x_2 identisch" oder „aus den Individuen von x_2 zusammengesetzt" ist. Das Feststellen, ob zwei potentielle Modelle de facto identisch sind oder ob die Objekte des einen aus den Objekten des anderen zusammengesetzt sind, ist eine empirische Angelegenheit, welche natürlich mit Problemen behaftet ist, die sich einer rein formalen Analyse entziehen. Dementsprechend sind die formalen Kriterien von Adams und Sneed, ebenso wie alle anderen in diesem Kapitel diskutierten, nur als notwendige Bedingungen für das intuitive Bestehen einer Reduktion zwischen zwei Theorien gemeint.[6] Die Rede von der Existenz einer Funktion **F** soll also auch im folgenden voraussetzen, daß diese Funktion den eben erwähnten intuitiven Anspruch erfüllt.[7] Damit blende ich die Debatte mit all jenen Kritikern aus, die die strukturalistische Idee der Reduktion als viel zu schwach und von vornherein unannehmbar betrachten, weil die Existenz einer wie in (S) angegebenen Funktion (mit Eigenschaften der unten diskutierten Art) höchstens Kardinalitätsabschätzungen liefern kann, was für eine echte Reduktion sicherlich völlig ungenügend ist.[8]

Nun ist der Weg bereitet für eine genauere Sichtung der Kriterien, die Reduktionen nach Ansicht verschiedener Strukturalisten erfüllen sollen, und ihrer Formalisierung im strukturalistischen Ansatz. Für diesen Zweck können wir die Aufsätze von Tuomela, Niiniluoto und Hoering, die auf einer

[6] Vgl. Adams (1959, S. 261f) und Sneed (1971, S. 231f). Bemerkungen allgemeinerer Art macht Stegmüller (1979, S. 42f).

[7] Ein gewisser Minimalschutz gegen allzu beliebiges Herumhantieren an Reduktionsfunktionen wird durch die formalen Bedingungen gewährleistet, daß **F** M^o_{p2} *auf* M_{p1} abbilden soll und daß dieses M_{p1} als die Klasse *aller* Strukturen vom Typ τ_1 definierbar sei.

[8] Vgl. Mayr (1976, S. 286f), Tuomela (1978, S. 220, 226) und insbesondere Hoering (1984, S. 37–39). Hoerings (1984, S. 35f) Besprechung des syntaktischen Ansatzes von Eberle (1971) kann als Warnung dienen, daß der Statement view von ganz ähnlichen Gefahren bedroht ist wie der Non-statement view. Vgl. auch Fußnote 21 von Pearce (1987, S. 112).

allgemeineren Ebene argumentieren, im Hintergrund halten; durchzugehen bleiben dann noch die Beiträge von Mayr, Pearce, Kamlah und Mormann. Diese Arbeiten sind interessant und wichtig, aber größtenteils ziemlich kompliziert, und die Art der Darstellung spiegelt die jeweils persönliche Vorliebe eines Autors für einen bestimmten technischen Aufbau wider. Ich möchte die disparat dargebotenen Ideen als ein zusammengehöriges Ganzes präsentieren. Dabei lasse ich, trotz eines kleinen Verlusts an systematischer Stringenz, die zeitliche Aufeinanderfolge der Kritiken unverändert. Mein Hauptanliegen ist es, die Entwicklung der Diskussion kohärent und in einfacher Sprache darzustellen, so daß sie — ungeachtet beträchtlicher Irrungen und Verwirrungen — leicht zu verfolgen ist. Aus dem Verlauf der Diskussion wird sich ein mehrdeutiges Fazit ziehen lassen: Entweder ist die logische Struktur des Reduktionsbegriffs im Non-statement view bisher nicht ausreichend klar geworden, oder es gibt mehrere, konkurrierende logische Typen strukturalistischer Reduktion. Die wichtigsten Typisierungen wurden von Pearce (zusammen mit Rantala) und von Scheibe vorgeschlagen und kommen am Ende dieses Kapitels zur Sprache. Jedoch — dies sei als Warnung vorausgeschickt — werden auch sie keine endgültige Klarheit über „den" strukturalistischen Reduktionsbegriff bringen können.

1.2 Definierbarkeit und Ableitbarkeit, intendierte Anwendungen und Modelle: Reduktion nach Adams und Sneed

Ernest Adams nennt in seinem wegweisenden Artikel genau die vom Statement view Nagelscher (1949; 1961) und Hempelscher (1965; 1966) Prägung her bekannten Adäquatheitsbedingungen. Danach ist T_1 nur dann auf T_2 reduzierbar, wenn die folgenden Bedingungen erfüllt sind:

(K1) Die grundlegenden Begriffe von T_1 sind durch die grundlegenden Begriffe von T_2 definierbar[9]

(K2) die grundlegenden Gesetze von T_1 sind aus den grundlegenden Gesetzen von T_2, zusammen mit den in (K1) erwähnten Defini-

[9]Hier variieren die Formulierungen: Adams (1959, S. 260) verwendet das Wort „definieren", Sneed (1971, S. 217) spricht von „correspondence" und Stegmüller sagt „überführen" (1985, S. 145) oder „zueinander in Beziehung setzen" (1986, S. 130). — Zur Angemessenheit des Definitionsbegriffs hier siehe auch den Anfang von Kapitel 2.

tionen, ableitbar[10].

Adams betont unter Berufung auf Nagel, daß (K1) den Charakter einer empirischen Hypothese hat, während (K2) eine a priori entscheidbare Behauptung ist. Diese Aufspaltung von Reduktion in einen „angewandten" und einen „formalen" Teil kann in Adams' strukturalistischen „Analoga" wiedererkannt werden, welche unter Zuhilfenahme einer Funktion **F** gemäß (S) sehr leicht ausgedrückt werden können:

(K1^s) $\mathbf{F}[I_2] \supseteq I_1$;
(K2^s) $\mathbf{F}[M_2] \subseteq M_1$.

(K1^s) besagt, daß es für jede intendierte Anwendung der reduzierten Theorie (mindestens) eine korrespondierende intendierte Anwendung der reduzierenden Theorie gibt. Ich kann mich Adams' (1959, S. 260) und Sneeds (1971, S. 217) Meinung, daß (K1^s) ein genaues Gegenstück zu (K1) sei, nicht anschließen. Denn während (K1) die Möglichkeit einer präzisen Anleitung präsupponiert, *wie* man T_2-Anwendungen (oder allgemeiner: potentielle T_2-Modelle) aus vorgegebenen T_1-Anwendungen (bzw. potentiellen T_1-Modellen) „konstruieren" kann, sagt (K1^s) lediglich, *daß* man korrespondierende T_2-Anwendungen finden kann. Weiter gibt es in (K1) keine Beschränkung auf *intendierte* Anwendungen. Deshalb ist (K1^s) schwächer als (K1), und man würde erwarten, daß hier Kritik ansetzt. Dagegen scheint (K2^s), nach dem T_2-Modelle nur auf T_1-Modelle abgebildet werden können, ein sehr getreues Abbild des fundamentalen Kriteriums (K2) zu sein: Die syntaktische Relation der Ableitbarkeit wird widergespiegelt durch die Inklusionsrelation auf der Modellebene, wobei die zwischen T_1 und T_2 vermittelnden Definitionen sozusagen in **F** eingebaut sind und der Definitionsbereich M_{p2}^{o} von **F** den Anwendungsbereich von T_1 in T_2 charakterisiert. Wir werden aber sehen, daß überraschenderweise nicht (K1^s), sondern (K2^s) ins Kreuzfeuer der Kritik geraten ist.[11]

---

[10] Adams (1959, S. 261) sagt „derive", Sneed (1971, S. 220) „deduce" und Stegmüller „abbilden" (1985, S. 145) oder „ableiten" (1986, S. 129).

[11] Es ist das Verdienst von Dieter Mayr, hervorgehoben zu haben, daß Sneed (1971, S. 229) und Stegmüller (1985, S. 151) ursprünglich *nicht* (K2^s), sondern
(K2^{s*}) $\forall \emptyset \neq N_1 \subseteq M_1 \; \exists N_2 \subseteq M_2 \; (\emptyset \neq \mathbf{F}[N_2] \subseteq N_1)$
als Kriterium für die Reduktion von Theorien (oder „Theorienelementen") vorgeschlagen hatten. Die Bedingungen, daß N_1 und $\mathbf{F}[N_2]$ nichtleer seien, sind von Sneed und Stegmüller sicherlich intendiert und wurden von mir ergänzt. Andernfalls wäre ja (K2^{s*}) völlig trivial (man nehme einfach ein $N_2 \subseteq M_2 \setminus M_{p2}^{o}$). Die Mayrschen (1976, S. 284, 286) Theoreme 2.7 und 2.8, nach denen (K2^{s*}) stärker als (K2^s) wäre, sind nicht korrekt; ihre Beweise zeigen jedoch, daß (K2^{s*}) äquivalent zu $\mathbf{F}[M_2] \supseteq M_1$ ist — ein im Lichte der nachfolgenden Diskussion kurioses Ergebnis! In Sneed (1976, S. 137f) und in Stegmüller

Es gibt noch weitere Kriterien, die bei der Motivation des durch (S), ($K1^s$) und ($K2^s$) charakterisierten strukturalistischen Reduktionskonzepts eine Rolle spielten. Adams (1959, S. 261) betrachtet die folgende (notwendige) Bedingung als die wichtigste für eine Reduktion von T_1 auf T_2:

(K3) Wenn T_2 wahr („korrekt") ist, dann auch T_1.

Zur Beurteilung von (K3) müssen wir wissen, was es heißen soll, daß eine Theorie wahr oder korrekt ist. Adams' plausible Antwort lautet so:

1.2.1. Definition Eine Theorie $T=\langle M,I\rangle$ ist genau dann *wahr* (*korrekt*), wenn $I\subseteq M$ (oder äquivalent, wenn $I\backslash M=\emptyset$).

Damit kann (K3) in strukturalistischer Übersetzung wie folgt ausgedrückt werden:

($K3^s$) $I_2\subseteq M_2 \Rightarrow I_1\subseteq M_1$ (oder äquivalent $I_2\backslash M_2=\emptyset \Rightarrow I_1\backslash M_1=\emptyset$)

Es ist klar, daß ($K3^s$) durch ($K1^s$) und ($K2^s$) garantiert ist. Aus $I_2\subseteq M_2$ folgt nämlich $\mathbf{F}[I_2]\subseteq\mathbf{F}[M_2]$, und deshalb gilt mit ($K1^s$) und ($K2^s$) $I_1\subseteq \mathbf{F}[I_2]\subseteq\mathbf{F}[M_2]\subseteq M_1$. Also stützt das Kriterium (K3) den Adamsschen Vorschlag.

Ein weiteres Kriterium kann man bei Sneed (1971, S. 218) und Stegmüller (1985, S. 143) finden:

(K4) Alles, was durch T_1 erklärt („systematisiert") werden kann, kann auch durch T_2 erklärt (systematisiert) werden.

Zur Anwendung dieser Forderung müßte uns ein angemessener Begriff der wissenschaftlichen Erklärung oder, allgemeiner, der Systematisierung zur Verfügung stehen.[12] Sneed und Stegmüller geben uns aber keine Hinweise darauf, wie man einen solchen Begriff passend in den strukturalistischen Ansatz einführen kann. Deshalb müssen sie sich wohl vorhalten lassen, mit (K4) zwar ein beachtenswertes Kriterium vorgebracht zu haben, aber nicht zu wissen, was sie im strukturalistischen Kontext damit anfangen sollen. An dieser Stelle gibt es also kein ($K4^s$). Wir werden erst in den Abschnitten 1.6 und 1.8 auf einen Formalisierungsvorschlag für (K4) zu sprechen kommen.

Schließlich gibt Sneed (1976, S. 139) die sogenannte „Erhaltungseigenschaft" als ein zusätzliches Desideratum an. Danach soll, wenn T_1 auf T_2

(1979, S. 96) kann man dann (kompliziertere Versionen von) ($K2^s$) als Kriterium finden. Es ist erstaunlich, daß sowohl Sneed als auch Stegmüller ihren doch beträchtlichen Sinneswandel (und die Mayrsche Beobachtung) unkommentiert ließen. Vgl. auch die Diskussion des Kriteriums (K5) unten.

[12] Tuomela (1978, S. 220f) meint, daß der Begriff der Reduktion nur über den Begriff der Erklärung zu analysieren ist. Im Gegensatz dazu ist Kamlah (1985, S. 124f) der Ansicht, daß (K4) auf ein alternatives Explikandum hindeutet.

reduzierbar ist, die folgende Bedingung erfüllt sein:

(K5) Für jede Spezialisierung T_1' von T_1 gibt es eine Spezialisierung T_2' von T_2, so daß T_1' auf T_2' reduzierbar ist (und zwar auf eine der Reduktion von T_1 auf T_2 analoge Art und Weise).

Zunächst wollen wir uns in Erinnerung rufen, was eine Spezialisierung ist:

1.2.2. Definition $T'=\langle M',I'\rangle$ ist genau dann eine *Spezialisierung* von $T=\langle M,I\rangle$ (in Zeichen: $T'\ll T$), wenn $M'\subseteq M$ und $I'\subseteq I$.

Diese Definition soll zum Ausdruck bringen, daß eine Spezialisierung mehr oder speziellere Gesetze, dafür aber weniger intendierte Anwendungen hat als die ihr übergeordnete allgemeinere Theorie. Eine Unklarheit in (K5) ist leicht zu beseitigen durch den Hinweis, daß dieses Kriterium eigens auf die strukturalistischen Begrifflichkeit zugeschnitten ist: „auf analoge Art und Weise" soll schlicht heißen, daß dieselbe Reduktionsfunktion (i.a. mit eingeschränktem Definitionsbereich) benutzt werden kann. Wenn **F** nun also die Reduktionsfunktion zwischen T_1 und T_2 ist, kann (K5) folgendermaßen übersetzt werden:

(K5^s) $\forall T_1'\ll T_1\ \exists T_2'\ll T_2\ (\mathbf{F}[I_2']\supseteq I_1' \wedge \mathbf{F}[M_2']\subseteq M_1')$.

Sneed behauptet — und Balzer und Sneed (1978, S. 188–192) beweisen —, daß ihr Reduktionsbegriff, d.h. eine durch die Berücksichtigung von theoretischen Termen und Constraints kompliziertere Version von (S), (K1^s) und (K2^s), das Kriterium (K5^s) erfüllt.[13] Daß dies als Irrtum angesehen werden muß, ist eines der Hauptpunkte von Mayrs Kritik, der wir uns nun zuwenden wollen.

1.3 Spezialisierungen und Anomalien: der Vorschlag von D. Mayr

Dank unserer Vereinfachung des strukturalistischen Theorienkonzepts haben wir es viel leichter als Mayr (1976, S. 285), eine Reduktionsrelation à la Adams und Sneed anzugeben, die ein Gegenbeispiel zu (K5^s) darstellt. Seien $T_1=\langle M_1,I_1\rangle$ und $T_2=\langle M_2,I_2\rangle$ Theorien und **F** eine Reduktionsfunktion derart, daß $\mathbf{F}[M_2]\subseteq M_1$, aber $\mathbf{F}[M_2]\neq M_1$, und $\mathbf{F}[I_2]=I_1$, und sei $T_1'=\langle M_1',I_1'\rangle:=\langle M_1\setminus\mathbf{F}[M_2],I_1\rangle$. Offenbar gilt $T_1'\ll T_1$, aber es gibt kein *nichtleeres* $M_2'\subseteq M_2$ mit $\mathbf{F}[M_2']\subseteq M_1'$, denn $\mathbf{F}[M_2']\subseteq\mathbf{F}[M_2]$ und $\mathbf{F}[M_2]\cap M_1'=\emptyset$. Mithin gibt es hier kein geeignetes T_1 für (K5^s). Tatsächlich ist das M_2' im Beweis von Balzer und Sneed, in unsere Bezeichnun-

[13] Die Behauptung findet sich auch bei Stegmüller (1986, S. 134) in Th4-2

gen übertragen, definiert als $M_2 \cap \mathbf{F}^{-1}[M_1']$, und diese Menge ist im angegebenen Beispiel leer. Die Autoren scheinen nicht bemerkt zu haben, daß Mayr im wesentlichen gezeigt hatte, daß in gewissen Fällen jedes M_2', das als Kandidat für ($K5^s$) dienen könnte, leer sein *muß*. Jedenfalls kann es nicht erwünscht sein, das Kriterium ($K5^s$) für die Adams-Sneedsche Reduktion auf triviale Weise zu retten, indem man inkonsistente „reduzierende" Spezialisierungen T_2' einsetzt. Da nur widerspruchsfreie Theorien wissenschaftlich sinnvoll sind, habe ich gleich zu Beginn inkonsistente Theorien aus unseren Betrachtungen ausgeschlossen und vorausgesetzt, daß die Modellmenge M für jede Theorie T nichtleer sein soll. Aus ähnlichen Gründen sollte außerdem das Kriterium ($K5^s$) durch die zusätzliche Forderung $\mathbf{F}[M_2'] \neq \emptyset$ verbessert werden; im folgenden beziehe sich die Signatur ($K5^s$) auf diese verbesserte Version von (K5). Zusammengefaßt ist festzustellen, daß das Kriterium (K5) keine Stütze der Reduktionsexplikation mittels (S), ($K1^s$) und ($K2^s$), die zunächst so befriedigend erschien, abgibt. Ganz im Gegenteil, (K5) wirft deutliche Zweifel an dieser Explikation auf.

Ein zweites intuitives Argument, welches Mayr vorbringt, ist dieses:[14]

(K6) Wenn T_1 eine Spezialisierung von T_2 oder wenn T_2 eine Spezialisierung von T_1 ist, dann stehen T_1 und T_2 nicht in einer Reduktionsrelation.

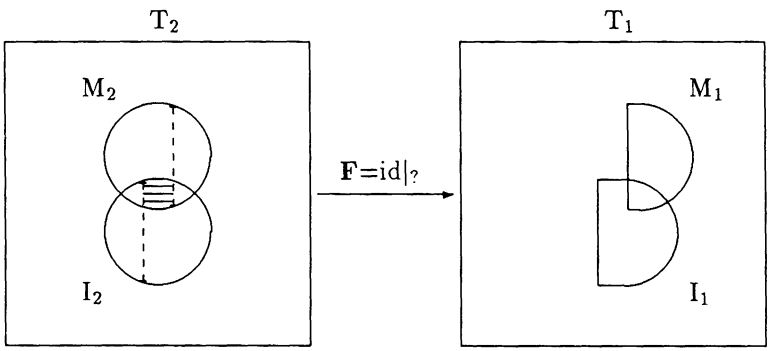
Wir werden diese Form des Kriteriums benutzen, doch wir sollten es der Einheitlichkeit halber auch als notwendige Bedingung für reduzierte und reduzierende Theorien T_1 und T_2 notieren:

($K6^s$) Weder $T_1 \ll T_2$ noch $T_2 \ll T_1$.

Auch (K6) bereitet einige Probleme, aber nicht eigentlich für das durch die Kombination von ($K1^s$) und ($K2^s$) charakterisierte Reduktionskonzept. Trotzdem ist die Diskussion von (K6) lehrreich. Angenommen, T_2 ist eine Spezialisierung von T_1. Wenn wir nun nichts weiter betrachten würden als die durch (S) und ($K2^s$) bestimmte Reduktion der sogenannten Theorienkerne (welche in unserem vereinfachten Theorienkonzept mit den Modellklassen identisch sind), dann wäre T_1 auf T_2 reduzierbar: Man wähle einfach die identische Abbildung auf $M_{p2}^{\circ} = M_{p2}$, d.h. $\mathbf{F} = \mathrm{id}|M_{p2}$. Offenbar gilt dann $\mathbf{F}[M_2] = M_2 \subseteq M_1$. Nehmen wir nun umgekehrt an, daß T_1 eine Spezialisierung von T_2 ist. Dann wäre nach dem *ursprünglichen* Reduktionskonzept von Sneed und Stegmüller (vgl. Fußnote 11) ebenfalls vermittels $\mathbf{F} = \mathrm{id}|M_{p2}$ T_1 auf T_2 reduzierbar. Adams' alte Version des Reduktionsbegriffs bleibt gegen Mayrs Kritik immun. Es ist interessant nachzusehen,

[14] Zur Motivation dieses Kriteriums siehe Mayr (1976, S. 279f, 287).

woran das liegt. Während man mit dem ersten Fall ($T_2 \ll T_1$) mit Hilfe von ($K1^s$) leicht fertig wird, könnte man im zweiten Fall ($T_1 \ll T_2$) vermuten, daß man vermittels einer geschickt gewählten Einschränkung von $id|M_{p2}$ schlußendlich doch eine Reduktion zustande bringen könnte. Aber weder $id|I_1$ noch $id|M_1 \cup \mathcal{C}M_2$ (wobei $\mathcal{C}M_2$ des Komplement von M_2 ist), die bei Optimierung der Chancen für das jeweils andere Kriterium direkt auf ($K1^s$) bzw. ($K2^s$) zugeschnitten sind, können garantieren, daß die beiden Kriterien gleichzeitig erfüllt werden.[15] Eine kleine Skizze zeigt uns, daß dies daran liegt, daß die Klasse $I_1 \setminus (M_1 \cup \mathcal{C}M_2)$ nicht leer sein muß:[16]



SKIZZE 1.1

Es ist ratsam, die Klassen von intendierten Anwendungen und Modellen nicht nur einzeln, sondern auch kombiniert in Betracht zu ziehen. Zu diesem Zweck erweist sich die folgende Aufteilung der intendierten Anwendungen als nützlich:

1.3.1. Definition Sei $T=\langle M,I\rangle$ eine Theorie. Dann heißen die Elemente von $I \cap M$ *erfolgreiche Anwendungen von T* und die Elemente von $I \setminus M$ *Anomalien von T*.

Nun kann die Situation so dargestellt werden: (K6) stellt ein Problem für ($K1^s$) und ($K2^s$) dar, insofern als eine Spezialisierung T_1 von T_2, die

[15] Ganz zu schweigen davon, daß entweder M_{p1} nicht wie gewünscht als die Klasse aller Strukturen vom Typ der T_1-Modelle definiert werden könnte oder daß man die Forderung der Surjektivität für **F** aufgeben müßte. Vgl. Fußnote 7.

[16] Es ist jedoch denkbar, daß die von Stegmüller (1985, S. 224-231) so genannte „Regel der Autodetermination" bei der Konstruktion von Spezialisierungen perfekt wirksam ist, d.h. daß schon $I_1 \setminus M_1$ leer ist. In diesem Fall würde (K6) definitiv gegen das Reduktionskonzept nach ($K1^s$) und ($K2^s$) sprechen.

keine erfolgreiche T_2-Anwendung in eine T_1-Anomalie umwandelt, auf T_2 reduzierbar ist. Da die Existenz solcher Spezialisierungen recht plausibel ist, muß man darauf gefaßt sein, daß (K6) auch dem Adamsschen Vorschlag gefährlich werden kann. Dies scheint Mayr entgangen zu sein (aber vgl. (K7) unten).

Bevor wir uns nun dem dritten und, soweit ich sehen kann, letzten Argument der Mayrschen Kritik zuwenden, seien zwei Bemerkungen angebracht. Erstens zeigt (K6), wie wichtig es ist, sich über die zweischneidige Rolle des Definitionsbereichs M°_{p2} von **F** Klarheit zu verschaffen. Auf der einen Seite übernimmt M°_{p2} die Einschränkung des breiteren Anwendungsbereichs der reduzierenden Theorie T_2, was in etwa den — nicht gesetzesartigen — Anfangsbedingungen für T_1 im Statement view entspricht. Auf der anderen Seite erlaubt es M°_{p2} aber, zusätzliche *Gesetze*, die nicht zu T_2 gehören, in T_1 „hineinzuschmuggeln", so daß die „reduzierte" Theorie viel stärker als die „reduzierende" sein kann — nach meinem Gefühl eine eindeutig kontraintuitive Konsequenz. Man muß also nach Regeln für die Auswahl geeigneter Bereiche M°_{p2} Ausschau halten, und es scheint, daß hierzu der Begriff der Gesetzesartigkeit vonnöten ist. Zweitens führt uns (K6) vor Augen, daß es unerläßlich ist, sich über das Verhältnis von Spezialisierung — die bisweilen als für den Theorienwandel in Perioden der Normalwissenschaft typisch angesehen wird — und Reduktion — die eher als für wissenschaftliche Revolutionen charakteristisch angesehen wird — weitere Gedanken zu machen.

Kommen wir nun zu Mayrs letztem Desiderat für Reduktionen:

(K7) Die reduzierende Theorie kann Anomalien der reduzierten Theorie auflösen („erklären"[17]).

Nach Mayr heißt dies, daß es erfolgreiche Anwendungen von T_2 geben sollte, denen Anomalien von T_1 entsprechen. Unter Verwendung von Definition 1.3.1 können wir (K7) umschreiben in

(K7^s) $\mathbf{F}[I_2 \cap M_2] \cap (I_1 \setminus M_1) \neq \emptyset$.

Die Gültigkeit von (K7^s) wird durch das zur Diskussion stehende Reduktionskonzept aber völlig ausgeschlossen. Denn gemäß (K2^s) gilt $\mathbf{F}[M_2] \cap \mathcal{C}M_1 = \emptyset$, und da $\mathbf{F}[I_2 \cap M_2] \subseteq \mathbf{F}[M_2]$ und $I_1 \setminus M_1 \subseteq \mathcal{C}M_1$, gilt a fortiori $\mathbf{F}[I_2 \cap M_2] \cap I_1 \setminus M_1 = \emptyset$.

Mayr beschränkt sich nicht auf destruktive Kritik, sondern macht auch, vor allem im Hinblick auf (K5) und (K7), einen konstruktiven Gegenvorschlag. Seine Alternative zu (K1^s) und (K2^s) besteht aus den folgenden

[17] Zum Erklärungsbegriff vgl. Abschnitt 1.6.

beiden Forderungen (Mayr 1976, S. 289):

($K1^{s\prime}$) $\mathbf{F}[I_2]=I_1$,

($K2^{s\prime}$) $\mathbf{F}[M_2]\supseteq M_1$.

Ich bin mir nicht sicher, ob Mayr bemerkt hat, daß sein Versuch, die Sneedsche mehrmehrdeutige Relation $\bar{\varrho}$ durch seine mehreindeutige Relation $\bar{\varrho}$ zu ersetzen, die intuitive Verschärfung von ($K1^s$) zu ($K1^{s\prime}$) zur Folge hat. Es sollte darauf hingewiesen werden, daß (nur) diese Abänderung eine Verletzung des Mayrschen Kriteriums (K6) verhindert, weil ($K1^s$) und ($K2^{s\prime}$) es erlauben würden, jede beliebige Spezialisierung T_1 von T_2 auf T_2 vermittels $id|M_{p2}$ zu reduzieren.[18] Weiterhin stellt sich Mayrs ($K2^{s\prime}$) als äquivalent mit der Grundidee ($K2^{s}*$) des ursprünglichen Sneed-Stegmüllerschen Vorschlags heraus (vgl. Fußnote 11), die Mayr selbst getadelt hat. Dann sollte man, wenn man für ($K1^{s\prime}$) und ($K2^{s\prime}$) plädiert, ($K5^{s\prime}$) so modifizieren, daß die Relation zwischen T_2' und T_1' ebenfalls durch ($K1^{s\prime}$) und ($K2^{s\prime}$) bestimmt ist. Glücklicherweise garantieren ($K1^{s\prime}$) und ($K2^{s\prime}$) diese Umformalisierung von (K5) ebenso wie das alte ($K5^s$) (man nehme $M_2':=\mathbf{F}^{-1}[M_1']\cap M_2$ und $I_2':=I_2$). Schließlich ist die zentrale Reduktionsbedingung (K2) durch ($K2^{s\prime}$) nicht länger gewährleistet, sondern im Gegenteil notwendig verletzt, wenn (K7) zum Tragen kommen soll. Mayr hat dies übersehen, denn er behauptet das Gegenteil (Mayr 1976, S. 276, 286f).

1.4 Wahrheit und Übersetzung: der Vorschlag von D. Pearce

An dieser Stelle kann man natürlich noch kein Urteil darüber fällen, welcher Vorschlag denn nun vorzuziehen sei: der von Adams und Sneed oder der von Mayr.[19] Der Beitrag von Pearce (1982) wirft zusätzliches Licht auf die Situation. Wir brauchen hier weder auf seine entschieden prolinguistischen Tendenzen noch auf seinen starken logischen Apparat einzugehen, sondern können ganz auf der Modellebene verbleiben. Pearce (1982, S. 308) wählt

[18] Die Ersetzung von ($K1^s$) durch ($K1^{s\prime}$) ist nicht wirksam, wenn ($K2^s$) die zweite Bedingung ist; ein $\mathbf{F}$, welches ($K1^s$) und ($K2^s$) erfüllt, könnte leicht in eine Reduktionsfunktion umgewandelt werden, die ($K1^{s\prime}$) erfüllt, indem man den Definitionsbereich von $\mathbf{F}$ auf $\mathbf{F}^{-1}[I_1]$ einschränkt. Aber vergleiche Fußnote 7.

[19] Balzer (1982, S. 298) sieht ($K2^s$) und ($K2^{s\prime}$) nicht als konkurrierende Kriterien, sondern benutzt die Kombination beider. Die Identität $\mathbf{F}[M_2]=M_1$ folgt auch schon aus der Art und Weise, wie Pearce und Rantala in ihren Arbeiten die Funktion $\mathbf{F}$ einführen. Sie betrachten heute diesen Fall allerdings als für nur einen Typ von Reduktion charakteristisch, den sie „Einbettung" nennen. Vgl. Abschnitt 1.8.

eine Formulierung von Stegmüller (1985, S. 146) als seinen Ausgangspunkt. Wenn T_1 auf T_2 reduzierbar ist, dann soll gelten:

(K8) Ist A ein Satz von T_2 und B der entsprechende Satz von T_1, so ist A nur dann wahr, wenn B wahr ist.

Stegmüller hält (K8) für eine anwendungsorientierte Paraphrase von (K2). Den erkennbar dem Statement view entlehnten Begriff „Satz von T" brauchen wir hier nicht problematisieren. Alles, was für das Folgende benötigt wird, ist, daß man jedem Satz A einer Theorie T als Extension eine Teilmenge $||A||$ von M_p zuordnen kann, nämlich die Klasse derjenigen potentiellen Modelle von T, in denen A gilt. Um (K8) aber zur Gänze beurteilen zu können, fehlen uns noch ein paar präzise strukturalistische Begriffe. Die Wahrheit eines Satzes expliziert Pearce (1982, S. 323) als theorienrelativen Begriff, und zwar auf zweifache Weise:

1.4.1. Definition Ein Satz A einer Theorie $T=\langle M,I\rangle$ heißt

(a) *wahr in M* genau dann, wenn $M \subseteq ||A||$;

(b) *wahr in T* genau dann, wenn $M \cap I \subseteq ||A||$.

Terminologisch erschiene es mir genauer zu sagen, daß Klausel (a) definiert, daß A eine *Folgerung aus T* ist. Wenn A aus T folgt und T wahr ist, dann ist A wahr. Legt man Definition 1.2.1 zugrunde, so könnte man also sagen, die (immer noch theorienrelative) Wahrheit eines in der Sprache von T formulierten Satzes A bestehe darin, daß alle intendierten Anwendungen von T den Satz A erfüllen (wobei zu berücksichtigen ist, daß intendierte Anwendungen vom selben Typ wie T-begrifflich gefaßte, real existierende Systeme sind). Der intuitive Gehalt des Definiens von Klausel (b) der Definition 1.4.1, wonach alle erfolgreichen Anwendungen A erfüllen sollen, ist mir nicht ganz klar. Dennoch will ich die Diskussion hier mit den beiden Pearceschen Wahrheitsbegriffen fortsetzen und erst in Abschnitt 1.8 das Thema erneut aufgreifen.

Die Interpretation „des entsprechenden Satzes" in (K8) muß offenbar eine Art Übersetzung (und deshalb eine Art Kommensurabilität) zwischen T_1 und T_2 voraussetzen. Wenn wir uns nichtsdestotrotz weiterhin auf Betrachtungen in der Modellebene beschränken, dann scheint (K8) sehr gut in unser vorliegendes Rahmenwerk hineinzupassen:

(K8^s) Für alle Sätze A von T_2 gilt

(a) $M_2 \subseteq ||A|| \Rightarrow M_1 \subseteq \mathbf{F}[||A||]$,

(b) $M_2 \cap I_2 \subseteq ||A|| \Rightarrow M_1 \cap I_1 \subseteq \mathbf{F}[||A||]$.

Hierbei verlangen wir von „dem entsprechenden Satz" B, der in (K8) erwähnt wird, lediglich, daß $||B||=\mathbf{F}[||A||]$. Versuchsweise könnten wir eine

„Übersetzung von T_2 in T_1 (relativ zu **F**)" abstrakt als eine Funktion von T_2-Sätzen in T_1-Sätze definieren, die jedem A ein B mit der Eigenschaft $||B|| = \mathbf{F}[||A||]$ zuordnet. Warum wir das nicht wirklich tun, wird unten gleich klar werden; für einen Moment verweilen wir aber noch bei dieser Konstruktion.

Ein konsequenter Vertreter des Strukturalismus wird sich von allen sprachlichen Überresten in ($K8^s$) befreien wollen. Dazu muß er nicht einmal annehmen, daß jede „Proposition" einer Theorie, d.h. jede Teilklasse ihrer potentiellen Modelle, sprachlich ausdrückbar ist. Es genügt die Voraussetzung, daß die Klasse M_2 aller Modelle von T_2 bzw. die Klasse $M_2 \cap I_2$ aller erfolgreichen Anwendungen von T_2 in T_2 definierbar ist, um zu zeigen, daß (K8) ohne jeden Rückgriff auf sprachliche Entitäten reformuliert werden kann. Denn dann ist ($K8^s$) äquivalent zu

($K8^{s\prime}$) (a) $\mathbf{F}[M_2] \supseteq M_1$,
(b) $\mathbf{F}[M_2 \cap I_2] \supseteq M_1 \cap I_1$.

Um ($K8^{s\prime}$)(a) aus ($K8^s$)(a) zu erhalten, setze man für A in ($K8^s$)(a) einfach einen die Klasse der T_2-Modelle definierenden Satz mit $||A||=M_2$ ein. Für die andere Richtung sei A derart, daß $M_2 \subseteq ||A||$; dann ist aber $\mathbf{F}[M_2] \subseteq \mathbf{F}[||A||]$ und wegen ($K8^{s\prime}$)(a) auch $M_1 \subseteq \mathbf{F}[||A||]$. Fall (b) ist natürlich analog zu Fall (a).

Auf diese Weise würde (K8) also, nach den beiden Pearceschen Interpretationen der Wahrheit eines Satzes, zwei Bedingungen liefern: Das Adams-Sneedsche Kriterium ($K2^s$) wird umgekehrt in ($K2^{s\prime}$) — ein Kriterium, das ja schon von Mayr den Vorzug erhalten hatte. Zusätzlich zu diesem Kriterium für Modelle ergibt sich eine entsprechende Bedingung für die erfolgreichen Anwendungen: Jeder erfolgreichen Anwendung der reduzierten Theorie soll — vermittelt durch die Reduktionsfunktion **F** — zumindest eine erfolgreiche Anwendung der reduzierenden Theorie entsprechen.

Leider kann dieses Ergebnis nicht in seiner ganzen Einfachheit aufrechterhalten werden. Da die reduzierende Theorie immer die ausdrucksreichere sein und sprachliche Differenzierungen erlauben soll, die in der reduzierten Theorie nicht nachvollziehbar sind, kann die obige Idee keine intuitiv adäquate Explikation einer Übersetzung sein. „Der" in (K8) präsupponierte „entsprechende Satz" existiert im allgemeinen überhaupt nicht. Folgerichtig geht bei Pearce (1982, S. 314) die Übersetzung gerade in die andere Richtung. Wir wollen hier nur eine notwendige Bedingung für Übersetzungen angeben:

1.4.2. Quasidefinition Sei die Theorie $T_1=\langle M_1,I_1\rangle$ durch eine geeignete Funktion **F** auf die Theorie $T_2=\langle M_2,I_2\rangle$ reduziert.

(a)Ein T_2-Satz A heißt nur dann eine *Übersetzung des T_1-Satzes B relativ zu* **F**, wenn $||A|| \cap M^{\circ}_{p2} = \mathbf{F}^{-1}[||B||]$.
(b)Eine Funktion $\ddot{\mathbf{U}}$ von T_1-Sätzen in T_2-Sätze heißt nur dann eine *Übersetzung von T_1 in T_2 relativ zu* **F**, wenn für jeden T_1-Satz B $\ddot{\mathbf{U}}(B)$ eine Übersetzung von B relativ zu **F** ist.[20]

Folgendes ist eine plausible Konsequenz von 1.4.2: Wenn eine Übersetzung $\ddot{\mathbf{U}}$ von T_1 in T_2 existiert, dann können potentielle Modelle x_2 und y_2 von T_2, die auf dasselbe potentielle Modell von T_1 abgebildet werden, nicht durch Übersetzungen von T_1-Sätzen voneinander unterschieden werden, d.h. $\mathbf{F}(x_2)=\mathbf{F}(y_2)$ impliziert $x_2 \in ||\ddot{\mathbf{U}}(B)|| \Leftrightarrow y_2 \in ||\ddot{\mathbf{U}}(B)||$ für jeden in der Sprache von T_1 formulierten Satz B.

Diese Änderung bringt es mit sich, daß ($K8^s$) durch

($K8^{s\prime\prime}$) Für alle Sätze B von T_1 gilt
(a) $M_2 \subseteq ||\ddot{\mathbf{U}}(B)|| \Rightarrow M_1 \subseteq ||B||$,
(b) $M_2 \cap I_2 \subseteq ||\ddot{\mathbf{U}}(B)|| \Rightarrow M_1 \cap I_1 \subseteq ||B||$.

ersetzt werden muß. Hier kommt die Präsupposition, daß eine Übersetzung von T_1 nach T_2 existiert, explizit zum Ausdruck. Die linguistisch inspirierte Bedingung ($K8^{s\prime\prime}$) ist nicht mehr gleichwertig mit den einfachen Inklusionen von ($K8^{s\prime}$), welches nämlich nur mehr hinreichende, aber keine notwendigen Bedingungen für ($K8^{s\prime\prime}$) gibt. Betrachen wir wieder den Fall (a). Einerseits ist ($K8^{s\prime}$) hinreichend: Aus $M_2 \subseteq ||\ddot{\mathbf{U}}(B)||$ folgt nach der Quasidefinition 1.4.2 $M_2 \cap M^{\circ}_{p2} \subseteq \mathbf{F}^{-1}[||B||]$, also $\mathbf{F}[M_2] = \mathbf{F}[M_2 \cap M^{\circ}_{p2}] \subseteq \mathbf{F}[\mathbf{F}^{-1}[||B||]] \subseteq ||B||$, und mit ($K8^{s\prime}$) folgt $M_1 \subseteq ||B||$. Andererseits wäre ($K8^{s\prime}$) nur unter der Vorausetzung notwendig, daß $\mathbf{F}[M_2]$ in T_1 definierbar ist, d.h., daß es einen T_1-Satz B mit $||B|| = \mathbf{F}[M_2]$ gibt; dann ist das Konsequens von ($K8^{s\prime\prime}$) identisch mit ($K8^{s\prime}$). Aber das Antezedens von ($K8^{s\prime\prime}$), nämlich $M_2 \subseteq ||\ddot{\mathbf{U}}(B)||$, können wir höchstens dann verifizieren, wenn $M_2 \subseteq M^{\circ}_{p2}$, da die Übersetzung eines T_1-Satzes nur innerhalb von M°_{p2} fixiert ist. Wenn $M_2 \subseteq M^{\circ}_{p2}$, dann genügt es, $M_2 \subseteq \mathbf{F}^{-1}[\mathbf{F}[M_2]]$ zu haben (denn $||\ddot{\mathbf{U}}(B)|| \cap M^{\circ}_{p2} = \mathbf{F}^{-1}[||B||]$), was trivial erfüllt ist. Aber weil die Annahme der T_1-Definierbarkeit von $\mathbf{F}[M_2]$ sehr stark und die Annahme $M_2 \subseteq M^{\circ}_{p2}$ direkt unerwünscht ist, darf man ($K8^{s\prime}$) und ($K8^{s\prime\prime}$) nicht als äquivalent betrachten.

[20] Ich habe hier gegenüber Pearces Formulierung eine zweifache Abschwächung vorgenommen, indem ich erstens nur eine notwendige Bedingung angebe und zweitens die Extension von A bzw. $\ddot{\mathbf{U}}(B)$ nur innerhalb von M°_{p2} als festgelegt betrachte. — Pearce (1981, S. 25; 1982, S. 314) verfährt übrigens, wie auch Balzer (1982, S. 220), abstrakter und nennt **F** selber eine Übersetzung von T_1 in T_2 (und zwar genau dann, wenn es für jeden T_1-Satz B ein A mit $||A||=\mathbf{F}^{-1}[||B||]$ gibt).

Dennoch wird Mayrs Bedingung ($K2^{s\prime}$) von Pearce auf einem ganz und gar unabhängigen Begründungsweg gestützt. Welches der Kriterien — ($K2^{s}$) oder ($K2^{s\prime}$) — man auch immer favorisieren will, es scheint, daß Stegmüllers Intuitionen hinsichtlich der Adäquatheitsbedingung (K8) für Reduktionen in zweierlei Hinsicht korrigiert werden müssen. Zunächst ist (K8) nicht genau genug formuliert, weil die Übersetzung in die andere Richtung laufen muß; dann ist, wie bereits Pearce (1982, S. 328) betont hat, (K8) nicht einmal annähernd gleichbedeutend mit (K2).

Schließlich sei festgehalten, daß Pearce (1982, S. 236) Mayrs ($K1^{s\prime}$) für zu stark erachtet und wieder ($K1^{s}$) an dessen Stelle setzt. Er fügt ($K8^{s\prime}$)(b) als eine dritte Bedingung hinzu, welche sich ebenso wie bei ihm ($K2^{s\prime}$) von (K8) herleitet. ($K8^{s\prime}$)(b) folgt nicht bereits aus den anderen beiden Pearceschen Bedingungen, welche gewährleisten, daß eine erfolgreiche Anwendung x_1 von T_1 eine **F**-korrespondierende intendierte Anwendung $x_2 \in I_2$ und ein **F**-korrespondierendes Modell $y_2 \in M_2$ hat, wobei aber $x_2 = y_2$ nicht gesichert ist. Man sollte weiter beachten, daß durch ($K1^{s}$), ($K2^{s\prime}$) und ($K8^{s\prime}$)(b) nicht verhindert wird, daß jede Spezialisierung T_1 von T_2 vermittels $\mathrm{id}|M_{p2}$ auf ihre allgemeinere „Muttertheorie" T_2 reduziert werden kann. Da in diesem Fall die informativere, also bessere Theorie auf die schwächere Theorie reduzierbar wäre, ist (K6) in seiner kritischen Richtung verletzt.[21]

1.5 Empirische Behauptungen und erlaubte Modelle: der Vorschlag von A. Kamlah

Auch der Aufsatz von Kamlah kann in den vorliegenden Rahmen eingepaßt werden, wenn man seine individuellen Eigenheiten und seine Behandlung von Approximationen außen vor läßt. Ausgehend von einer vielzitierten Stelle bei Hempel (1965, S. 344), wonach die reduzierende Theorie die Gesetze der reduzierten Theorie nur innerhalb eines beschränkten Bereichs impliziert,[22] läuft Kamlahs Argumentation darauf hinaus, daß es nicht wie in (K2) die Gesetze, sondern die sogenannten empirischen Behauptungen der Theorien T_1 und T_2 sind, die in einer Konsequenzrelation stehen sollen. Das richtige Kriterium für Reduktionen muß nach Kamlah also so lauten:

[21] Vgl. aber Abschnitt 1.8.

[22] Dieses Zitat kann man fast wörtlich auch schon bei Kemeny und Oppenheim (1956, S. 13) finden, wobei die Autoren in einer Fußnote zum entsprechenden Absatz Hempel für seine klärenden Hinweise danken. S. Kapitel 2, Fußnote 34.

(K9) Die empirische Behauptung von T_1 folgt aus der empirischen Behauptung von T_2.

In unserem vereinfachten Theorienmodell ist die empirische Behauptung einer Theorie $T=\langle M,I\rangle$ einfach mit der in Definition 1.2.1 angegebenen Wahrheitsbedingung $I\subseteq M$ zu identifizieren. Kamlahs Idee ist es nun, diese metatheoretische Forderung auf die T-theoretische Ebene „herunterzuziehen", indem er die Menge $\mathcal{C}I\cup M$ der *von T erlaubten potentiellen Modelle* betrachtet. Dies ist die Menge derjenigen $x\in M_p$, die den Satz „Alle I's sind M's", formalisiert durch $\forall x(Ix\rightarrow Mx)$, erfüllen. Auf der Grundlage dieser Idee kann man (K9) wie folgt strukturalistisch präzisieren:

($K9^s$) $\mathbf{F}[\mathcal{C}I_2\cup M_2]\subseteq \mathcal{C}I_1\cup M_1$.[23]

Zunächst einmal ist natürlich der Zusammenhang mit dem verwandten Kriterium (K3) interessant. Wenn wir, wie es üblich ist, voraussetzen, daß die Reduktionsfunktion $\mathbf{F}$ surjektiv ist, dann impliziert ($K9^s$) ($K3^s$): Aus $I_2\backslash M_2=\emptyset$ erhalten wir mit ($K9^s$) $\mathbf{F}[M_{p2}]\subseteq \mathcal{C}I_1\cup M_1$, und da $\mathbf{F}$ surjektiv ist, folgt $I_1\backslash M_1=\emptyset$.

Wir können ($K9^s$) auch als ein Kriterium für Anomalien formulieren. Tatsächlich ist es nämlich äquivalent zu der Bedingung

($K9^{s\prime}$) $I_2\backslash M_2 \supseteq \mathbf{F}^{-1}[I_1\backslash M_1]$.

Der Beweis ist trivial: Für die Richtung „($K9^s$)$\Rightarrow$($K9^{s\prime}$)" sei x_2 in $\mathbf{F}^{-1}[I_1\backslash M_1]$; das heißt $\mathbf{F}(x_2)$ ist in $I_1\backslash M_1=\mathcal{C}(\mathcal{C}I_1\cup M_1)$, also folgt mit ($K9^s$), daß x_2 nicht in $\mathcal{C}I_2\cup M_2$ ist, das heißt x_2 ist in $I_2\backslash M_2$. Für die Richtung „($K9^{s\prime}$)$\Rightarrow$($K9^s$)" sei x_1 in $\mathbf{F}[\mathcal{C}I_2\cup M_2]$; das heißt, es gibt ein x_2 in $\mathcal{C}I_2\cup M_2$ mit $\mathbf{F}(x_2)=x_1$, also ist $\mathbf{F}^{-1}(x_1)$ keine Teilmenge von $I_2\backslash M_2$, also folgt mit ($K9^{s\prime}$), daß x_1 nicht in $I_1\backslash M_1$ enthalten ist, das heißt, x_1 ist in $\mathcal{C}I_1\cup M_1$ enthalten. □

($K9^{s\prime}$) besagt, daß jede Anomalie der reduzierten Theorie nur Anomalien der reduzierenden Theorie als $\mathbf{F}$-Korrelate hat. Damit ist sofort klar, daß diese Interpretation von (K9) im Widerspruch zu (K7) steht. Aus ($K9^s$) folgt wegen $I_2\cap M_2\subseteq M_2\subseteq \mathcal{C}I_2\cup M_2$ sofort $\mathbf{F}[I_2\cap M_2]\subseteq \mathcal{C}I_1\cup M_1$, und dies heißt nichts anderes als $\mathbf{F}[I_2\cap M_2]\cap (I_1\backslash M_1)=\emptyset$.

Weiter hat Kamlah (1985, S. 138) ein Gegenstück zu (K1) in seiner endgültigen Definition des Reduktionsbegriffs, und zwar

[23] Das Kriterium (K9) spielt jetzt auch bei Balzer, Moulines und Sneed (1987, z.B. S. 264, 274, 279) eine große Rolle. Man vergleiche die Formalisierung von (K9) dort (S. 279) im vollentwickelten strukturalistischen Theorienmodell, die ($K3^s$) entspricht.

$(K1^{s\prime\prime})$ $I_2 \supseteq \mathbf{F}^{-1}[I_1]$. [24]

Wenn wir wieder die Surjektivität von **F** voraussetzen, dann impliziert $(K1^{s\prime\prime})$ $\mathbf{F}[I_2] \supseteq \mathbf{F}[\mathbf{F}^{-1}[I_1]] = I_1$, und das ist $(K1^s)$. Umgekehrt jedoch verbürgt $(K1^s)$ keineswegs $(K1^{s\prime\prime})$. Dies wäre nur dann der Fall, wenn $I_2 \supseteq \mathbf{F}^{-1}[\mathbf{F}[I_2]]$ gelten würde.[25]

Deshalb ist $(K1^{s\prime\prime})$ stärker als $(K1^s)$ und entsprechend schwieriger zu rechtfertigen. Kamlahs eigene Anmerkungen zu $(K1^{s\prime\prime})$ sind in dieser Hinsicht nicht zufriedenstellend.

Kamlahs Kriterien $(K1^{s\prime\prime})$ und $(K9^s)$ lassen sich über die Adamsschen Kriterien $(K1^s)$ und $(K2^s)$ begründen, wenn man eine Zusatzbedingung akzeptiert, die schwächer ist als $(K1^{s\prime\prime})$ und wesentlich weniger problematisch aussieht. Nach meiner Ansicht ist es intuitiv durchaus plausibel, daß Reduktionsfunktionen **F** im allgemeinen so bestimmt sind, daß die Menge der **F**-Urbilder jeder intendierten Anwendung der reduzierten Theorie entweder ganz in der Menge der intendierten Anwendungen der reduzierenden Theorie oder ganz außerhalb davon liegt.[26] Die Zusatzannahme lautet

(Z) $\forall x_1 \in I_1 \; (\mathbf{F}^{-1}(x_1) \subseteq I_2 \vee \mathbf{F}^{-1}(x_1) \subseteq \mathcal{C}I_2)$.

Unter der Voraussetzung von (Z) impliziert $(K1^s)$ die stärkere Bedingung $(K1^{s\prime\prime})$: Wennn es für jedes $x_1 \in I_1$ zumindest ein $x_2 \in I_2$ mit $\mathbf{F}(x_2) = x_1$ gibt, dann muß wegen (Z) auch schon das ganze $\mathbf{F}^{-1}(x_1)$ in I_2 liegen, und wir haben $(K1^{s\prime\prime})$. Wichtiger noch ist, daß sich nun auch Kamlahs $(K9^s)$ aus Adams' Kriterien ableiten läßt, wenn wir die Gültigkeit von (Z) unterstellen. Wie Kamlah (1985, S. 136–138) ausführt, kann man mit $(K1^{s\prime\prime})$ als Prämisse die Äquivalenz von $(K9^s)$ mit der Bedingung

$(K9^{s\prime\prime})$ $I_1 \cap \mathbf{F}[M_2] \subseteq M_1$

zeigen, die offensichtlich eine Abschwächung von $(K2^s)$ darstellt. Um zu illustrieren, wie furchterregend aussehende Theoreme und Herleitungen in unserem vereinfachten Theorienmodell auf ein übersichtliches Maß schrumpfen, sei hier ein eigener kleiner Beweis dieser Äquivalenz angegeben. Für die Richtung „$(K9^s) \Rightarrow (K9^{s\prime\prime})$„ sei x_1 in $I_1 \cap \mathbf{F}[M_2]$; also gibt es

[24] Als intuitive Forderung nennen $(K1^{s\prime\prime})$ auch Balzer und Sneed (1977, S. 202f), die in ihrer darauf folgenden Definition dann aber doch $(K1^s)$ den Vorzug geben.

[25] Falls $(K2^s)$ oder $(K9^s)$ — aber nicht $(K2^{s\prime})$ — die zweite Bedingung sein soll, kann man $(K1^{s\prime\prime})$ aus $(K1^s)$ dadurch erhalten, daß man $\mathbf{F}|I_2$ anstelle von **F** nimmt. — Aber vergleiche Fußnote 7.

[26] Diese Annahme ist zum Beispiel schon dann als begründet erwiesen, wenn die Klasse der intendierten Anwendungen von T_2 (innerhalb von M^o_{p2}) T_1-definierbar ist, d.h. genauer, wenn relativ zu **F** eine Übersetzung $\ddot{\mathbf{U}}$ von T_1 in T_2 und ein T_1-Satz B mit $||\ddot{\mathbf{U}}(B)|| \cap M^o_{p2} = I_2 \cap M^o_{p2}$ existiert. Vgl. die Bemerkung nach Definition 1.4.2.

ein x_2 in M_2 mit $\mathbf{F}(x_2)=x_1$, mit ($K9^s$) ist deshalb x_1 in $\mathcal{C}I_1 \cup M_1$ und — weil x_1 auch in I_1 liegt — genauer in M_1. Für die Richtung "($K9^{s\prime\prime}$)⇒($K9^s$)" sei x_1 in $\mathbf{F}[\mathcal{C}I_2 \cup M_2]$; also gibt es ein x_2 in $\mathcal{C}I_2 \cup M_2$ mit $\mathbf{F}(x_2)=x_1$; falls x_2 in $\mathcal{C}I_2$ liegt, dann ist wegen ($K1^{s\prime\prime}$) $x_1=\mathbf{F}(x_2)$ in $\mathcal{C}I_1$; falls andererseits x_2 in M_2 liegt, dann ist wegen ($K9^{s\prime\prime}$) $x_1=\mathbf{F}(x_2)$ in $\mathcal{C}I_1 \cup M_1$. □

Für den, der (Z) zu akzeptieren bereit ist, werden also mit ($K9^s$) keine über die Adamsschen hinausgehende Eigenschaften von Reduktionen erfaßt. Da auf der anderen Seite ($K2^s$) aber offensichtlich nicht aus ($K1^{s\prime\prime}$) und ($K9^s$) ableitbar ist, führt Kamlahs Idee zu einer echten Liberalisierung des ursprünglichen strukturalistischen Reduktionskonzepts — eine Liberalisierung, die freilich stark genug ist, um mit dem Mayrschen Anomalienkriterium (K7) im Widerspruch zu stehen.

Kommen wir noch einmal auf Kamlahs Neuerung (K9) zurück. Die grundlegende Idee wird, zumindest teilweise, sehr genau vorweggenommen durch Niiniluotos Wiedergabe der Sneed-Stegmüllerschen Reduktionsdefinition. Er formuliert als Definiens:

> there is a many-one relation $\mathbf{F}^{-1}$ from M_{pp1} to M_{pp2} such that the intended applications of T_1 are correlated with intended applications of T_2 and what T_1 says *about these applications* is entailed by what T_2 says *about the corresponding applications*. (Niiniluoto 1980, S. 36, Bezeichnungen abgeändert, Hervorhebungen von mir.)

Niiniluoto macht keinen Unterschied zwischen ($K2^s$) und ($K9^s$). Wenn man nach der Quelle für diese Ungenauigkeit forscht, so stößt man bei Sneed (1971, S. 218, Zeile 6–3 von unten; 1976, S. 136, Zeile 18–22), bei Balzer und Sneed (1977, S. 202, Zeile 8–5 von unten) und bei Stegmüller (1979, S. 36, Zeile 23–25; 1986, S. 131, Zeile 9-7 von unten) auf ähnliche Sätze, die noch viel verwirrender sind als der Niiniluotos. Stellvertretend zitiere ich Balzer und Sneed (nach Balzer und Heidelberger 1983, S. 127):

(K10) Alles, was die reduzierte Theorie [T_1] über eine gegebene Anwendung sagt, ist enthalten in („entailed by") dem, was die reduzierende Theorie [T_2] über jede entsprechende Anwendung sagt.

Wir können hier davon absehen, daß im *vollen* strukturalistischen Modell eine Theorie wegen der Constraints nicht über *einzelne* Anwendungen spricht. Weil nach (K10) offenbar nicht passieren darf, daß die reduzierende Theorie über verschiedene „entsprechende" Anwendungen etwas Verschiedenes sagt, präsupponiert schon die Formulierung dieses Kriteri-

ums zumindest[27] das Folgende:

(Z′) $\forall x_1 \in I_1\ (\mathbf{F}^{-1}(x_1) \subseteq M_2 \vee \mathbf{F}^{-1}(x_1) \subseteq \mathcal{C}M_2)$.

Ich bezweifle, daß dies eine erwünschte Zusatzbedingung sein kann.[28] Gehen wir dennoch davon aus, daß (Z′) erfüllt ist, so kann man (K10) wohl auf zumindest dreierlei Art präzisieren:

(K10^s) $\forall x_1 \in I_1\ ((\mathbf{F}^{-1}(x_1) \subseteq M_2 \Rightarrow x_1 \in M_1) \wedge (\mathbf{F}^{-1}(x_1) \subseteq \mathcal{C}M_2 \Rightarrow x_1 \in M_1))$.

(K10$^{s\prime}$) $\forall x_1 \in I_1\ (\mathbf{F}^{-1}(x_1) \subseteq M_2 \Rightarrow x_1 \in M_1)$.

(K10$^{s\prime\prime}$) $\forall x_1 \in I_1\ ((\mathbf{F}^{-1}(x_1) \subseteq M_2 \Rightarrow x_1 \in M_1) \wedge (\mathbf{F}^{-1}(x_1) \subseteq \mathcal{C}M_2 \Rightarrow x_1 \in \mathcal{C}M_1))$.

(K10^s) stimmt am besten mit der strukturalistischen Standardauffassung überein, daß die empirische Behauptung von T_1 $I_1 \subseteq M_1$ lautet, d.h. daß T_1 für jedes $x_1 \in I_1$ lediglich „sagen" kann, x_1 sei ein Element von M_1. Die Lesart (K10$^{s\prime}$) von (K10) verschärft diese Idee durch die Forderung, daß jede intendierte Anwendung von T_1 auch „im Lichte der reduzierenden Theorie T_2" ein Modell sein muß, d.h. daß für jedes $x_1 \in T_1$ auch T_2 nur „sagen" kann, daß $\mathbf{F}^{-1}(x_1) \subseteq M_2$. (K10$^{s\prime\prime}$) erlaubt hingegen wieder, daß einige $x_1 \in I_1$ „im Lichte von T_2" keine Modelle sind, und darüber hinausgehend interpretiert es (K10) so, daß T_1 „dasselbe" über diese x_1 „sagt", nämlich $x_1 \notin M_1$. Somit geht (K10$^{s\prime\prime}$) von einer Verletzung der Vorschrift aus, daß $I_1 \subseteq M_1$ die empirische Behauptung von T_1 sein muß. Vielleicht kann man den Verstoß durch den Verweis darauf rechtfertigen, daß aufgeklärte T_1-Theoretiker sehr wohl anzuerkennen bereit sind, daß ein gewisser Anteil der intendierten Anwendungen von T_1 aus Anomalien besteht, und daß diese Theoretiker also, obgleich sie nach wie vor an ihrer Theorie T_1 festhalten wollen, durchaus nicht unbedingt „$I_1 \subseteq M_1$" behaupten.

Wie dem auch sei, keine der oben aufgelisteten Bedingungen läßt uns auf den ersten Blick erkennen, daß die Protagonisten des Strukturalismus fast immer (beachte aber die Fußnoten 11 und 19) die zentrale Adamssche Bedingung (K2^s) im Auge haben. Auf den zweiten Blick, den wir nun auf (K10^s) bis (K10$^{s\prime\prime}$) werfen wollen, wird sich tatsächlich herausstellen, daß (K10) nicht geeignet ist, (K2^s) zu motivieren.

---

[27] Sneeds Formulierung von 1971 präsupponiert darüber hinaus, daß $\mathbf{F}^{-1}$ eine Funktion ist.

[28] (Z′) kann wohl nicht durch das in Fußnote 26 erwähnte Argument gerechtfertigt werden, denn die Klasse der T_2-Modelle (innerhalb von M_{p2}°) wird im allgemeinen vermutlich nicht T_1-definierbar sein. Aber (Z′) könnte durch den formalen Trick garantiert werden, (mit Hilfe des Auswahlaxioms) für jedes $x_1 \in I_1$ *ein* $x_2(x_1)$ aus I_2 mit $\mathbf{F}(x_2) = x_1$ auszuwählen (ein solches $x_2(x_1)$ existiert nach (K1^s)) und dann anstelle von $\mathbf{F}$ die Funktion $\mathbf{F}|\{x_2(x_1) : x_1 \in I_1\}$ zu verwenden. Aber vgl. Fußnote 7.

Unsere der strukturalistischen Doktrin von der empirischen Behauptung am meisten verpflichtete Formalisierung ($K10^s$) kann offenbar kein Kriterium für eine nichttriviale *inter*theoretische Relation sein. Wegen (Z') ist nämlich eine der beiden Antezedenzien immer erfüllt, weshalb ($K10^s$) einfach äquivalent zur sogenannten empirischen Behauptung „$I_1 \subseteq M_1$" von T_1 ist.

Die zweite Version von (K10) ist schwächer und führt ebenfalls nicht zu ($K2^s$), sondern überraschenderweise zu ($K9^s$). Denn es gilt:

$$
\begin{aligned}
(K10^{s\prime}) \quad &\Leftrightarrow \forall x_1 \in I_1 \; (\mathbf{F}^{-1}(x_1) \not\subseteq M_2 \vee x_1 \in M_1) \\
&\Leftrightarrow \forall x_1 \in I_1 \; (\mathbf{F}^{-1}(x_1) \subseteq \mathcal{C}M_2 \vee x_1 \in M_1) \quad \text{(wegen (Z'))} \\
&\Leftrightarrow \forall x_1 \in I_1 \; (x_1 \notin \mathbf{F}[M_2] \vee x_1 \in M_1) \\
&\Leftrightarrow I_1 \subseteq M_1 \cup \mathbf{C}\mathbf{F}[M_2] \\
&\Leftrightarrow I_1 \cap \mathbf{F}[M_2] \subseteq M_1 \quad (\equiv (K9^{s\prime\prime})) \\
&\Leftrightarrow (K9^s) \; .
\end{aligned}
$$

Die dritte Lesart, wiewohl stärker als ($K9^s$), ist genausowenig hinreichend für ($K2^s$). Für das folgende Argument setzen wir wieder voraus, daß $\mathbf{F}$ surjektiv ist.

$$
\begin{aligned}
(K10^{s\prime\prime}) \quad &\Leftrightarrow I_1 \cap \mathbf{F}[M_2] \subseteq M_1 \wedge I_1 \cap \mathbf{F}[\mathcal{C}M_2] \subseteq \mathcal{C}M_1 \quad \text{(wie eben)} \\
&\Leftrightarrow I_1 \cap \mathbf{F}[M_2] \subseteq M_1 \wedge I_1 \cap \mathbf{C}\mathbf{F}[M_2] \subseteq \mathcal{C}M_1 \quad \text{(wegen (Z') und der Surjektivität von } \mathbf{F}) \\
&\Leftrightarrow I_1 \cap \mathbf{F}[M_2] \subseteq M_1 \wedge I_1 \cap M_1 \subseteq \mathbf{F}[M_2] \\
&\Leftrightarrow I_1 \cap \mathbf{F}[M_2] = I_1 \cap M_1 \; .
\end{aligned}
$$

Diese Bedingung bestimmt $I_1 \cap \mathbf{F}[M_2]$ genauer als ($K10^{s\prime}$). Grob gesprochen, sagt ($K10^{s\prime\prime}$), daß T_1 und T_2 relativ zum intendierten Anwendungsbereich I_1 von T_1 „dieselben" Modelle auszeichnen. Außerhalb von I_1 jedoch gibt es keine Forderung für das Verhältnis von $\mathbf{F}[M_2]$ und M_1. ($K10^{s\prime\prime}$) ist keine notwendige Bedingung mehr für ($K2^s$).

Wir haben nun gesehen, daß sich die Hauptvertreter des Strukturalismus über die Unterschiede zwischen absoluten und anwendungsrelativen Kriterien für die Modellklassen nicht völlig im Klaren waren. Leider ist aber auch Kamlahs (1985, S. 140) eigener Vergleich seines Ansatzes mit dem von Adams und Sneed keine große Hilfe. Denn erstens hält er offenbar ($K1^{s\prime\prime}$) für äquivalent mit ($K1^s$), was, wie wir sahen, ein Irrtum ist. Wichtiger ist, daß sein einziges Argument gegen ($K2^s$) gar kein Gegenargument ist: Die Existenz von T_1-Modellen, die „den mathematischen Teilen von T_2 widersprechen", d.h. die höchstens T_2-Nichtmodelle als Urbilder bzgl. $\mathbf{F}$ haben, ist bestens mit ($K2^s$) verträglich. Und bei dem Versuch, aus Kamlahs Idee ein Gegenargument zu machen, indem man die Rollen von

T_1 und T_2 vertauscht, stößt man nur dann auf ein Problem, wenn man mit Kamlah (1985, S. 133) davon ausgeht, daß $M^{\circ}_{p2}=M_{p2}$ — aber dies wird üblicherweise explizit verneint.

1.6 Erklärungen: der Vorschlag von T. Mormann

Der letzte Beitrag zur Diskussion des strukturalistischen Reduktionsbegriffs stammt von Mormann. Er betrachtet intuitive Adäquatheitsbedingungen für Reduktionen und bemerkt hier (genau) eine Inkonsistenz. Eines seiner Kriterien kennen wir noch nicht, und mit ihm kommt nun auch eine strukturalistische Minimalexplikation des Erklärungsbegriffs ins Spiel. Ausgehend von einer Stelle aus Sneed (1976, S. 138f) und Balzer und Sneed (1977, S. 204), formuliert Mormann (1984, S. 14) ein Kriterium, welches er „die Bedingung der potentiellen Erklärungsverschärfung" nennt:

(K11) Nicht alle Erklärungen von T_1 behalten im Lichte von T_2 ihre Gültigkeit.

(K11) interpretiert die intuitive Anforderung der Überlegenheit von T_2 gegenüber T_1 dahingehend, daß T_2 genauer als T_1 ist und deshalb manche T_1-Erklärungen als Scheinerklärungen entlarvt. Mormann (1984, S. 15) schlägt den folgenden Erklärungsbegriff vor:

1.6.1. Definition Sei $T=\langle M,I\rangle$ eine Theorie. (a) Ein *potentielles Modell* $x\in M_p$ heißt genau dann *von T erklärt*, wenn $x\in M$; (b) eine *intendierte Anwendung* $x\in I$ heißt genau dann *von T erklärt*, wenn $x\in M$.

Der Terminus „Erklärung" klingt hier zugegebenermaßen viel besser, wenn man den vollen Sneedschen Apparat mitverwendet. Je nachdem, ob man (a) oder (b) bevorzugt, kann man *Erklärungen von T* einfach mit (Mengen von) Modellen oder (Mengen von) erfolgreichen Anwendungen gleichsetzen. Mormanns Gebrauch schwankt hier ein bißchen, weshalb wir beide Varianten verfolgen werden.[29] Der folgende Explikationsvorschlag Mormanns versetzt uns schließlich in die Lage, (K11) auf eine strukturalistische Art und Weise niederzuschreiben:

1.6.2. Definition Eine Erklärung $E_1 \subseteq M_1(\cap I_1)$ von T_1 heißt genau dann *im Lichte von T_2 gültig*, wenn es ein $E_2\subseteq M_2$ mit $\mathbf{F}[E_2]=E_1$ gibt.

[29] Mir scheint Teil (b) von Definition 1.6.1 um eine Winzigkeit genauer zu sein. Fragt man sich, was die Theorie T *nicht* erklärt, so ergibt sich nach (a) $\mathcal{C}M$, während (b) die Menge $I\backslash M$ der Anomalien auszeichnet, was den gängigen Sprachgebrauch wohl präziser trifft.

Offensichtlich ist das Definiens von Definition 1.6.2 mit $\mathbf{F}[M_2] \supseteq E_1$ äquivalent (man nehme $E_2{:=}\mathbf{F}^{-1}[E_1]\cap M_2$). Und auch die Quantifikation in (K11) kann man loswerden, indem man die „maximale Erklärung" von T_1, nämlich $M_1(\cap I_1)$, verwendet. Damit lauten die beiden Reformulierungen von (K11)

(K11^s) $\mathbf{F}[M_2] \not\supseteq M_1$,
(K11$^{s\prime}$) $\mathbf{F}[M_2] \not\supseteq M_1\cap I_1$.

(K11^s)] ist in der Tat die Bedingung, auf die in den erwähnten Stellen von Sneed und Balzer/Sneed angespielt wird. Die stärkere Bedingung (K11$^{s\prime}$) hingegen ist neu. Allerdings halte ich Definition 1.6.2 nicht für die natürlichste Definition und möchte eine naheliegendere Alternative vorschlagen. Indem wir von Constraints absehen, brauchen wir Erklärungen nicht zu Mengen zusammenfassen, sondern es genügt, wenn wir sie sozusagen punktweise betrachten.

1.6.3. Definition Eine Erklärung $x_1\in M_1$ (oder $x_1\in M_1\cap I_1$) von T_1 heißt genau dann *im Lichte von T_2 gültig*, wenn es eine Erklärung $x_2\in M_2$ (bzw. $x_2\in M_2\cap I_2$) mit $\mathbf{F}[x_2]=x_1$ gibt.

Mit dieser neuen Bestimmung ändert sich gar nichts, wenn man an Definition 1.6.1(a) anknüpft, jedoch wird (K11$^{s\prime}$) zu einer dritten Fassung von (K11) abgeschwächt, wenn man die intuitiv wohl um eine Spur genauere Definition 1.6.1(b) anwendet:

(K11$^{s\prime\prime}$) $\mathbf{F}[M_2\cap I_2] \not\supseteq M_1\cap I_1$.

Schon die natürlichsprachliche Formulierung des Kriteriums (K11) macht klar, daß es mit (K4) inkonsistent sein muß. Gemäß unserer doppelten Buchführung in Definition 1.6.1 geht nun (K4) in die folgenden Inklusionen über:

(K4^s) $\mathbf{F}[M_2] \supseteq M_1$,
(K4$^{s\prime}$) $\mathbf{F}[M_2\cap I_2] \supseteq M_1\cap I_1$.

Wie zu erwarten war, sind (K4^s) und (K4$^{s\prime}$) die Negationen von (K11^s) und (K11$^{s\prime\prime}$). Eine interessantere Feststellung ist, daß die beiden Versionen von (K4) bereits unabhängige Stützung gefunden haben: (K4^s) stellt sich als die Mayr-Pearcesche Bedingung (K2$^{s\prime}$) heraus, und (K4$^{s\prime}$) ist genau die Pearcesche Zusatzbedingung (K8$^{s\prime}$)(b). Insofern liefert das Kriterium (K4) überraschenderweise Argumente für die Kritiker des Adams-Sneedschen Reduktionskonzepts, welches aber wiederum von (K11) gestützt wird. Man beachte aber, daß die stärkste Interpretation von (K11), nämlich Mormanns (K11$^{s\prime}$), im Widerspruch steht zur stärksten Interpretation des Balzer-Sneed-Kriteriums (K10), nämlich zur Bedingung (K10$^{s\prime\prime}$).

Mormann (1984) glaubt die Inkonsistenz von ($K1^s$) (und ($K11^{s\prime}$)) mit ($K2^{s\prime}$) auflösen zu können. Zu diesem Zweck macht er ausgiebigen Gebrauch von Constraints, so daß sein Versuch nicht mehr in den Rahmen dieses Kapitels gehört. Ohnehin ist die hier diagnostizierte Widersprüchlichkeit der verschiedentlich vorgebrachten Kriterien weit schlimmer als von Mormann angenommen.

1.7 Zwischenbilanz

Es ist nicht ganz einfach, den Überblick über die Diskussion zu behalten. Aus den verschiedenen Kriterien ergaben sich Bedingungen für intendierte Anwendungen (die Formalisierungen zu (K1)), für Modelle ((K2), Versionen von (K4), (K8) und (K11)), für intendierte Anwendungen *und* Modelle ((K5) und (K6)), für Anomalien ((K3), (K9) und eine Version von (K10)), für erfolgreiche Anwendungen (Versionen von (K4), (K9), (K10) und (K11)) und für Anomalien *und* erfolgreiche Anwendungen ((K7)). Fassen wir die Beziehungen zwischen diesen Bedingungen noch einmal zusammen:

1.7.1 Theorem (Implikationen)

(a) $(K1^{s\prime}) \Rightarrow (K1^s)$, $(K1^{s\prime\prime}) \Leftrightarrow (K1^s) \wedge (Z)$;
(b) $(K1^{s\prime\prime}) \wedge (K2^s) \Rightarrow (K9^s) \Rightarrow (K3^s)$;
(c) $(K1^s) \wedge (K2^{s\prime}) \Leftrightarrow (K5^s)$;
(d) $(K2^{s\prime}) \Leftrightarrow (K4^s) \Leftrightarrow (K8^{s\prime})(a) \Rightarrow (K8^{s\prime\prime})(a)$;
(e) $(K4^{s\prime}) \Leftrightarrow (K8^{s\prime})(b) \Rightarrow (K8^{s\prime\prime})(b)$;
(f) $(K10^{s\prime\prime}) \Rightarrow (K10^{s\prime}) \Leftrightarrow (K9^s)$;
(g) $(K11^{s\prime}) \Rightarrow (K11^s)$, $(K11^{s\prime}) \Rightarrow (K11^{s\prime\prime})$;
(h) $(K8^{s\prime})(b) \wedge (K9^s) \Rightarrow (K1^s)$.

Soweit ich sehe, ist diese Liste vollständig. (Die oben nicht gegebenen Beweise mag der interessierte Leser selbst nachtragen.) Die in Theorem 1.7.1 unerwähnt gebliebenen oder nicht eingebundenen Kriterien werden im nächsten Theorem wirksam, welches man leicht mit Hilfe von 1.7.1 vervollständigen kann. Man beachte insbesondere die gewissermaßen komplementären Rollen von (K7) und (K11) in der Kritik des Adams-Sneedschen bzw. des Mayr-Pearceschen Vorschlags:

1.7.2. Theorem (Inkompatibilitäten) Die folgenden Kriterienpaare sind logisch unvereinbar:

(a) $(K2^s)$ — $(K7^s)$;
(b) $(K2^{s\prime})$ — $(K11^s)$ (oder äquivalent $(K4^s)$ — $(K11^s)$) ;
(c) $(K1^s) \wedge (K2^{s\prime})$ — $(K6^s)$;

(d) (K8$^{s\prime}$)(b) — (K11$^{s\prime\prime}$) (oder äquivalent (K4$^{s\prime}$) — (K11$^{s\prime\prime}$));
(e) (K7^{s}) — (K9^{s}) ;
(f) (K10$^{s\prime\prime}$) — (K11$^{s\prime}$) ;
das folgende Kriterienpaar ist praktisch unvereinbar:
(g) (K3^{s}) — (K7^{s}) (denn es macht $I_2 \subseteq M_2$ unmöglich) .

Wem nun — wie mir — die von Mayr, Pearce, Kamlah und Mormann vorgetragenen Kritiken des strukturalistischen Reduktionskonzepts trotz ihrer Widersprüchlichkeit allesamt irgendwie plausibel erscheinen, der kann versuchen, die traditionellen Adams-Sneedschen Kriterien durch neuangesetzte Kriterien zu ersetzen. Einerseits kann man (K1$^{(s)}$) wohl ohne Verlust aufgeben, weil prima facie nicht jede intendierte Anwendung von T_1 ein **F**-Korrelat haben muß: *Zunächst* braucht man sich noch gar nicht darum sorgen, daß die Anomalien von T_1 in T_2 widergespiegelt werden. Andererseits muß man nicht auf (K2$^{(s)}$) bestehen, wenn man die — künstliche? — Trennung von Anwendungs- und Formalaspekt einer Theorie aufhebt:[30] man kann Kamlahs Idee folgen und fordern, daß anstelle der Gesetze die empirischen Behauptungen von T_1 und T_2 in der Folgerungsrelation stehen sollen. Kombinieren wir also (K9^{s}) mit dem von Mayr und Pearce bevorzugten Haupkriterium (K2$^{s\prime}$) zu

(K12) $\quad M_1 \subseteq \mathbf{F}[M_2] \subseteq \mathbf{F}[\mathcal{C}I_2 \cup M_2] \subseteq \mathcal{C}I_1 \cup M_1$.

Die mittlere Inklusion ist natürlich trivial. Zur Veranschaulichung füge ich noch einmal hinzu, was nach (K12) bezüglich der Modelle und der Anomalien nicht passieren darf, wenn T_1 auf T_2 reduzierbar ist: Erstens darf kein T_2-Modell vermittels **F** eine Anomalie von T_1 repräsentieren (falls dies vorkommen sollte, so entferne das betreffende T_2-Modell aus M°_{2p}), und zweitens darf kein T_1-Modell nur T_2-Anomalien als Urbilder haben.

Da (K11$^{s\prime\prime}$) die beste Lesart von (K11) sein dürfte, ist der Widerspruch von (K12) mit Mormanns (K11^{s}) (oder mit (K11$^{s\prime}$)) nicht sehr störend. Wir wissen aber auch, daß (K12) mit der Mayrschen Idee der Erklärung von T_1-Anomalien durch T_2 unverträglich ist. *Nachdem* man sich jedoch der Existenz einer intuitiv überzeugenden Reduktionsfunktion **F** gemäß (K12) versichert hat, kann man daran gehen zu prüfen, ob sich für einige T_1-Anomalien $x_1 \in I_1 \setminus M_1$ T_2-Gegenstücke in $(I_2 \cap M_2) \setminus M^{\circ}_{p2}$ finden lassen. Verläuft diese Prüfung positiv, kann man in einem zweiten Schritt den Definitionsbereich M°_{p2} der Reduktionsfunktion **F** geeignet erweitern, so daß eine Funktion **F*** entsteht, die (K7^{s}) — aber natürlich nicht mehr

[30] Dies tut zum Beispiel auch Niiniluoto (1980, S. 26f) bei seiner Minimaltransformation des Non-statement in den Statement view, indem er einer Struktur (M,I) die Aussage „Jedes I ist ein M" zuordnet.

(K12) — erfüllt.[31]

Ich muß allerdings gestehen, daß mir auch dieser Zweistufenplan zur Rettung des strukturalistischen Reduktionskonzepts sehr ad hoc und durchaus nicht hieb- und stichfest vorkommt. Sicherlich ist in *jeder* der vorgestellten Variationen der Adamsschen Bedingungen die so einfache und klare Idee des alten empiristischen Konzepts von Definierbarkeit und Ableitbarkeit verloren gegangen. Deshalb bin ich geneigt, (K12) zu verwerfen und bzgl. des Reduktionsbegriffs ganz andere Konsequenzen aus der obigen Diskussion zu ziehen.

Vergegenwärtigen wir uns noch ein letztes Mal unser Ergebnis: Die von verschiedener Seite vorgeschlagenen (notwendigen) Adäquatheitsbedingungen für Reduktionen widersprechen sich auf vielfältige Weise. Hält man sich vor Augen, daß wir in wirklich fortschrittlichen Reduktionsverhältnissen anstelle von $\subseteq$ und $\supseteq$ die strikten Inklusionen $\subsetneq$ und $\supsetneq$ zu erwarten haben, dann ist offensichtlich, daß schon ($K2^s$) und ($K2^{s\prime}$) miteinander unverträglich sind. Ich habe in der Literatur keine Hinweise darauf gefunden, daß man sich dieses doch ziemlich drastischen Problems bewußt ist.[32] Man kann nur darüber spekulieren, warum das bisher nicht klar geworden ist. Erstens verstellen wohl die ständigen Wechsel der Darstellung (Notationen, Konventionen und Definitionen) den unmittelbaren Blick auf die Sachlage. Zweitens wird diese Wirkung verstärkt durch den z.T. beträchtlichen technischen Aufwand, der durch die Unterscheidung von partiellen potentiellen und potentiellen Modellen und durch Constraints verursacht wird. Obschon diese Unterscheidung und Constraints an sich wohlbegründet sind, dürften ihre Betrachtung für die speziellen Probleme des Reduktionsbegriffs abdingbar sein.[33] Vielleicht waren die Ergebnisse dieses Kapitels nur

[31] Vgl. aber Fußnote 7.

[32] Hoerings (1984, 47f) Vermutung, daß die „requirements of definability and derivability" (K1) und (K2) manchmal abgeschwächt werden müßten, um Inkonsistenzen zu vermeiden, ist hier nicht einschlägig. Er nennt als Gründe Approximationen und Undefinierbarkeiten theoretischer Funktionen — Aspekte, die ich in diesem Kapitel ausgeklammert habe.

[33] So lernt man etwa aus den Überlegungen, die Sneed (1971) auf S. 224f anstellt, zwar etwas über die technischen Probleme seines raffinierteren Theorienbegriffs, aber kaum etwas Substantielles über Probleme der Reduktion von Theorien. — Im Gegensatz zu mir ist Mormann (1984) der Meinung, daß die von ihm aufgespürte Inkonsistenz (vgl. Theorem 1.7.2(b)) gerade durch besondere Berücksichtigung von Constraints zu überwinden sei. In diesem Sinne modifiziert er die strukturalistischen Formulierungen von (K8) und (K5). Ich denke aber, daß man — wenn überhaupt — dann konsequenterweise *alle* Kriterien entsprechend reformulieren müßte und daß damit die Probleme in analoger Weise wieder auftauchen würden.

deshalb so mühelos zu gewinnen, weil ich mich auf das vereinfachte Theorienmodell von Adams beschränkt habe und alle Kriterien als Relationen zwischen bequem zu handhabenden Bild- und Urbildmengen von **F** formuliert habe. Aber dies ist noch keine befriedigende Erklärung dafür, daß die Spezialisten auf diesem Gebiet ein schiefes[34] oder gar unrichtiges[35] Bild der Zusammenhänge zwischen den Kriterien gegeben haben. Erklären könnte man dies durch die Annahme, daß sich der Reduktionsbegriff zumindest in seinem strukturalistischen Gewand auf zu wenig ausgegorene Intuitionen gründet; daß all die vorgebrachten Kriterien auf die eine oder andere Art und Weise legitim sind, aber doch nur Ausdruck einer Familienähnlichkeit der Relationen zwischen einander ablösenden Theorien sind; kurz, daß nach einem anderen Raster gesucht werden muß, welches Theorienverdrängung adäquat und als rational zu rekonstruieren gestattet.

Doch es gibt noch eine andere, gutwilligere Annahme, die die Verwirrung in der Reduktionsdiskussion erklären könnte. Bisher haben wir — wie die Autoren in den referierten Arbeiten — stillschweigend vorausgesetzt, daß es nur *einen* logischen Typ von Reduktion gibt. Es wäre aber auch möglich, daß diese Präsupposition falsch ist und daß mehrere logische Typen von Reduktion unterschieden werden müssen. Dieser Idee wollen wir uns im Rest dieses Kapitels zuwenden.

Ein naheliegender Gedanke ist es, eine Dichotomie von zwei Arten von Reduktion zu postulieren, für die ($K2^s$) und ($K2^{s\prime}$) die grundlegenden charakteristischen Kriterien wären. Diese Idee findet zwar keine unmittelbare Stütze in der Diskussion der Abschnitte 1.2–1.6, da die verschiedenen Kriterien, wie wir sahen, doch in ziemlich komplizierter Weise miteinander verflochten sind. Dennoch scheinen die beiden in ($K2^s$) und ($K2^{s\prime}$) zum Ausdruck kommenden Intuitionen die strukturalistische Reduktionsdiskussion von den allerersten Anfängen an geprägt zu haben. Adams (1959) war ein ganz klarer Befürworter des Ableitbarkeitskonzepts (genauer: des D-Konzepts; vgl. Kapitel 2) der Reduktion, und dementsprechend stammt von ihm die Bedingung ($K2^s$). Von vielen Autoren wird der strukturalistische Begriff der Reduktion aber auf die Logik-Einführung von Suppes

[34] So ist Pearces Berufung auf Stegmüllers (K8) wie erwähnt zumindest ungenau. Außerdem konnte ich Pearces (1982, S. 309, 323) Behauptung, daß Mayr seine Kritik am Adams-Sneed-Konzept aufgrund ebendieses Kriteriums (K8) vortrage, nicht verifizieren. Hoerings (1984, S. 38f) „Pearce-Darstellung" (eigentlich handelt es sich hier um eine Darstellung von Pearces Sneed-Darstellung — aber Pearce favorisiert andere Kriterien als Sneed!) kann hier oder auch in der nächsten Fußnote genannt werden.

[35] Ich habe Stegmüller (1985, S. 146), Sneed (1976, S. 139), Mayr (1976, S. 276, 284, 286f) und Kamlah (1985, S. 140) erwähnt.

(1957) zurückgeführt. Dies ist eigentlich erstaunlich, denn Suppes macht in diesem Buch nur eine kleine Bemerkung zum Thema:

(1.7.3) To show in a sharp sense that thermodynamics may be reduced to statistical mechanics, we would need to axiomatize both disciplines by defining appropriate set-theoretical predicates, and then show that given any model T of thermodynamics we may find a model of statistical mechanics on the basis of which we may construct a model isomorphic to T. (Suppes 1957, S. 271)

Suppes verlangt also, daß es für jedes $x_1 \in M_1$ ein $x_2 \in M_2$ geben muß, aus dem ein zu x_1 „isomorphes" Modell „konstruiert" werden kann. Über Isomorphie und die Art der Konstruktion sagt Suppes nichts, es scheint aber angemessen, diese Dinge im gegenwärtigen Rahmen als inhaltliche Vorbedingungen an die Reduktionsfunktion **F** zu deuten. Vom logischen Standpunkt aus ist wichtig, daß zu jedem $x_1 \in M_1$ ein irgendwie — über **F** — korrespondierendes $x_2 \in M_2$ existieren soll. Dies ist aber das Kriterium ($K2^{s\prime}$).

So kurz und vage die Formulierung von (1.7.3) auch ist, so hat sie doch genauso viel Beachtung gefunden wie der Adamssche Beitrag. Schaffner (1967, S. 139) spricht vom „Suppes-Paradigma" oder „Suppes-Adams-Paradigma" der Reduktion, Eberle (1971, S. 491–494) betrachtet „Suppes-Adams-Reduktionen" und Pearce (1987, S. 22f, 93f) widmet dem „Suppes-Adams-Konzept" (Pearce 1987, S. 26) einige Seiten. Natürlich sind die Reduktionskonzepte von Suppes und Adams insofern verwandt, als sie beide im strukturalistischen Rahmen vorgestellt wurden. Aber der prinzipielle logische Unterschied zwischen dem Adamsschen ($K2^s$) und dem Suppesschen ($K2^{s\prime}$) sollte nicht verwischt werden. Schaffners (1967, S. 145) Theorem, wonach eine Reduktion nach dem D-konzept eine Suppessche Reduktion impliziert, ist, wie auch Pearce (1987, S. 27) bemerkt, nicht korrekt.[36] Richtig stellt dagegen Eberle (1971, S. 493) bei Suppes eine Umkehrung der intuitiv erwarteten Verhältnisse fest. Day (1985, S. 169f) schlägt vor, Suppes' Reduktionsbegriff durch ($K2^s$) *und* ($K2^{s\prime}$) zu formalisieren — ohne jedoch anzugeben, wo er bei Suppes das Kriterium ($K2^s$) herausgelesen

[36] Während Pearce bei Schaffner eine falsche Verwendung von Isomorphismen und Reduktionen à la Suppes konstatiert, sehe ich andere Fehlerquellen: Schaffner benützt bei seinem „Beweis" nur die Definitions-, nicht aber die Ableitbarkeitsbedingung des D-Konzepts, und er wendet sich gar nicht der Frage zu, ob ein gerade in Rede stehendes potentielles Modell auch ein Modell ist. Schaffners Beweisskizze ist unklar, aber was sie meiner Ansicht nach zeigt, ist lediglich, daß unter der Voraussetzung einer D-Reduktion für jedes (potentielle) Modell von T_1 ein korrespondierendes *potentielles* Modell von T_2 gefunden werden kann.

hat.[37]

Schließlich kann ich David Pearce (1987, S. 22) nicht zustimmen, wenn er sagt, daß Suppes' Kriterium (1.7.3) „im wesentlichen eine semantische Variante des üblichen syntaktischen Begriffs der (relativen) Interpretierbarkeit" ist und daß Adams „eine Verfeinerung und auch eine Verallgemeinerung" des Suppesschen Kriteriums angibt.

Zusammengefaßt: Wir haben mit den alten Kriterien von Suppes und Adams offenbar zwei alternative, deutlich verschiedene Ansätze zur Analyse von Reduktionen vorliegen und gleichzeitig möglicherweise die Kerne einer Separation der Intuitionen in zwei kohärente Reduktionsbegriffe gefunden. In den nächsten beiden Abschnitten sollen zwei bemerkenswerte Versuche, eine solche Dichotomisierung auf systematische Weise durchzuführen, besprochen werden. Es handelt sich dabei um die Vorschläge von Pearce und Rantala und von Scheibe. Wir wollen prüfen, ob beide Seiten, die Suppessche und die Adamssche Idee, gleichermaßen gut motiviert sind und ob eine logische Zweiteilung des Reduktionsbegriffs wirklich zwingend ist.

1.8 Typen von Reduktion I: Pearce und Rantala über theoretische und erklärende Reduktion

David Pearce und Veikko Rantala (z.B. 1983b; 1984a) können für sich in Anspruch nehmen, seit einigen Jahren schon mit ihrer formalen Präzisierung des Begriffs der Korrespondenz die Grundlage für eine Typologie verschiedener Reduktionsbegriffe gelegt zu haben. Wir wollen kurz und in vereinfachter Form ihre fundamentalen Definitionen wiedergeben. Auch bei Pearce und Rantala steht eine Reduktionsfunktion **F** im Zentrum, die allerdings weder im Definitions- noch im Wertebereich über die Modellmengen der in Rede stehenden Theorien hinausgeht:

(1.8.1) $\mathbf{F}$: $M_2^\circ \rightarrow M_1$, wobei $M_2^\circ \subseteq M_2$.

Wir verwenden im folgenden weiterhin die Schreibweisen $\mathbf{F}[X_2]$ und $\mathbf{F}^{-1}[X_1]$ mit denselben Festlegungen wie oben; es ist hier aber stets daran zu denken, daß $\mathbf{F}[X_2]$ für beliebiges $X_2 \subseteq M_{p2}$ eine Teilmenge von M_1 (nicht nur von M_{p1}) und daß $\mathbf{F}^{-1}[X_1]$ für beliebiges $X_1 \subseteq M_{p1}$ eine Teilmenge von M_2° (nicht von M_{p2}°) ist. Intuitiv jedoch, so scheint es mir, sollte eine

[37] In seiner Fußnote 3 bemerkt Day (1985) richtig, daß Adams nicht mit Suppes (sondern eher mit Nagel) über einen Kamm geschert werden darf.

Reduktionsfunktion stets „dieselben" *potentiellen* Modelle von T_1 und T_2 verknüpfen,[38] unabhängig davon, was die einzelnen Theorien über diese potentiellen Modelle sagen, d.h. welche sie als Modelle auszeichnen. Deshalb werde ich bei den folgenden Überlegungen *intuitiv* (nicht formal) immer von einem $\mathbf{F}$ mit Definitionsbereich M_{p2}° und Wertebereich M_{p1} ausgehen.

Zu einer Korrespondenz nach Pearce und Rantala gehört als ebenso wesentlicher zweiter Bestandteil eine Übersetzungsfunktion $\ddot{\mathbf{U}}$, die allen Sätzen der Sprache von T_1 als Übersetzung Sätze der Sprache von T_2 zuordnet. Die Übersetzung soll — dies ist entscheidend — die Reduktionsfunktion $\mathbf{F}$ „respektieren", d.h. es soll gelten:

(1.8.2) $\forall\, x_2 \in M_2^{\circ}\ \forall\ T_1\text{-Sätze } A\ (\ \mathbf{F}(x_2) \models A \text{ gdw. } x_2 \models \ddot{\mathbf{U}}(A)\ )$

Was hat nun eine solchermaßen mit $\mathbf{F}$ abgestimmte Übersetzung mit dem in Abschnitt 1.4 eingeführten Übersetzungsbegriff zu tun? Um dies zu sehen, müssen wir die eben angeschriebene Bedingung in eine vertrautere Form bringen:

(1.8.3)
$$\begin{aligned}
&\forall x_2 \in M_2^{\circ}\ \forall T_1\text{-Sätze } A\ (\mathbf{F}(x_2) \models A \Leftrightarrow x_2 \models \ddot{\mathbf{U}}(A)) \Leftrightarrow \\
&\Leftrightarrow \forall\, x_2 \in M_2^{\circ}\ \forall T_1\text{-Sätze } A\ (\mathbf{F}(x_2) \in ||A|| \Rightarrow x_2 \in ||\ddot{\mathbf{U}}(A)||) \wedge \\
&\qquad\qquad x_2 \in ||\ddot{\mathbf{U}}(A)|| \Rightarrow \mathbf{F}(x_2) \in ||A||) \\
&\Leftrightarrow \forall T_1\text{-Sätze } A\ (\mathbf{F}^{-1}(||A||) \subseteq ||\ddot{\mathbf{U}}(A)|| \wedge \mathbf{F}(||\ddot{\mathbf{U}}(A)||) \subseteq ||A||) \\
&\Leftrightarrow \forall T_1\text{-Sätze } A\ (\mathbf{F}^{-1}(||A||) \subseteq ||\ddot{\mathbf{U}}(A)|| \wedge \\
&\qquad\qquad \mathbf{F}^{-1}(\mathbf{F}(||\ddot{\mathbf{U}}(A)||)) \subseteq \mathbf{F}^{-1}(||A||)) \\
&\Leftrightarrow \forall T_1\text{-Sätze } A\ (\mathbf{F}^{-1}(||A||) \subseteq ||\ddot{\mathbf{U}}(A)|| \wedge \\
&\qquad\qquad ||\ddot{\mathbf{U}}(A)|| \cap M_2^{\circ} \subseteq \mathbf{F}^{-1}(||A||)) \\
&\Leftrightarrow \forall T_1\text{-Sätze } A\ (\mathbf{F}^{-1}(||A||) = ||\ddot{\mathbf{U}}(A)|| \cap M_2^{\circ}).
\end{aligned}$$

Entsprechend dem kleineren Definitions- und Wertebereich ihrer Reduktionsfunktion $\mathbf{F}$ stellen Pearce und Rantala also kleinere Ansprüche an Übersetzungen als wir in Quasidefinition 1.4.2: $||\ddot{\mathbf{U}}(A)||$ und $\mathbf{F}^{-1}(||A||)$ müssen nicht auf M_{p2}°, sondern nur auf M_2°, was im allgemeinen eine echte Teilmenge von M_{p2}° ist, miteinander übereinstimmen.

Pearce und Rantala listen in ihren Aufsätzen eine ganze Anzahl von Spezialfällen von Korrespondenzen auf. Pearce (1987, S. 112) meint, sie alle seien „reduktiver Art". Was sind nun, im Zusammenhang der Diskussion in den vorangehenden Abschnitten, die wichtigsten Typen von Reduktion? Formal ist die Angelegenheit bei Pearce und Rantala sehr einfach

[38] Genauer soll das heißen, daß $\mathbf{F}$ jedem *potentiellen* Modell x_2 von T_2, das reale Systeme eines bestimmten Typs beschreibt, ein *potentielles* Modell x_1 von T_1 zuordnet, welches eine (im allgemeinen gröbere) Beschreibung der Systeme desselben Typs in der Sprache von T_1 darstellt. Dies setzt voraus, daß die Mittel einer T_1-Beschreibung in T_2 rekonstruierbar sind.

und befriedigend angelegt:

1.8.4. Definition Sei **F** die Reduktionsfunktion einer Korrespondenz $\langle \mathbf{F}, \ddot{\mathbf{U}} \rangle$. Wenn der Definitionsbereich von **F** gleich der Modellmenge von T_2 ist, d.h. wenn $M_2^\circ = M_2$, dann heißt $\langle \mathbf{F}, \ddot{\mathbf{U}} \rangle$ eine *Interpretation*; wenn M_2° nicht nur in, sondern sogar auf die Modellmenge von T_1 abgebildet wird, d.h. wenn **F** surjektiv ist, heißt $\langle \mathbf{F}, \ddot{\mathbf{U}} \rangle$ eine *Einbettung*.[39]

Der Zusammenhang von Interpretationen und Einbettungen mit unserer bisherigen Thematik kann so hergestellt werden: Verändert man den Definitionsbereich einer *interpretierenden* Reduktionsfunktion **F** von M_2° zu einer beliebigen Teilmenge M_{p2}° von M_{p2}, wobei $\mathbf{F}|M_2^\circ \cap M_{p2}^\circ$ erhalten bleibe, dann erfüllt das neue **F** in jedem Fall $\mathbf{F}[M_2] \subseteq M_1$, d.h. ($K2^s$). Erweitert man den Definitionsbereich einer *einbettenden* Reduktionsfunktion **F** von M_2° auf eine Obermenge M_{p2}°, dann erfüllt das neue **F** in jedem Fall $\mathbf{F}[M_2] \supseteq M_1$, d.h. ($K2^{s\prime}$).[40]

Schränkt man umgekehrt ein ($K2^s$) erfüllendes **F** mit Definitionsbereich $M_{p2}^\circ \supseteq M_2$ auf M_2 ein, so erhält man eine Interpretation.[41] Schränkt man ein ($K2^{s\prime}$) erfüllendes **F** mit Definitionsbereich M_{p2}° auf $M_2^\circ := M_{p2}^\circ \cap M_2$ ein, so erhält man eine Einbettung.

Wenn die Parallele zwischen Interpretationen und Einbettungen einerseits und ($K2^s$) und ($K2^{s\prime}$) andererseits auch nicht ganz perfekt durchgezogen werden kann, so wird man doch sagen dürfen, daß Pearce und Rantala mit ihrer sehr einfach und elegant darzustellenden Dichotomie „Interpretation vs. Einbettung" der in den Abschnitten 1.2–1.6 rekonstruierten Diskussion eine klare theoretische Grundlage geben können — eine Grundlage, die ihr offensichtlich gefehlt hat. Ich werde mich in diesem Abschnitt nur noch mit den Begriffen von Pearce und Rantala beschäftigen. Im Gegensatz zu den früheren Abschnitten, wo ich auf eine ausführliche intuitive Diskussion der Kriterien verzichtet habe, will ich nun — wo es um die Begründung einer systematischen Typologie geht — die Motivation der

[39] Pearce und Rantala verwenden in den meisten Arbeiten für diesen zweiten Fall den Terminus „Reduktion". „Reduktion" erscheint erst bei Pearce (1985; 1987) als Überbegriff von „Interpretation" und „Einbettung".

[40] Hinter einer reinen *Erweiterung* von M_2° auf M_{p2}° steckt als Idee natürlich die naheliegende Gleichung $M_2^\circ = M_{p2}^\circ \cap M_2$. — Allerdings erfüllt gemäß dieser Idee eine beliebige Erweiterung *jeder* Korrespondenz das Kriterium ($K2^s$).

[41] Schränkt man ein ($K2^s$) erfüllendes **F** mit beliebigem Definitionsbereich M_{p2}° auf $M_2^\circ := M_{p2}^\circ \cap M_2$ ein, so erhält man eine gewöhnliche Korrespondenz. Reduktionsfunktionen mit beliebigen Definitionsbereich M_{p2}°, die *nicht* ($K2^s$) erfüllen, sind ganz ausgeschlossen, falls man verlangen will, daß ihre Einschränkung auf $M_2^\circ := M_{p2}^\circ \cap M_2$ eine Korrespondenz sein muß (vgl. Fußnote 40).

formalen Definitionen von Pearce und Rantala eingehend unter die Lupe nehmen. Hierbei halte ich mich vor allem an Pearce (1987), wo eine aktuelle, zusammenfassende und nicht allzu knappe Darstellung dieser Motivation gegeben ist.[42] Der Schwerpunkt meiner folgenden Überlegungen liegt ganz eindeutig auf der Seite der Einbettung. Während ich gegen die intuitive Basis von Interpretationen[43] nichts einzuwenden habe, werde ich einige Zweifel an der Notwendigkeit und Begründbarkeit von Einbettungen und damit auch an der Notwendigkeit und Begründbarkeit einer Aufspaltung des Reduktionsbegriffs in zwei hauptsächliche logische Typen anmelden.

Das formale Begriffspaar „Interpretation vs. Einbettung" ist von Pearce und Rantala dazu gedacht, ein intuitives Begriffspaar „theoretische Reduktion vs. erklärende Reduktion" zu präzisieren.[44] Dies letztere Paar wird von Pearce (1987, S. 92) folgendermaßen vorgestellt:

> *explanatory* reduction ... can be said to occur when the problems solved and explanations proffered within one framework or theory can be reduced to problems and explanations handled by another framework or theory. ... *theoretical* reduction ... cover[s] those cases where a scientific principle, a theory, or even an entire branch of science, is reduced to a more fundamental principle, theory or science. What are explained here are therefore not (sets of) problems, but rather (sets of) laws.

Bezüglich der theoretischen Reduktion, wo das „reduced to" im Explikat wohl mit „derivable from" gleichgesetzt werden darf, habe ich, wie erwähnt, nichts zu sagen. Meine Aufmerksamkeit soll von nun an der erklärenden Reduktion gelten, bei der zunächst nicht klar ist, wie das „reduced to" im Explikat aufgefaßt werden soll. Näheres darüber erfahren wir an anderer Stelle in Pearces Buch. Die „potentielle Erklärung einer 'Tatsache' oder eines 'Problems' ", welche(s) durch einen T_1-Satz E ausgedrückt werde, durch

[42] Die Mehrzahl der einschlägigen, im Buch von Pearce vorgebrachten Argumente kann man verstreut schon in den zahlreichen gemeinsamen Aufsätzen von Pearce und Rantala finden; siehe die Anmerkungen und Bibliographie in Pearce (1987).

[43] Genauer gesagt, gegen die intuitive Basis des Kriteriums ($K2^s$), welches auf die oben erwähnte Weise mit dem Begriff der Interpretation zusammenhängt.

[44] Genauere Ausführungen hierzu findet man nur bei Pearce (1985; 1987), doch ist die Idee schon in den frühen Arbeiten von Pearce und Rantala formuliert, z.B. in Pearce und Rantala (1983b, S. 368): „... if we have a reduction [=Einbettung], then, e.g., any 'explanation' in T_1 can be transformed to T_2 ... In the case of interpretation, the translations of the axioms of T_1 are L_2-consequences of the axioms of T_2, without any additional hypotheses." (Indizierungen von mir.)

die Theorie T_1 besteht nach Pearce (1987, S. 102) in einer Ableitung von E aus Gesetzen $G_1,\ldots,G_n$ und 'Anfangsbedingungen' $A_1,\ldots,A_k$. Da die Gesetze wohl aus T_1 sein sollen[45] und da man die Anfangsbedingungen zu ihrer Konjunktion $A_1\wedge\ldots\wedge A_k$ — abgekürzt durch „A" — zusammenfassen kann, schreiben wir dies einfach so:

(1.8.5) $T_1, A \vdash E$.

Pearce beginnt seine Argumentation mit der Feststellung, daß das Explanans einer Einzelfallerklärung konsistent sein muß, daß es also ein $x_1\in M_1\cap||A||$ geben muß. Falls nun eine *Einbettung* **F** von T_1 in T_2 vorliegt, gibt es auch ein $x_2\in M_2\cap||\ddot{\mathbf{U}}(T_1)||\cap||\ddot{\mathbf{U}}(A)||$ mit $\mathbf{F}(x_2)=x_1$, also ist die Konjunktion von $\ddot{\mathbf{U}}(T_1)$ und $\ddot{\mathbf{U}}(A)$ mit T_2 konsistent, was nach Pearce wiederum heißt, daß $\ddot{\mathbf{U}}(T_1)$ und $\ddot{\mathbf{U}}(A)$ „die Basis für eine Erklärung des (übersetzten) Explanandums $\ddot{\mathbf{U}}(E)$ durch T_2 bilden."[46] Damit beschließt Pearce den relevanten Absatz. Ich halte diese Formulierung aber erstens für allzu vage, und zweitens scheint sie mir, was sie auch immer heißen mag, den eigentlichen Punkt einer erklärenden Reduktion nicht genau zu treffen.

Näher an die Pointe einer erklärenden Reduktion kommt Pearce (1987, S. 128) meines Erachtens an anderer Stelle seines Buch: Die Erklärung von E in T_1 gemäß (1.8.5) wird übertragen auf eine Erklärung in T_2, wenn

(1.8.6) $T_2, A' \vdash \ddot{\mathbf{U}}(E)$

für „eine geeignete Wahl der Bedingungen" A' gilt. Doch auch dies ist nicht genügend aussagekräftig: Was kann ein geeignetes A' sein? Nach meiner Intuition ist die einleuchtendste Idee die, daß man einfach die Anfangsbedingungen A in die Sprache von T_2 übersetzt, d.h.

(1.8.7) $T_2, \ddot{\mathbf{U}}(A) \vdash \ddot{\mathbf{U}}(E)$.

Man könnte sich allerdings wohl — wie Pearce (1985, S. 266) — auch mit der Idee

(1.8.8) $T_2, \ddot{\mathbf{U}}(T_1), \ddot{\mathbf{U}}(A) \vdash \ddot{\mathbf{U}}(E)$

[45] Es würde an den folgenden Überlegungen nichts Wesentliches ändern, wenn Pearce dies nicht intendiert hätte.

[46] „T_1" stehe hierbei für „das" (einzige) Axiom der Theorie T_1. — Genaugenommen sagt Pearce (1987, S. 102) daß $\ddot{\mathbf{U}}(G)$ (G sei eine Abkürzung für $G_1\wedge\ldots\wedge G_n$) und $\ddot{\mathbf{U}}(A)$ diese Basis bilden. Wegen $M_1\subseteq||G||$ gilt aber $\mathbf{F}^{-1}[M_1]\subseteq\mathbf{F}^{-1}[||G||]$, also wegen der Surjektivität einer Einbettung **F** und obiger Rechnung für Übersetzungen auch $M_2^\circ\subseteq||\ddot{\mathbf{U}}(G)||$. Dies heißt aber, daß, grob gesagt, M_2 (bzw. T_2) innerhalb von M_2° — und nur hier sind die Übersetzungen von T_1-Sätzen festgelegt — mindestens genauso viel leistet wie $||\ddot{\mathbf{U}}(G)||$.

begnügen. Leider geht aber weder der erstere noch der letztere Ansatz auf. Denn „die Bedeutung" sämtlicher Übersetzungen von T_1-Sätzen in die Sprache von T_2 ist, wie wir in (1.8.3) sahen, nur hinsichtlich ihres Verhaltens innerhalb von $M_2^\wp$ festgelegt, und wenn M_2 eine echte Obermenge von $M_2^\wp$ ist, kann mithin eine Ableitbarkeit im gewünschten Sinne nicht mehr gewährleistet sein.

Eine Garantie, daß die Erklärung von E in T_2 vermittels „derselben" Anfangsbedingungen nachvollzogen werden kann, gibt es nur dann, wenn entweder $M_2{=}M_2^\wp$ gilt, d.h. wenn von vornherein eine *Interpretation* vorliegt, oder aber wenn wir als weitere Prämisse einen T_2-Satz C dazunehmen, welcher (zusammen mit den Axiomen von T_2) $M_2^\wp$ definiert. Um bei reinen Einbettungen bleiben zu können, betrachten wir erst einmal die zweite Alternative:

(1.8.9) $T_2, C, \ddot{\mathbf{U}}(T_1), \ddot{\mathbf{U}}(A) \vdash \ddot{\mathbf{U}}(E)$.

Durch Betrachtung der „Modellebene" kann man sofort sehen, daß sich aus der modelltheoretischen Version von (1.8.5), nämlich aus

(1.8.10) $M_1 \cap ||A|| \subseteq ||E||$

wegen (1.8.3) und $\mathbf{F}^{-1}[M_1]{=}M_2^\wp$ unmittelbar

(1.8.11) $M_2 \cap ||C|| \cap ||\ddot{\mathbf{U}}(A)|| = M_2^\wp \cap \mathbf{F}^{-1}(||A||) = \mathbf{F}^{-1}(M_1 \cap ||A||) \subseteq \mathbf{F}^{-1}(||E||) \subseteq ||\ddot{\mathbf{U}}(E)||$

ergibt. Die Übertragung der Ableitbarkeit *innerhalb von* $M_2^\wp$ folgt somit trivial für *jede* Korrespondenz, nicht nur für Einbettungen (auf die Prämisse $\ddot{\mathbf{U}}(T_1)$ kann sogar verzichtet werden). Das Besondere an Einbettungen besteht allein in der Garantie, daß auf der linken Seite der letzten Beziehung keine leere Menge steht.

Eine interessante Frage ist nun, ob T_2 wirklich *alle* Erklärungen, insbesondere alle *potentiellen* Erklärungen (d.h. wenn A zwar möglich, aber in der Realität nie verifiziert sein wird) und alle *gescheiterten* Erklärungen (d.h. wenn sich A tatsächlich als falsch herausstellt) von T_1 auf diese Weise *konsistent* widerspiegeln können sollte. Ich meine, es gibt keinen Grund für eine solche Forderung. Im Gegenteil, T_2 ist nur dann wirklich fortschrittlich gegenüber T_1, wenn es insofern „informativer" oder „genauer" als T_1 ist, als es einige von T_1 für möglich gehaltene, aber in Wirklichkeit nicht realisierbare Modelle ausschließt. Und dies kann dadurch geschehen, daß das A aus (1.8.5) im Lichte von T_2 keine korrekte Beschreibung der Anfangsbedingungen mehr ist, d.h. daß $M_2 \cap ||\ddot{\mathbf{U}}(A)||$ leer ist (d.h. nach (1.8.3) auch $\mathbf{F}^{-1}(||A||){=}\emptyset$). Ist diese Überlegung richtig, dann können wir die folgende überraschende Schlußfolgerung ziehen: Wenn Reduktionen von T_1

auf T_2 einen wissenschaftlichen Fortschritt im Übergang von T_1 auf T_2 nicht ausschließen sollen, dann darf die Pearce-Rantalasche Reduktionsfunktion $\mathbf{F}$: $M_2^\circ \rightarrow M_1$ *nicht* surjektiv, also *keine* Einbettung sein.

Es bleibt noch anzumerken, daß wir eben durch das Hinzuziehen von C eigentlich schon wieder bei einer Art *theoretischer* Reduktion gelandet sind, denn bei Einbettungen gilt natürlich

(1.8.12) $M_2^\circ \subseteq \mathbf{F}^{-1}[M_1]$,

und dies impliziert, syntaktisch gewendet, nicht weniger als

(1.8.13) $T_2, C \vdash \ddot{U}(T_1)$.[47]

Mit der Zusatzbedingung C ist also die (Übersetzung der) ganze(n) *Theorie* T_1 und nicht „nur" die Menge der Problemlösungen von T_1 konsistent aus T_2 ableitbar. Man kann daher sagen, daß die zweite Garantiemöglichkeit dafür, daß die Erklärung von E durch T_1 in T_2 vermittels „derselben" Anfangsbedingungen nachvollzogen werden kann, unsere Aufmerksamkeit weg von einer erklärenden hin auf eine theoretische Reduktion lenkt, ebenso wie die erste Alternative weg von Einbettungen hin zu Interpretationen führt.

Pearce stützt sich noch auf einen zweiten, prima facie völlig unabhängigen Argumentationsstrang zugunsten von Einbettungen. Wenn wir auch in Abschnitt 1.4 schon ein gut Teil dieser Idee kennengelernt haben, so erscheint es doch sinnvoll, ihre aktuelle Fassung in Pearces und Rantalas eigenem Rahmen noch einmal einer genaueren intuitiven Prüfung zu unterziehen. Drei relevante Formulierungen werden in Pearces Buch (1987, S. 92, 94) angeboten. Die erste stammt von Sneed (1971, S. 220):[48]

(1.8.14) If S_2 is a statement of T_2 about a certain physical system and S_1 is a statement of T_1 about a *corresponding* physical system, then S_2 is true only if S_1 is true.

David Pearces eigene Formulierungen lauten

(1.8.15) the truth in T_2 of any given statement in the language of T_1 is sufficient to ensure its truth in T_1 (1987, S. 92)[49]

und

[47] Dies hält auch Pearce (1987, S. 36) fest.

[48] Die Bezeichnungen von Theorien und Sätzen sind im folgenden von mir abgeändert worden.

[49] Genau die umgekehrte Richtung dieser Bedingung findet man als Kriterium der Wahrheitserhaltung bei Eberle (1971, S. 487) erwähnt: „If we are interested in the truth of some theory T_1, then we shall not regard it as replaceable by a theory T_2 if some sentence should be true in T_1 while its counterpart is not true in T_2."

(1.8.16) for any sentence ϕ_1 in the language of T_1, if the translation of ϕ_1 is a consequence of T_2, then ϕ_1 is itself a consequence of T_1. (1987, S. 94)

Das Zitat von Sneed (der anscheinend vergessen hat, einen Zusammenhang zwischen S_2 und S_1 vorauszusetzen) läuft — dies darf man wohl behaupten — auf das oben angeführte Kriterium (K10) hinaus (und nicht auf (K8)). (1.8.14) ist laut ausdrücklicher Angabe von Sneed dazu gedacht, das Adamssche Reduktionskonzept der Ableitbarkeit von Gesetzen, d.h. also die *theoretische* Reduktion, zu explizieren.[50] In Abschnitt 1.5 haben wir gesehen, daß (K10) nur in seiner Interpretation ($K10^{s\prime}$) (zwar nicht hinreichend, aber immerhin) notwendig für ($K2^s$) ist. Demgegenüber ist die Pearcesche Ausdrucksweise in (1.8.15) und (1.8.16) offenbar an (K8) ausgerichtet,[51] was, wie wir aus Abschnitt 1.4 wissen, jedenfalls in der plausiblen Interpretation ($K8^{s\prime\prime}$) (zwar nicht hinreichend, aber immerhin) notwendig für ($K2^{s\prime}$) ist.

Anscheinend wurde irgendwo im Übergang von Sneeds zu Pearces Formulierungen das logische Verhältnis von reduzierender und reduzierter Theorie umgedreht: Bei Sneed soll T_2 tendenziell stärker als (oder zumindest genauso stark wie) T_1 sein, bei Pearce stellt sich heraus, daß gerade der umgekehrte Fall gefragt ist.[52] Man beachte hier auch den Wechsel der metatheoretischen Prädikate: Während bei Sneed — wie auch in Stegmüllers (K8) — nur einfach von Wahrheit die Rede ist, spricht Pearce von Wahrheit-in-T_i oder, noch etwas genauer, von Konsequenzen-von-T_i. Möglicherweise liegt in dieser nicht nur terminologischen Verschiebung der Schlüssel für das Auseinanderdriften der intuitiv gefaßten Kriterien zu formal quasi komplementären Bedingungen.

Insofern Pearce also an die Sneedsche Intuition anknüpfen will, scheint er sein Ziel zu verfehlen; insofern er eigenständige Kriterien für Ein-

[50] Jedenfalls gilt das für S. 220 von Sneed (1971). Auf S. 218 formuliert Sneed praktisch wörtlich die begriffliche Charakterisierung einer *erklärenden* Reduktion und expliziert sie dann vermittels der Konjunktion von ($K1^s$) und ($K10^s$).

[51] Freilich mit den anderen, auf die Modellmengen eingeschränkten Reduktions- und Übersetzungsfunktionen **F** und **Ü**. Man kann aber leicht nachprüfen, daß sich die in Abschnitt 1.4 angestellten Überlegungen zum Zusammenhang zwischen ($K8^{s\prime}$)(a) und ($K8^{s\prime\prime}$)(a) unverändert übertragen lassen, denn die Ersetzung der Bedingung $M_2 \subseteq M_{p2}^o$ durch die Bedingung $M_2 \subseteq M_2^o$ als (unplausible) Voraussetzung für die Richtung ($K8^{s\prime\prime}$)(a)$\Rightarrow$($K8^{s\prime}$)(a) ist unwesentlich.

[52] Pearce ist sich dessen voll bewußt. Er hält es für durchaus zutreffend, daß bisweilen auch stärkere Theorien auf schwächere Theorien reduziert werden (vgl. Pearce 1987, besonders S. 104f).

bettungen als Explikate von Reduktionen geben will, wäre es zumindest wünschenswert gewesen, die intuitive Basis für (1.8.15) und (1.8.16) besser erläutert zu bekommen. (1.8.16) macht es ganz klar, daß die reduzierte Theorie T_1 bei manchen (nämlich den erklärenden) Reduktionen mindestens genauso stark sein soll wie die reduzierende Theorie T_2, eine Forderung, für die ich mir keine überzeugende Motivation denken kann.

Gibt es aber nicht doch irgendeine richtige und wichtige Idee, die hinter Einbettungen stecken könnte? Ich glaube schon. Man kann sich gut vorstellen, daß T_2 den Einfluß eines Parameters oder Faktors offenlegt, der von T_1 noch überhaupt nicht in Betracht gezogen worden war, und daß man T_1 aus T_2 dadurch erhält, daß man in T_2 ganz bestimmte Werte dieses Parameters fixiert. Dann kann jedem T_1-Modell ein T_2-Modell als Gegenstück zugeordnet werden, indem man den neuen Parameter in T_2 eben diese Werte „annehmen läßt"; T_2 kann daneben aber viele weitere Modelle besitzen, in denen der Parameter andere Werte annimmt.

Zur Illustration dieser Idee verlassen wir nun die Ebene abstrakter Argumentation und gehen kurz und an dieser Stelle nicht allzu tiefschürfend auf das Verhältnis zwischen den Keplerschen Gesetzen der Planetenbewegung („**KGP**") und der Newtonschen Theorie der Gravitation („**NTG**") ein.[53] Wir wollen dabei die für intertheoretische Relationen nötige Genauigkeit der Übereinstimmung von **NTG** und **KGP** bewußt niedrig ansetzen und jegliche noch irgendwie erträgliche Abweichung **NTG**s von **KGP** vernachlässigen. So soll insbesondere davon ausgegangen werden, daß **KGP** und **NTG** für unser Sonnensystem dasselbe besagen (obwohl ja bekanntlich die Newtonschen Anziehungskräfte der Planeten auf Sonne und auf andere Planeten merkliche Abweichungen von Keplers Ellipsenbahnen hervorrufen). Unter dieser Voraussetzung liegt hier ein geradezu paradigmatisches Beispiel von Reduzierbarkeit vor: Wenn überhaupt irgendeine Theorie der Wissenschaftsgeschichte auf eine andere reduzierbar ist, dann ist **KGP** auf **NTG** reduzierbar. Man wird hier zuversichtlich sowohl auf eine theoretische als auch auf eine erklärende Reduktion hoffen dürfen.

Eine erklärende Reduktion ist auch leicht nachweisbar. Jedes Planetensystem, das sich Keplersch verhält,[54] verhält sich tatsächlich auch (ungefähr) Newtonsch (als Planetenmassen der Newtonschen Erklärung darf

[53] Genaueres und für das Verständnis des folgenden hilfreiche Einzelheiten findet der Leser in Kapitel 8.1.

[54] Ich setze hier voraus, daß **KGP** eine Theorie über Planetensysteme im allgemeinen — und nicht nur über „unser" Plantensystem — ist, obgleich mir diese Interpretation letztlich nicht ganz korrekt erscheint (vgl. Kapitel 8.1).

man einfach die wirklichen Planetenmassen hernehmen[55]). **NTG** weist den Planeten einen neuen Parameter „Masse" zu, der in **KGP** noch gänzlich unbekannt war, und hat durch die neugewonnene Variationsmöglichkeit (eben dieses Parameters) in der Tat „mehr" Modelle als **KGP**. Umgekehrt gehorcht nämlich auch bei großzügigster Auslegung von „ungefähr" nicht jedes sich Newtonsch verhaltende Planetensystem den Keplerschen Gesetzen; man stelle sich etwa vor, in unserem Planetensystem hätten alle Planeten eine Masse von der Größenordnung der zehnfachen Jupitermasse. Dies heißt natürlich nicht, daß keine (approximative) theoretische Reduktion von **KGP** auf **NTG** möglich wäre, sondern nur, daß man zu diesem Zweck **NTG** mit Randbedingungen verstärken muß, die das Massenverhältnis zwischen Planeten und Sonne in geeigneter Weise beschränken. Außerdem braucht man mindestens noch Bedingungen, welche ausschließen, daß sich die Planeten auf ihren Bahnen allzu nahe kommen — geschweige denn kollidieren — oder daß sie zuviel kinetische Energie haben, um in beschränkter Entfernung von der Sonne zu verbleiben.[56]

Wenn das Finden von geeigneten Zusatzbedingungen für eine „angenäherte Ableitbarkeit" von **KGP** aus **NTG** auch kein unüberwindliches Problem zu sein scheint (vgl. Abschnitt 8.1),[57] so haben wir doch mit dem Aufweisen der theoretischen Reduktion sicherlich mehr Schwierigkeiten als mit dem Aufweisen der erklärenden Reduktion. Erlaubt dies den Schluß, daß hier die erklärende Reduktion à la von Pearce für das Vorliegen einer Reduktion wichtiger ist als die theoretische Reduktion? Keineswegs! Entscheidend ist vor allem die letztere. Eine erklärende Reduktion im Sinne einer Einbettung bleibt erhalten, wenn T_1 verstärkt wird, eine theoretische Reduktion im Sinne einer Interpretation bleibt erhalten, wenn T_1 abgeschwächt wird. Betrachten wir nun einerseits die Theorie **KGP**$^+$, die aus den drei heute als „die Keplerschen" bekannten Gesetzen und dem „Gesetz"

[55] Sind Orte und Beschleunigungen von Sonne und Planeten eines Newtonschen n-Körper-Systems zu einem gegebenen Zeitpunkt bekannt, so ist das Ausrechnen der n Massen von Sonne und Planeten aus den n vektoriellen Gleichungen

$\ddot{\mathbf{r}}_i = -\sum_{j \neq i} (m_j/|\mathbf{r}_j - \mathbf{r}_i|^3) \cdot (\mathbf{r}_j - \mathbf{r}_i)$, $i = 1, \ldots, n$,

im allgemeinen höchstens ein praktisches, aber kein theoretisches Problem.

[56] Bekanntlich sind nach **NTG** auch (angenähert) parabolische und (angenähert) hyperbolische Bahnen möglich. Allerdings könnten die entsprechenden Himmelskörper wohl kaum mehr als Planeten bezeichnet werden. Da ein Planet sich *qua Planet* nicht weiter und weiter von der Sonne entfernen darf, ist diese oft als Argument für die Notwendigkeit von Zusatzannahmen verwendete Beobachtung im Kontext von *Planeten*systemen nicht sonderlich wichtig.

[57] Für die Anfangsbedingungen der meisten Planetensysteme wird die ungefähre Übereinstimmung zwischen **KGP**- und **NTG**-Bewegungen zeitlich begrenzt sein.

aus Keplers Frühwerk, dem *Mysterium Cosmographicum* (1596), besteht, nach welchem sich die Proportionen der (mittleren) Entfernungen der seinerzeit bekannten sechs Planeten von der Sonne durch Ein- und Umschreiben der fünf „vollkommenen" oder „regulären" Körper erhalten (und erklären) lassen.[58] Und betrachten wir auf der anderen Seite die Theorie $\mathbf{KGP}^-$, welche nur die 1609 in der *Astronomia Nova* veröffentlichten ersten beiden Keplerschen Gesetze umfasse. *Formal* ist dann eine erklärende (aber keine theoretische) Reduktion von $\mathbf{KGP}^+$ auf **NTG** und eine theoretische (aber keine erklärende) Reduktion von $\mathbf{KGP}^-$ auf **NTG** möglich. *Intuitiv* hingegen ist die Angelegenheit ganz eindeutig: $\mathbf{KGP}^-$ wird man als auf **NTG** reduzierbar bezeichnen, keinesfalls jedoch $\mathbf{KGP}^+$. Deshalb glaube ich, daß nur die theoretische, nicht aber die erklärende Reduktion nach Pearce und Rantala das wiedergibt, was man gemeinhin mit dem Begriff „Reduktion" verbindet.

1.9 Typen von Reduktion II: Scheibe über (bereichseinschränkend-)deduktive und empirische Erklärung von Theorien

Erhard Scheibe hat zwei Arten von exakter intertheoretischer Erklärung oder Reduktion[59] unterschieden und analysiert, die den Pearceschen Begriffen von theoretischer und erklärender Reduktion ziemlich nahe kommen. Scheibe hat für seine Dichotomie mehrere Namen: auf der einen Seite kennt er den „Deduktionsbegriff (D-Begriff) der Erklärung" (1982, S. 303f), auch genannt die „potentielle Erklärung" (1983, S. 76f) oder „deduktive Erklärung" (1984, 82–85), auf der anderen Seite den „Bewährungsbegriff (B-Begriff) der Erklärung" (1982, S. 304), den er später als „faktische Erklärung" (1983, S. 77) oder als „empirische Erklärung" (1984, S. 90–92)

[58] Kepler hat sich übrigens von der Idee dieses „Gesetzes" zeitlebens nicht lossagen mögen. Interessanterweise wird auch heute noch in den meisten einschlägigen Handbüchern ein (viel einfacheres) Gesetz über die Proportionen der mittleren Planetenentfernungen von der Sonne angeführt, nämlich das Gesetz von Titius und Bode (aufgestellt 1766/72).

[59] Ein wesentlicher Unterschied zwischen (intertheoretischer) Erklärung und Reduktion ist hier weder von Scheibe intendiert noch von der Sache her notwendig. Scheibe (1982; 1983; 1984) betrachtet neben der exakten auch einen Begriff der „approximativen Erklärung". Kernpunkt seiner Ausführungen ist jedoch der exakte Fall, den wir i.f. ausschließlich untersuchen und uns dabei das Attribut „exakt" ersparen.

benennt. Wir wählen für das Folgende einheitlich die letzten Bezeichnungen, sprechen also von *deduktiver* und *empirischer* Erklärung.

Den Ursprung seiner Definitionen verfolgt Scheibe zurück auf den Nagelschen Reduktionsbegriff bzw. einen Verbesserungsvorschlag von Kemeny und Oppenheim.[60] Die Scheibeschen Verbalisierungen dieser Ausgangspunkte lauten wörtlich für die deduktive Erklärung:

(1.9.1) T_1 *folgt logisch* aus T_2 und eventuellen Verknüpfungsbedingungen, die in T_1, aber nicht in T_2 vorkommende Begriffe an die Begriffe von T_2 knüpfen;

und für die empirische Erklärung:

(1.9.2) Jeder durch T_1 *erklärbare* Teil der Gesamtheit aller Beobachtungsdaten ist auch durch T_2 erklärbar.

(Scheibe 1982, S. 298, 300; Bezeichnungen der Theorien von mir.) Scheibe versäumt auch nicht, seine Verpflichtung gegenüber Popper offenzulegen. Zum einen nämlich kann man bei Popper und ebenso bei Lakatos eine Bedingung finden, die von (1.9.2) praktisch nicht zu unterscheiden ist.[61] Zum anderen — und sehr viel wichtigeren — beruft sich Scheibe auf Popper, wenn er seine Präzisierung der Idee der Einschränkung des Geltungsbereiches von T_1 durch T_2 vorbringt. Wie wir in Abschnitt 1.3 gesehen haben, schließt eine *unbeschränkte* deduktive Erklärung von T_1 durch T_2 aus, daß T_2 Anomalien von T_1 erklärt und daß T_2 ihrer Vorgängerin T_1 widerspricht.[62] Dies bedeutet aber nach Popper — und nach den von Scheibe zitierten Physikern Boltzmann und Nernst — gerade, daß eine unbeschränkte deduktive Erklärung echt progressive Nachfolgertheorien, d.h. wirklichen wissenschaftlichen Fortschritt ausschließt. Die Idee der deduktiven Erklärung ist jedoch zu retten, indem man sie abschwächt zu der Forderung, daß eine deduktive intertheoretische Erklärung *nur innerhalb eines gewissen, von T_2 aus zu bestimmenden Geltungsbereichs für T_1* geleistet wird. Ich werde gleich darauf zu sprechen kommen, auf welche Weise

[60] Siehe hierzu Kapitel 2, Abschnitt 2.2. Wie Scheibe (1982, S. 300) betont, geht er nicht von der endgültigen Version des Reduktionsbegriffs von Kemeny und Oppenheim, sondern von einem ihrer provisorischen Vorschläge aus, den sie sofort wieder verwerfen.

[61] Vgl. Scheibes (1982, S. 300) Verbalisierung: „T_2 *erklärt* (oder reproduziert) die empirischen Erfolge von T_1." Sonderbarerweise kommentiert Scheibe diese Bedingung (mit der Hervorhebung von „erklärt") weder in ihrer Beziehung zum Kemeny-Oppenheimschen Vorschlag (mit der Hervorhebung von „erklärbar") noch greift er sie überhaupt wieder auf.

[62] Die Möglichkeit des Widerspruchs zwischen aufeinander folgenden Theorien wird von Scheibe immer wieder hervorgehoben. Siehe Scheibe (1982, S. 299, 305; 1983, S. 77; 1984, S. 78, 88).

hierdurch eine Verträglichkeit mit der Bedingung der Anomalienerklärung gewährleistet wird.

Gehen wir zunächst daran, eine Verbindung zwischen den beiden fundamentalen Begriffen Scheibes und der vorangehenden Diskussion herzustellen. Offensichtlich sind die oben genannten Zitate fast identisch mit (K2) und (K4) aus Abschnitt 1.2. Ein wichtiger — in der Tat der wichtigste — Unterschied in der Interpretation kommt jedoch durch die eben erwähnte Idee der Bereichseinschränkung zustande. Die Abgrenzung des Geltungsbereiches von T_1 im Lichte von T_2 wird offenbar in T_2 vorgenommen, was man modellieren kann durch die Auszeichnung einer Teilmenge A_2 von M°_{p2} deren Bilder bezüglich der Reduktionsfunktion **F** intuitiv in eben diesem Geltungsbereich $A_1 := \mathbf{F}[A_2]$ liegen.[63] Obgleich also aus philosophischen Gründen die Abgrenzung von A_1 durch T_2 (und in der Sprache von T_2) geleistet werden muß, erscheint doch die Annahme plausibel, daß mit einem Urbild eines x_1 aus A_1 auch alle anderen Urbilder dieses x_1's in A_2 enthalten sein müssen, d.h. daß gilt:

(Z'') $\quad \forall x_1 \in M_{p1}\ (\mathbf{F}^{-1}(x_1) \subseteq A_2 \text{ v } \mathbf{F}^{-1}(x_1) \subseteq \mathbf{C}A_2)$.[64]

Nun kann man mit Scheibe davon reden, daß T_1 durch T_2 nur dann[65] *deduktiv erklärt* wird, wenn

$(K2^{sch})$ $\quad \mathbf{F}[M_2 \cap A_2] \subseteq M_1 \cap A_1$.

Wir sehen, daß $(K2^{sch})$ eine Abschwächung von $(K2^{s})$ ist: Aus $(K2^{s})$ folgt $\mathbf{F}[M_2] \cap A_1 \subseteq M_1 \cap A_1$, woraus wegen $\mathbf{F}[M_2 \cap A_2] \subseteq \mathbf{F}[M_2] \cap \mathbf{F}[A_2]$ und der Definition von A_1 unmittelbar $(K2^{sch})$ folgt.[66] Andererseits ist klar,

[63] Wir arbeiten jetzt wieder mit einer Reduktionsfunktion **F**: $M^{\circ}_{p2} \rightarrow M_{p1}$. — Es ist klar, daß man an die Abgrenzung der Teilmenge A_2 gewisse pragmatische Forderungen stellen muß, etwa so, wie sie in Scheibe (1984, S. 82f) als Zulässigkeitsbeschränkungen angesprochen werden. Wenn sie auch sehr wichtig sind, werde ich diese außerlogischen Probleme im folgenden nicht behandeln. Vgl. auch die Bemerkungen zur Rolle von M°_{p2} in Abschnitt 1.3.

[64] *Formal* hätte man demnach auch von A_1 ausgehen und A_2 durch $A_2 := \mathbf{F}^{-1}[A_1]$ definieren dürfen.

[65] Scheibe sagt nirgends, daß er nur notwendige Bedingungen angibt, würde dies aber auf Nachfrage wohl sofort zugestehen und etwa zusätzliche Bedingungen für den Bewährungsgrad von T_1 und T_2 angeben wollen. Wir müssen uns wegen der in Abschnitt 1.1 angesprochenen intuitiven Forderungen an **F** ohnehin mit notwendigen Bedingungen zufrieden geben.

[66] Diese Überlegung legt die Frage nahe, ob man denn (K2) mit Bereichseinschränkung nicht gleich einfach durch $\mathbf{F}[M_2] \cap A_1 \subseteq M_1 \cap A_1$ darstellen sollte, was ein formales Analogon zu $(K1^{s''})$ wäre. Ich glaube, die Formalisierung ist hier nicht eindeutig festzulegen, und sehe jedenfalls keinen unmittelbaren intuitiven Einwand gegen diese Möglichkeit. Allerdings bräuchte man dann unten entweder die schwerlich zu rechtfertigende Zusatz-

daß ($K2^{sch}$) eine echte Abschwächung von ($K2^{s}$) ist. T_2 kann denn auch durchaus im Sinne von ($K7^{s}$) Anomalien von T_1 erklären, wenn man ($K2^{sch}$) statt ($K2^{s}$) zugrunde legt: Für eine außerhalb des Geltungsbereichs von T_1 liegende Anomalie $x_1 \in (I_1 \setminus M_1) \cap CA_1$ ist es natürlich überhaupt nicht ausgeschlossen, daß es ein x_2 in $I_2 \cap M_2 \cap CA_2$ mit $\mathbf{F}(x_2)=x_1$ gibt; für innerhalb von A_1 liegende x_1 ist dies wegen (Z″) nicht möglich.

Der Begriff der empirischen Erklärung muß nach Scheibe auf eine „Bewährungs-" oder „Erfahrungslage", das heißt auf die vorliegenden Beobachtungsdaten relativiert werden. Insofern müssen wir hier vorsichtig sein, wenn wir die Ausdrucksweise Scheibes (1982, S. 304; 1983, S. 77; 1984, S. 90), daß T_2 die „Erfolge" von T_1 als eigene „Erfolge" reproduzieren soll, direkt in die Sprache der Strukturalisten übersetzen wollen. In unserem einfachen Theorienmodell waren die Erfolge oder genauer: die erfolgreichen Anwendungen einer Theorie T in der Menge $I \cap M$ zusammengefaßt. Es ist aber durchaus nicht klar, daß alle *intendierten* Anwendungen auch schon zur einschlägigen Erfahrungslage gezählt werden können (der Umkehrung dagegen wird man wohl zustimmen dürfen). Um uns nicht noch eine neue Parametermenge aufzubürden, wollen wir aber einmal so tun, als ob man immer schon alle intendierten Anwendungen der alten Theorie T_1 — nicht unbedingt aber die von T_2 — geprüft hätte und wüßte, wie sie in Erfolge und Mißerfolge (Anomalien) aufzuteilen sind. Wir identifizieren also die intendierten Anwendungen mit den Beobachtungsdaten und erhalten sofort die folgende (Quasi-)Definition, die auch insofern passend erscheint, als sie genau die oben (in Abschnitt 1.6) bevorzugte Version ($K4^{s\prime}$) des Kriteriums (K4)[67] als Definiens verwendet: T_1 wird nur dann durch T_2 *empirisch erklärt*, wenn

($K4^{sch}$) $\quad \mathbf{F}[I_2 \cap M_2] \supseteq I_1 \cap M_1$.

Das Problem, welchem sich Scheibe nun zuwendet, ist dieses: Unter welchen Bedingungen, d.h. unter welchen Zusatzannahmen folgt eine empirische Erklärung aus einer deduktiven? Scheibe ist sich völlig darüber im Klaren, daß ($K2^{sch}$) allein jedenfalls nicht hinreicht, um ($K4^{sch}$) zu garantieren. Er schlägt als Ergänzung im wesentlichen zwei Bedingungen vor. Die erste und wichtigere Forderung ist die, daß (K2) dahingehend zu verstärken ist, daß T_1 — innerhalb des ihm eigenen Geltungsbereichs — nicht nur aus T_2 folgt, sondern hier sogar die stärkste (1982, S. 304) oder „maximale" (1983,

voraussetzung $\mathbf{F}[M_2] \cap \mathbf{F}[I_2] \subseteq \mathbf{F}[M_2 \cap I_2]$ oder statt der Zusatzvoraussetzung $A_2 \subseteq I_2$ das doch um einiges stärkere $A_2 \subseteq I_2 \cap M_2$.

[67] Oder äquivalent die Version ($K8^{s}$)(b) des Kriteriums (K8). Der Zusammenhang hiermit ist allerdings nicht sehr klar.

S. 79) logische Folgerung aus T_2 ist. Diese zusätzliche Bedingung[68] lautet formalisiert offenbar so:

(K2$^{sch\prime}$) $\mathbf{F}[M_2 \cap A_2] \supseteq M_1 \cap A_1$.

Unter der Voraussetzung von (Z″) ist diese Bedingung eine Abschwächung von (K2$^{s\prime}$). Denn sei $x_1 \in M_1 \cap A_1$. Nach (K2$^{s\prime}$) gibt es ein $x_2 \in M_2$ mit $\mathbf{F}(x_2)=x_1$. Weiter gibt es nach der Definition von A_1 auch ein $x_2' \in A_2$ mit $\mathbf{F}(x_2)=x_1$, und deshalb gilt — wenn man (Z″) zugrunde legt — $\mathbf{F}^{-1}(x_1) \subseteq A_2$, also insbesondere $x_2 \in A_2$. Damit haben wir aber ein $x_2 \in M_2 \cap A_2$ mit $\mathbf{F}(x_2)=x_1$ gefunden, und (K2$^{sch\prime}$) ist gezeigt.

Die zweite von Scheibe (1982, S. 305; 1983, S. 79) vorgeschlagene Forderung (Scheibe: „Anwendungsbedingung") besagt, grob gesprochen, daß alle Erfolge von T_1 in dem durch T_2 abgesteckten Geltungsbereich von T_1 liegen, d.h.

(1.9.3) $I_1 \cap M_1 \subseteq A_1$.

Nun stellt sich natürlich die Frage, ob man mit diesen Bedingungen — deren Korrektheit als strukturalistische Rekonstuktion der Scheibeschen Intentionen einmal unterstellt sei — die Behauptung[69] Scheibes verifizieren kann, daß sich aus einer durch (K2^{sch}) präzisierten deduktiven Erklärung unter Zuhilfenahme der beiden Bedingungen (K2$^{sch\prime}$) und (1.9.3) eine empirische Erklärung gemäß (K4^{sch}) herleiten läßt. Die Antwort lautet: Ja, allerdings nur dann, falls wir uns noch einer weiteren plausiblen Zusatzbedingung bedienen dürfen, nämlich der, daß alle in der T_2-Charakterisierung des Geltungsbereichs von T_1 enthaltenen potentiellen Modelle von T_2 auch intendierte Anwendungen von T_2 sind. In Zeichen:

(1.9.4) $A_2 \subseteq I_2$.

Die Behauptung (K4^{sch}) ist nun leicht aus den Voraussetzungen (K2^{sch}), (K2$^{sch\prime}$), (1.9.3) und (1.9.4) zu erhalten. Denn nach (1.9.3) gilt $I_1 \cap M_1 \subseteq M_1 \cap A_1$, welch letzteres nach (K2^{sch}) und (K$^{sch\prime}$) identisch ist mit $\mathbf{F}[M_2 \cap A_2]$, und dies ist wiederum wegen (1.9.4) eine Teilmenge von $\mathbf{F}[I_2 \cap M_2]$, und wir sind fertig. Doch man sieht gleich, daß die deduktive Reduktion gar nicht vonnöten war, um die empirische Reduktion zu gewährleisten, man wäre mit (K2$^{sch\prime}$), (1.9.3) und (1.9.4) alleine genausogut zum Ziel gekommen. Insofern muß, wenn die obigen strukturalistischen

[68] In Scheibe (1982) gehört dies zum Begriff der deduktiven Erklärung selbst, was aber außer durch die angepeilte Gewährleistung einer empirischen Erklärung überhaupt nicht zu motivieren ist. In Scheibe (1983) ist dieser Schönheitsfehler behoben, und die Forderung wird als zusätzliche „crucial assumption" aufgeführt.

[69] Scheibe erwähnt für das folgende „Theoreme" und „Resultate", skizziert aber leider keine Beweisidee.

Formalisierungen nicht verfehlt sind, Scheibes Vorhaben, eine empirische Reduktion unter gewissen Voraussetzungen *aus einer deduktiven Reduktion* zu gewinnen, wohl als gescheitert betrachtet werden.[70]

Die Kritikpunkte sind also klar: Einerseits wird der Ausgangspunkt ($K2^{sch}$) gar nicht gebraucht. Andererseits gibt es für ($K2^{sch\prime}$) gar keine andere Motivation als eben die Möglichkeit einer Ableitung von ($K4^{sch}$). Damit muß man schon die Frage stellen, warum denn eigentlich T_2 nicht auch im Geltungsbereich von T_1 (und gerade dort) *echt* stärker als T_2 sein darf. Ich sehe keine Motivation für ($K2^{sch\prime}$). Man kann sich durchaus mit der — unter der Voraussetzung von (1.9.3) — schwächeren Bedingung

(1.9.5) $\mathbf{F}[M_2 \cap A_2] \supseteq I_1 \cap M_1$

begnügen, aus welcher — unter der Voraussetzung von (1.9.4) — ($K4^{sch\prime}$) auch schon ableitbar ist. Mit (Z″) ist diese Bedingung äquivalent zu $\mathbf{F}[M_2] \cap A_1 \supseteq I_1 \cap M_1$, was sich mit (1.9.3) wiederum zu

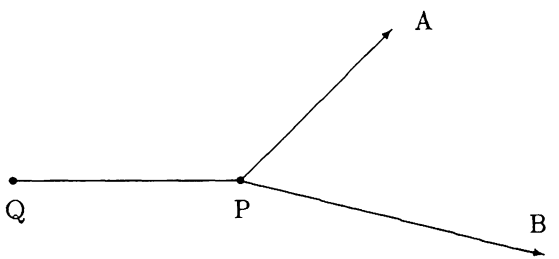
(1.9.6) $\mathbf{F}[M_2] \supseteq I_1 \cap M_1$

vereinfacht. Meiner Ansicht nach ist das die einzige Forderung an die Menge der T_2-Modelle, die man mit dem Hinweis auf empirische Betrachtungen vernünftigerweise rechtfertigen kann.[71] Warum soll jedes T_1-Modell ein T_2-Modell als Gegenstück haben? Es können, so scheint mir, doch nur die erfolgreichen T_1-Modelle darauf Anspruch erheben.

Vielleicht hat man vor lauter Jongliererei mit ähnlich aussehenden Formeln aus den Augen verloren, worin der Unterschied zwischen ($K2^{sch\prime}$) und (1.9.5), worin der Unterschied der Interpretation von I_1 und A_1 überhaupt liegt. Ein einfaches Beispiel soll dies noch einmal verdeutlichen. Man betrachte Teilchen, die von einer Quelle Q bis zum Punkt P eine Trägheitsbewegung ausführen, in P aber aus der Bahn geworfen werden:

[70] Ein Teilvorhaben von Scheibe (1982) wurde schon in Scheibe (1983) nicht mehr weiter verfolgt. Im früheren Aufsatz hatte Scheibe (1982, S. 304) nämlich eine zweite Bedingung für empirische Erklärung angeführt, die wir wohl so wiedergeben dürfen:
($K4^{sch\prime}$) $\mathbf{F}[(I_2 \setminus M_2) \cap A_2] \subseteq I_1 \setminus M_1$.
Scheibe (1982, Fußnote 17) nannte dies eine „automatische Ergänzung" von ($K4^{sch}$), ich kann aber keine Möglichkeit finden, ($K4^{sch\prime}$) aus einer ähnlichen Prämissenmenge wie ($K4^{sch}$) zu bekommen. Vielleicht hat Scheibe aus diesem Grund ($K4^{sch\prime}$) im späteren Aufsatz nicht mehr erwähnt.

[71] Und dies ist natürlich immer noch recht viel verlangt. Wenn sich alle Wissenschaftler einig sind, daß T_2 in vielen Bereichen und in vielerlei Hinsicht der Theorie T_1 eindeutig überlegen ist, wird man den Verlust einiger weniger erfolgreicher Anwendungen von T_1 leicht verschmerzen, in der Hoffnung, daß sie sich auf lange Sicht doch noch irgendwie in die neue Theorie einpassen lassen.



SKIZZE 1.2

Die Theorie T_1 sage nun aus, daß nach dem Punkt P, wo zum Beispiel eine bestimmte Interaktion mit anderen Teilchen stattfinde, die Bahn A oder die Bahn B durchlaufen werden kann. T_1 hat dementsprechende Modelle x_A und x_B. Versuche sollen nun ergeben, daß de facto alle Teilchen den Weg A einschlagen, d.h. daß nur x_A in I_1 liegt. Nichtsdestotrotz gehören Teilchen mit Bahnen zwischen Q und P natürlich zu dem Bereich, in dem die Theorie T_1 gültig ist, also müssen wohl x_A *und* x_B, die beide die Bahnen solcher Teilchen darstellen, in A_1 sein. Eine bessere Theorie T_2 sollte zwar x_A als Modell zulassen, im Einklang mit (1.9.5), sollte aber x_B als Modell ausschließen, im Widerspruch zu ($K2^{sch\prime}$).

Scheibe selbst scheint sein Projekt später ebenfalls sehr kritisch beurteilt zu haben. In Scheibe (1984, S. 91f) leistet er eine Art Offenbarungseid sowohl hinsichtlich seiner Idee, eine empirische Erklärung aus einer deduktiven Erklärung herzuleiten:

> Almost nothing is known at the general level about the relations [between deductive and empirical explanations]

als auch hinsichtlich seines Begriffs der empirischen Erklärung überhaupt:

> This problem of intertheoretic relations [of empirical theory explanation] can hardly be tackled at present, since the situation has not even been clarified for single cases of sufficient interest.

Gerade nach einer solch selbstkritischen Schlußbilanz sollte man sich vor Augen führen, worum es Scheibe eigentlich ging. Die angestrebte Gewinnung einer empirischen aus einer deduktiven Erklärung sollte nicht zur Rechtfertigung der empirischen, sondern zur Rechtfertigung der deduktiven Erklärung dienen! Scheibe betont immer wieder, daß nur die erstere eine im Sinne des Fortschrittsgedankens wirklich wichtige intertheoretische

Relation sei: Eine deduktive Erklärung „interessiert ... doch nur dann" (1982, S. 304), wenn aus ihr eine empirische folgt, ohne diese gäbe er für jene „überhaupt keine physikalische Rechtfertigung" (1983, S. 79), und schließlich, „warum sollten wir uns wegen zusätzlicher deduktiver Beziehungen Umstände machen" (1984, S. 90), wenn wir eine empirische Erklärung schon haben?

Ohne daß auf mögliche Erwiderungen eingegangen werden muß, erscheint Scheibes Typologie in diesem Licht zumindest nicht mehr von eigenständigem Interesse. Im Grunde genommen gibt es für ihn nur eine relevante intertheoretische Relation, und das ist die empirische Erklärung. Die andere, die deduktive Erklärung ist *nur* insofern interessant, als sie sozusagen Helfershelferdieste für die eine, die empirische Erklärung leistet. Von hier aus ist es nur ein Katzensprung zur der monistischen Position, die einen Pluralismus von logisch verschiedenen intertheoretischen Erklärungs- oder Reduktionsbegriffen überhaupt ablehnt.

Mir ist bis jetzt kein überzeugendes Argument begegnet, welches schlüssig zeigt, daß die in diesem Kapitel diskutierten Problemstellungen mehrere Typen von Reduktion erzwängen. Einen Schluß von der in den Abschnitten 1.2–1.6 dargestellten Lage der Reduktionsdiskussion bei den Strukturalisten auf eine logische „Mehr-Typen-Lösung" jedenfalls hielte ich für völlig verfehlt. Im Unterschied zu Scheibe glaube ich jedoch, daß die deduktive und nicht die empirische Erklärung der richtige Kandidat für die Explikation von fortschittlichem Theorienwandel ist.[72] Eine (bereichseinschränkend-)deduktive Erklärung gestattet in vielen Fällen eine — informativere! — Erklärung der erfolgreichen Anwendungen der alten Theorie. Alles, was aus T_1 folgt, soll auch aus T_2 folgen, wenn möglich sogar mehr. Im obigen Beispiel war T_1 eine erfolgreiche Theorie, weil sie voraussagte, daß die Teilchen den Weg A oder den Weg B durchlaufen würden, Theorie T_2 war aber besser, weil sie die B-Wege ausschloß. Wie schon im letzten Abschnitt gesagt, muß T_2 nicht alle *möglichen* Erklärungen von T_1 (und erst recht nicht alle *gescheiterten* Erklärungen von T_1) nachvollziehen können. Deshalb ist es *nicht* schlimm, wenn einige T_1-Modelle nicht durch $\mathbf{F}[M_2 \cap A_2]$ abgedeckt werden. Allerdings ist es wünschenswert, daß T_2 alle geglückten Erklärungen von T_1, d.h. alle erfolgreichen Anwendungen von T_1 widerspiegeln kann. Dies ist eine über die deduktive Relation hinausgehende, separate Forderung, die wir mit (1.9.5) schon formuliert haben. Zusammengefaßt ist meine These, daß es nur *einen* logischen Typ

[72] Genauer: daß ($K2^s$), ($K2^{sch}$) oder Interpretationen die richtigen Ansatzpunkte der Explikation sind, nicht ($K2^{s\prime}$), ($K2^{sch\prime}$) bzw. Einbettungen.

von Reduktion gibt, der eine wesentliche Ähnlichkeit mit der Reduktion nach Adams und Sneed, der Pearceschen theoretischen Reduktion und der Scheibeschen deduktiven Erklärung hat. Als empirische Zusatzbedingung sollte wohl (1.9.5) hinzugenommen werden. Also schlage ich vor, für bereichseinschränkende Reduktionen das folgende Kriterium als notwendige Bedingung anzusetzen:

(1.9.7) $I_1 \cap M_1 \subseteq \mathbf{F}[M_2 \cap A_2] \subseteq M_1 \cap A_1$

Im Gegensatz zu Scheibe glaube ich nicht, daß $\mathbf{F}[M_2 \cap A_2]$ gleich $M_1 \cap A_1$ sein sollte, sondern halte T_2 für desto besser, je kleiner $\mathbf{F}[M_2 \cap A_2]$ ist, d.h. je näher es an $I_1 \cap M_1$ herankommt.[73] Denn dann ist T_2 auch (und gerade) im Geltungsbereich von T_1 logisch stärker und hat eine größere Aussagekraft, ohne daß die Erfolge von T_1 verloren gehen. Im Vorgriff auf später zu Untersuchendes ist noch anzumerken, daß es in vielen Fällen gar nicht richtig ist, daß T_2 die erfolgreichen Anwendungen von T_1 erklärt. Vielmehr zeigt T_2 häufig, daß auch die angeblich erfolgreichen Anwendungen von T_1 *genaugenommen* Anomalien von T_1 sind und nur aufgrund von Meßungenauigkeiten, „Korrekturen" an Meßergebnissen, Idealisierungen oder ähnlichem als Erfolge durchgehen konnten.[74] Eine empirische Erklärung ist dann — jedenfalls ohne Approximationen[75] — nicht mehr gegeben. Dies ist typisch für echt progressiven Theorienwandel. Oft kann man sagen, daß T_2 eigentlich nicht die Theorie T_1, sondern ihr Scheitern erklärt. Inwiefern man in solchen Fällen von einer kontrafaktischen, immer noch deduktiven Erklärung von T_1 durch T_2 sprechen kann und wie dies zu analysieren ist, werde ich in Kapitel 7 darzutun versuchen.

1.10 Ein Miniaturbeispiel: Bereichseinschränkungen und kontrafaktische Konditionalsätze

Scheibe präsentiert seine Vorschläge nicht strukturalistisch, sondern im Rahmen des Statement view. Wie kann man hier die Idee einer Bereichsein-

[73] Eine wichtige, im strukturalistischen Konzept schwerlich ausdrückbare Forderung wäre, daß T_2 gesetzesartige Sätze enthält, also doch echt über $I_1 \cap M_1$ (worin nur reale Systeme und evtl. „ähnliche" potentielle Modelle enthalten sind) hinausgehen muß. Vgl. hierzu Balzer, Moulines und Sneed (1987, S. 15–20).

[74] Vgl. hierzu insbesondere auch die von Scheibe (1982, S. 294f; 1984, S. 75f) zitierten, sehr treffenden Bemerkungen Walther Nernsts (1922).

[75] Auf Approximationen werde ich ausführlich in den Kapiteln 2, 8 und 9 eingehen

schränkung (zur Ermöglichung von Anomalienerklärungen) formalisieren? Scheibes Antwort auf diese Frage kann durch das folgende Schema versinnbildlicht werden:

(1.10.1) $T_2 \vdash T_1|_A$.

Hier sei A eine offene Formel mit einer ungebundenen Variablen, und $T_1|_A$ stehe für die *Relativierung* der Theorie T_1 *auf eine Substruktur* des eigentlich „intendierten Universums", die als Objektbereich die Klasse aller x hat, welche A erfüllen.[76] Ich setze nun voraus, daß eine Verknüpfung der Begriffe von T_1 und T_2 möglich ist, verzichte jedoch auf eine explizite Nennung der hierzu normalerweise nötigen Brückenprinzipien. Ich werde die Idee von (1.10.1) an einem zwar extrem einfachen und insofern wissenschaftsfremden, aber trotzdem lehrreichen Beispiel Scheibes (1983, S. 76f) illustrieren. Dem stelle ich anschließend eine alternative Sichtweise des Beispiels gegenüber, die einen ersten Eindruck von später genauer zu entwickelnden, für dieses Buch zentralen Gedanken vermitteln soll.

Das Miniaturbeispiel Scheibes hat die sich am 17. April 1982 in der kanadischen Stadt Waterloo aufhaltenden Menschen zum Gegenstand. In Waterloo wurde zu dieser Zeit eine wissenschaftliche Tagung abgehalten. Wir betrachten das Theorienpaar T_1 und T_2 mit den folgenden „Gesetzen" über die Körpergröße dieser Menschen:

T_1: Alle Menschen in Waterloo sind mehr als einen Meter groß.

T_2: Alle Teilnehmer an der Tagung in Waterloo sind mehr als einen Meter groß. u n d
Alle Kinder unter zwei Jahren in Waterloo sind höchstens einen Meter groß.

oder formalisiert in der Sprache der Quantorenlogik erster Stufe:

T_1: T_1-1 $\forall x(Wx \rightarrow Gx)$
(die Sprache von T_1 besteht aus den nichtlogischen Prädikaten
W: sich in *W*aterloo befindender Mensch und
G: mehr als einen Meter *g*roß) ;

T_2: T_2-1a $\forall x(Wx \wedge Tx \rightarrow Gx)$,
T_2-1b $\forall x(Wx \wedge Kx \rightarrow \neg Gx)$

[76] Zu den Begriffen der Relativierung und der Substruktur vergleiche zum Beispiel Ebbinghaus, Flum und Thomas (1978, S. 160–163). — Eberle (1971, S. 496–498) verwendet ebenfalls Relativierungen und das Schema (1.10.1), aber zum Zwecke des Nachweises einer Reduzierbarkeit von „inakkuraten" auf akkuratere Theorien (zu Scheibe vgl. hingegen Fußnote 59). Eberle ist der Ansicht, daß A im allgemeinen, zumindest aber in dem von ihm besprochenen Galilei-Newton-Beispiel „ideale Bedingungen" angibt, die — von T_2 aus gesehen — kontrafaktisch sind.

(die Sprache von T_2 besteht aus den nichtlogischen Prädikaten
W, G,
T: Teilnehmer an der *T*agung und
K: *K*ind unter zwei Jahren.

In einem gewissen, sehr weit gefaßten Sinn wird man sagen können, daß T_1 auf T_2 reduzierbar ist, oder auch, daß T_1 durch T_2 erklärt wird (besonders wenn es in Waterloo viele Tagungsteilnehmer und wenig kleine Kinder gibt). Andererseits hat man ziemlich deutlich das Gefühl, daß die erste Theorie eigentlich mit der zweiten inkonsistent ist. [77]

Tatsächlich folgt aus der Vereinigung von T_1 und T_2 aber nur, daß es in Waterloo keine kleinen Kinder gibt, in Zeichen: $\forall x(Wx \rightarrow \neg Kx)$. Um nun dem intuitiven Gefühl der Inkonsistenz zu seinem Recht zu verhelfen, beschließen wir, die unwillkürlich gemachte Annahme, daß es in Waterloo doch kleine Kinder gibt, explizit in T_2 mit aufzunehmen:

T_2-2 $\quad \exists x(Wx \wedge Kx)$.

Man beachte, daß wir hiermit so etwas wie eine kontingente Randbedingung in unsere Theorie mit aufgenommen haben, das heißt, wir erlauben ausdrücklich, daß Theorien nicht nur aus in irgendeinem Sinne gesetzesartigen Aussagen bestehen sollen. Es ist sehr leicht zu sehen, daß T_1 und T_2 jetzt einerseits wirklich widersprüchlich sind, daß sich aus T_2 andererseits aber

$T_1|_T$: $\quad \forall x(Tx \rightarrow (Wx \rightarrow Gx))$

ableiten läßt ($T_1|_T$ ist natürlich äquivalent mit T_2-1a) oder, anders gesagt, daß T_1 durch T_2 im Sinne Scheibes bereichseinschränkend deduktiv erklärt wird.

Scheibe (1983, S. 79f) geht weiter und behauptet, das vorliegende Beispiel erfülle auch die oben durch ($K2^{sch\prime}$) formalisierte zusätzliche Forderung, die beiden Theorien seien also auf dem durch T abgesteckten Bereich „äquivalent". Wir wollen dies nachprüfen und sehen, ob sich $T_2|_T$ aus T_1 herleiten läßt. Eine erste Schwierigkeit besteht darin, daß unser T_2 mit T_2-2 jetzt einen Existenzsatz enthält, dessen T-Relativierung $\exists x(Tx \wedge (Wx \wedge Kx))$ aller Voraussicht nach falsch sein wird, weil an wissenschaftlichen Tagungen

[77] Ein ganz ähnliches Gefühl stellt sich bei den beiden in Kapitel 8 unten bearbeiteten wissenschaftlichen Vergleichspaaren ein (Keplers Gesetze der Planetenbewegung und Newtons Gravitationstheorie; ideales und van der Waalssches Gasgesetz), obgleich die jeweiligen Gleichungen in beiden Fällen gemeinsame Lösungen besitzen, d.h. obgleich die Vergleichspaare — zumindest nach *einer* Interpretation — nachweislich miteinander verträglich sind.

üblicherweise keine kleinen Kinder teilnehmen dürfen.[78] Deshalb sollten wir weder erwarten noch fordern, daß dieser Satz aus T_1 herleitbar ist, und uns auf die Allsätze in T_2 beschränken.[79] T_2-1a und T_2-1b lauten in ihrer auf die Menge der Tagungsteilnehmer relativierten Form

$T_2|_T$: $\forall x(Tx \rightarrow (Wx \wedge Tx \rightarrow Gx))$,
$\forall x(Tx \rightarrow (Wx \wedge Kx \rightarrow \neg Gx))$.

Die relativierte Form von T_2-1a ist sofort aus T_1 zu gewinnen, denn wenn alle Menschen in Waterloo groß sind, dann auch insbesondere die an der Tagung teilnehmenden. Was ist aber zur relativierten Form von T_2-1b zu sagen? Sie kann offenbar nur dann aus T_1 folgen, ja, sie ist sogar nur dann mit T_1 verträglich, wenn in T_1 ausgesagt wird, daß an der Tagung in Waterloo keine kleinen Kinder teilnehmen. Also müßte man, um die relativierte Form von T_2-1b zu erhalten, T_1 noch entsprechend erweitern. Da aber ein solches Gesetz nur in der Sprache von T_2 formuliert werden kann und auch einen nur für T_2 relevanten Sachverhalt ausdrückt (nach T_1 sind ja ohnehin alle Menschen in Waterloo groß), schlage ich vor, die Disjunktheit von Tagenden und Kleinkindern nicht in T_1, sondern in T_2 mit aufzunehmen:

T_2-3 $\forall x(Wx \wedge Kx \rightarrow \neg Tx)$.

T_2 wird im folgenden als aus T_2-1, T_2-2 und T_2-3 bestehend gedacht. Damit ist der Scheibeschen Intention, $T_2|_T$ aus T_1 herzuleiten, natürlich nicht gedient, und T_2 ist insbesondere innerhalb der Menge der Tagungsteilnehmer stärker als T_1. Doch erstens wird so der intuitive Gehalt von T_1 und T_2 recht gut getroffen, und zweitens halte ich — wie im letzten Abschnitt ausgeführt — das Nichterfülltsein der Bedingung ($K2^{sch\prime}$) für gar nicht so schlimm.

Kritisch ist zu diesen Vorschlägen aber anzumerken, daß sowohl $T_1|_T$ als auch $T_2|_T$ nur mehr sehr schwache Abbilder der ursprünglichen Theorien T_1 und T_2 sind und daß in ihnen vor allem wesentliche Teile der zugrundeliegenden „Gesetze“ T_1-1 und T_2-1 verloren gehen. Natürlich

[78] Ähnlich gibt es Existenzsätze, die aus der Newtonschen Gravitationstheorie (z.B. „Es gibt Himmelskörper mit angenähert hyperbolische Bahnen im Sonnensystem.“) oder aus der van der Waalsschen Zustandsgleichung (z.B. „Es gibt Systeme von Gasen, wo bei gleichbleibenden Temperaturen eine Volumenkompressionen keine Druckerhöhung nach sich zieht.“) folgen und falsch werden, wenn man sie auf Planeten in unserem Sonnensystem bzw. auf ideale Gase relativiert.

[79] Vgl. Scheibe (1982, S. 305): „Die Idee, daß eine physikalische Theorie durch *Erweiterung* ihres Anwendungsbereichs problematisiert und schließlich verbessert werden kann, impliziert, daß man sich nur für ihre Folgerungen von *universaler* Form interessiert.“

muß zugestanden werden, daß man ohne jeden Eingriff in die ursprünglichen Theorien *keine* Erklärung von T_1 durch T_2 zuwege bringen kann. Es bleibt aber die Frage: Gibt es nicht aber doch eine Rekonstruktion, die solche Eingriffe in die Identität der Theorien etwas behutsamer vornimmt? Ich glaube, es gibt sie, und möchte nun andeutungsweise — und im Vorgriff auf später Nachzuliefernde — meine Sichtweise des Problems darstellen. Wie Scheibe möchte auch ich mich nicht allzusehr von der Grundidee des deduktiv-nomologischen Erklärungsbegriffs entfernen. Anstelle der im Schema (1.10.1) angedeuteten Relativierung auf eine Substruktur will ich das Schema

(1.10.2) $T_2{}^*{}_A \vdash T_1$

der *hypothetischen Revision* $T_2{}^*{}_A$ von T_2 *durch eine kontrafaktische Annahme* A zur Diskussion stellen. Eine solche Revision soll „minimal" sein und möglichst wenig von der in T_2 enthaltenen Information preisgeben. Im vorliegenden konkreten Beispiel, so meine ich, müßte man, um T_1 zu erhalten, „in T_2" kontrafaktisch annehmen, daß alle Menschen in Waterloo an der Tagung teilnehmen, d.h. daß

$$A: \quad \forall x(Wx \rightarrow Tx) .$$

Es geht hier nicht darum, daß A eine vernünftige Annahme sein muß — T_1 ist von T_2 aus gesehen ja oft auch nicht besonders vernünftig. Sehr wohl berücksichtigt werden muß allerdings, daß A der Theorie T_2 widerspricht, denn $\neg A$ ist aus T_2-2 und T_2-3 ableitbar.[80] Ein einfaches *Dazunehmen* von A zu T_2 würde also (1.10.2) trivialisieren. Da das zu vermeiden ist, müssen wir T_2 wirklich *revidieren*, das heißt insbesondere auch Teile von T_2 *streichen*, um A konsistenterweise in die Theorie T_2 einpassen zu können. In diesem — und nur in diesem — Sinne kann man A nicht nur kontrafaktisch, sondern auch kontratheoretisch nennen. Es wird sich herausstellen, daß die Revision von T_2 weniger gravierend in T_2 eingreift als die Relati-

[80] $\neg A$ wäre natürlich auch aus T_1 ableitbar, wenn T_2-2 und T_2-3 Bestandteile von T_1 wären. Das im folgenden präsentierte Modell funktioniert aber nicht in der Form von (1.10.2), wenn $\neg A$ in T_1 enthalten ist. Deshalb ist es wichtig, Argumente zu finden, daß zumindest *entweder* T_2-2 *oder* T_2-3 nicht in T_1 sein sollte. Eine Möglichkeit bestünde, wie im Text angedeutet, in dem Verweis darauf, daß diese Bedingungen nicht in der Sprache von T_1, sondern in der von T_2 formuliert sind und schon allein deshalb nicht zu T_1 gehören können. Ich muß allerdings gestehen, daß ich bezüglich der Überzeugungskraft und der Reichweite dieses argumentativen Schachzugs skeptisch bin. In ernstzunehmenden Beispielsfällen, wo keine Rechtfertigungsmöglichkeit für $\neg A \notin T_1$ besteht, müßte (1.10.2) modifiziert werden, vermutlich zu $T_2{}^*{}_A \vdash T_1{}^*{}_A$ oder zu $T_2{}^*{}_A \vdash G(T_1)$, wobei $G(T_1)$ für die Menge der „Gesetze" von T_1 (in irgendeinem hier nicht problematisierten Sinn) stehen soll und als mit A verträglich vorausgesetzt ist.

vierung von T_2 auf die Substruktur der Tagenden. Im besonderen läßt sie den „Gesetzesteil" von T_2 unbehelligt.

Das Problem, das sich nun stellt ist natürlich: *Wie* muß man T_2 revidieren, d.h. *was* muß man aus T_2 streichen, um A ohne Widersprüche in T_2 einbauen zu können? Offenbar gibt es zwei naheliegende Möglichkeiten. Entweder man verzichtet in T_2 auf T_2-2 oder auf T_2-3. Ich möchte beide Möglichkeiten und die sich jeweils ergebenden Konsequenzen diskutieren.[81]

Der erste Fall sei der, in dem unter der Annahme von A T_2-2 aus T_2 gestrichen wird. Wenn alle Menschen in Waterloo an der Tagung teilnehmen würden, so könnte das Räsonnement zum Beispiel gehen, dann müßte Waterloo wohl ein künstlich aus dem Boden gestampftes Forschungszentrum sein, in der es keine Nichtwissenschaftler und erst recht keine kleinen Kinder gibt. Die Satzmenge ${T_2}^*{}_A$ ist in diesem Fall die Menge der Folgerungen der „Axiome" T_2-1, T_2-3 und A, und man sieht, daß in der Tat das einzige Gesetz von T_1, nämlich T_1-1, zu dieser Menge gehört (es folgt unmittelbar aus T_2-1 und A). Das heißt nichts anderes, als daß das Schema (1.10.2) erfüllt ist.

Es ist aber auch der zweite Fall denkbar, in welchem T_2-3 der hypothetischen Annahme von A geopfert wird. Wenn nämlich alle Menschen in Waterloo an der Tagung teilnehmen würden, dann, so könnte man sich vorstellen, wäre Waterloo wohl eine typische kanadische Kleinstadt, wäre die Tagung ein großer Kongreß über die Sozialstruktur typischer kanadischer Kleinstädte o.ä., und dann würden die „Objekte" der durchgeführten Studien ausnahmsweise einmal auch als Teilnehmer der Tagung geführt. Unter diesen Umständen gibt es selbstverständlich kleine Kinder in Waterloo, und die sind jetzt eben Teilnehmer an der wissenschaftlichen Tagung. Es ist allerdings nicht möglich, als Satzmenge ${T_2}^*{}_A$ einfach die Folgerungsmenge der „Axiome" T_2-1, T_2-2 und A herzunehmen, denn diese Axiome wären immer noch inkonsistent. Wir könnten uns helfen, indem wir statt von der alten von einer neuen, äquivalenten Axiomatisierung T_2's ausgehen, welche aus T_2-1a, T_2-2, T_2-3 und

T_2-1b′ $\quad \forall x(Wx \wedge Kx \wedge \neg Tx \rightarrow \neg Gx)$

[81] Natürlich sind noch mehr Revisionsmöglichkeiten denkbar. Man kann sich zum Beispiel vornehmen, auf die Konjunktion von T_2-2 und T_2-3 zu verzichten. Dies wird in der Regel (wenn T_2-2 und T_2-3 nicht ganz genau gleich wichtig sind; vgl. Kapitel 3, Bedingung (T-7∧8)) allerdings auf dasselbe hinauslaufen wie der Verzicht auf eines von beiden, denn es bedeutet normalerweise eine unnötige Preisgabe von in T_2 enthaltener Information, wenn man sowohl auf T_2-2 als auch auf T_2-3 verzichtet.

besteht. T_2-1b′ ist für sich genommen schwächer als T_2-1b, vor dem Hintergrund von T_2-3 jedoch gleichwertig. Die „Axiomenmenge“ T_2-1a, T_2-1b′, T_2-2 und A ist als Basis für die revidierte Theorie ${T_2}^*{}_A$ geeignet,[82] und wieder folgt T_1-1 aus ${T_2}^*{}_A$.

Unschönerweise ist dieses Resultat aber gar nicht einleuchtend, denn nach den zuletzt angedeuteten Überlegungen bei der hypothetischen Annahme von A würde T_1 intuitiv eben *nicht* aus T_2 folgen — kleine Kinder bleiben auch dann klein, wenn sie an wissenschaftlichen Tagungen teilnehmen! Auch formal ist das Verfahren der Ersetzung von T_2-1b durch T_2-1b′ angreifbar: Schon aus Symmetriegründen sollte man dann doch auch T_2-1a durch

$$T_2\text{-}1a' \qquad \forall x(Wx \wedge Tx \wedge \neg Kx \rightarrow Gx)$$

ersetzen, womit man die Ableitbarkeit von T_1 aus ${T_2}^*{}_A$ wieder verloren hat. Insofern können wir also unser intuives Urteil und die formale Analyse zwar einigermaßen zusammenbringen. Aber das Schema (1.10.2) ist in große Gefahr geraten, da nicht alle möglichen Revisionen von T_2 durch die hypothetische Annahme von A — und es sind in der Regel offensichtlich mehrere solche Revisionen möglich — die Ableitung von T_1 gestatten.

Was ist nun zu zun? Soll man resignieren und schließen, Revisionen durch kontrafaktische Bedingungen seien eben von unwiderruflicher Beliebigkeit und könnten deshalb keinen Platz in der Wissenschaft oder in der Wissenschaftstheorie beanspruchen? Diese Position hat viele Vertreter, zu deren prominentesten Quine und van Fraassen gehören. Ich meine aber, ein solcher Schluß wäre an dieser Stelle vorschnell. Wir müssen nur zeigen, daß nicht alle Revisionsmöglichkeiten gleich gut sind, in unserem Beispiel speziell, daß die erste Überlegung besser ist als die zweite. Es sind also Argumente dafür vorzubringen, daß der erste oben skizzierte Weg hypothetischen Räsonierens angemessener ist als der zweite, anders gesagt, daß es richtiger ist, T_2-2 aus T_2 herauszunehmen als T_2-3 herauszunehmen, oder, wieder anders ausgedrückt, daß T_2-3 für (die Identität von) T_2 wichtiger ist als T_2-2. Gibt es in diesem Beispiel aber Argumente dieser Art? Ich

[82] Wie kann es sein, daß die Art der Axiomatisierung einer Theorie eine Rolle für die Revisionsmöglichkeiten spielt? Die Antwort ist nicht leicht zu geben, denn im Gärdenforsschen Modell der Revision, welches ich später vorstellen und verwenden werde, spielt die Art der Axiomatisierung jedenfalls vorderhand keine Rolle, da sowieso immer nur deduktive Abschlüsse entscheidend sind. In Beispielsdiskussionen wie hier jedoch hat man immer nur endlich viele, meistens sogar recht wenige Sätze vor sich. Ein *in dieser Hinsicht* besseres, weil realitätsnäheres Modell der Theorienrevision wurde von Angelika Kratzer (1981a; 1981b) entwickelt und in Rott (1984) ausführlich besprochen und mit dem Gärdenforsschen Modell verglichen.

glaube: Nein. Warum sollte die Disjunktheit der Menge der Tagungsteilnehmer und der Menge der kleinen Kinder in Waterloo von T_2 aus gesehen wichtiger sein als die Existenz der kleinen Kinder? Ich kann mir nicht vorstellen, wie man das plausibel machen sollte.

Ist mein Vorschlag also letztlich doch gescheitert? Wieder möchte ich für „Nein" plädieren und die Schuld am Fehlen überzeugender Argumente der gesuchten Art der Theorie T_2 selbst zuschreiben. T_2 ist, im Gegensatz zu ihrer „Vorgängertheorie" T_1, ganz einfach keine wirkliche Theorie über Körpergrößen (der Menschen in Waterloo), da zumindest eines ihrer „Gesetze", nämlich T_2-1a, gar keinen Gesetzescharakter hat.[83] Jede nichttriviale Aussage über die Teilnehmer einer wissenschaftlichen Tagung muß, so scheint es, immer eine kontingente Aussage bleiben, und zwar aus Gründen, die vermutlich mit der Tatsache zusammenhängen, daß T kein projektierbares Prädikat (im Sinne von Goodman) ist oder keine natürliche Art definiert.[84] Wenn wir anstelle von T_2 die Theorie T_3 betrachten, welche aus T_2 dadurch entsteht, daß alle Vorkommnisse von „T" durch „E" ersetzt werden, wobei „E" für „(normalwüchsiger) *E*rwachsener" stehe, sieht die Sache viel besser aus. Dann kann nämlich

$$T_3\text{-}3 \qquad \forall x(Wx \wedge Kx \rightarrow \neg Ex)\ ,$$

was besagt, daß kein kleines Kind in Waterloo ein Erwachsener ist, als mehr oder weniger analytisch wahrer Satz durchaus den klaren Vorzug vor T_3-2 ($\equiv T_2$-2) erhalten. Die entsprechende kontrafaktische Hypothese, daß alle Menschen in Waterloo normalwüchsige Erwachsene sind, kann eindeutig und problemlos in T_3 mit aufgenommen werden: durch Streichen von T_3-2.[85] So kann T_3 T_1 kontrafaktisch erklären („Wenn alle Menschen in Waterloo erwachsen wären, dann wäre T_1 gültig."), und so ist T_3 viel eher eine gute Nachfolgertheorie für T_1 als T_2.

Zum Abschluß dieses Miniaturbeispiels möchte ich festhalten, daß auch T_3 keine optimale Nachfolgertheorie von T_1 ist, sondern immer noch einige

[83] Obgleich ich stets die Auffassung vertrete, daß rein Kontingent-Empirisches und Gesetzesartig-Theoretisches *vorderhand* unterschiedslos in einer wissenschaftlichen Theorie zusammengefaßt werden sollten, ist doch Gesetzesartigkeit ein wichtiges intuitives Kriterium. Der Begriff „Gesetzesartigkeit" soll in diesem Buch nicht als vorausgesetzter Begriff, sondern als Explikandum (welches u.a. über die Relation der theoretischen Wichtigkeit zu verstehen ist; s. Kapitel 3) betrachtet werden.

[84] In diesem Lichte betrachtet, ist wohl auch die Beschränkung unserer Theorien auf die Stadt Waterloo nicht zu rechtfertigen.

[85] Nach meiner Auffassung ist eine gewisse Hierarchisierung der Axiome, welche hilft, wohldefinierte Revisionen durchzuführen, (vgl. hierzu Kapitel 3) ein *integraler Bestandteil* von wissenschaftlichen Theorien (vgl. hierzu insbesondere Kapitel 9, Abschnitt 9.5).

Wünsche offen läßt. Oben hatten wir ein Problem mit Scheibes Forderung der Umkehrbarkeit von (1.10.1) zu $T_1 \vdash T_2|_A$. Wir erinnern uns nun daran, daß dieser Forderung im strukturalistischen Rahmen das Kriterium ($K2^{sch\prime}$) entsprach, welches ausschließlich dadurch motiviert war, daß über eine deduktive Erklärung (die wir im vorliegenden Beispiel nachweisen konnten) eine empirische Erklärung garantiert werden sollte. Ein wenig verwunderlich ist es deshalb, daß Scheibe (1983) gar nicht erwähnt, daß das von ihm gewählte Miniaturbeispiel *kein* Beispiel für eine erfolgreiche empirische Erklärung von T_1 durch T_2 abgibt. Denn es gibt viele Erfolge von T_1, denen kein Erfolg von T_2 — oder von T_3 - entspricht; auf beinahe jeden Schüler in Waterloo läßt sich T_1 erfolgreich, lassen sich T_2 oder T_3 aber überhaupt nicht anwenden. Das Nichtzustandekommen der empirischen Erklärung halte ich aber, wie in den vorangegangenen Abschnitten gesagt, für gar nicht schlimm.

Später (in Kapitel 7) wird aber auch eine Art Symmetrieforderung für wirklich überlegene Nachfolgertheorien vorgeschlagen werden. Während Scheibe für (1.10.1) eine Symmetrie in T_1 und T_2 verlangt, werde ich zu dem Ergebnis kommen, daß in den bestmöglichen Fällen von Nachfolgertheorien eine Symmetrie von (1.10.2) in T_1 und A vorliegen sollte. Was ist aber zu den Bedingungen ${T_2}^*{}_{T_1} \vdash A$ und ${T_3}^*{}_{T_1} \vdash A$ zu sagen? Unter der kontrafaktischen Annahme, daß alle Menschen in Waterloo mehr als einen Meter groß sind, kommt man von T_2 bzw. T_3 aus *nicht* zu der Folgerung, daß alle Menschen in Waterloo Tagungsteilnehmer bzw. normalwüchsige Erwachsene sind.[86] Der Grund dafür ist identisch mit dem Grund des Scheiterns von Scheibes empirischer Erklärung: Es sind eben nicht nur Tagungsteilnehmer bzw. normalwüchsige Erwachsene, die größer als einen Meter sind. Die Lücken in T_2 und T_3 sind dafür verantwortlich, daß T_2 und T_3 nicht ohne jeden Abstrich *besser* sind als T_1. Das Gefühl, daß man beim Übergang von T_1 auf T_2 oder T_3 trotzdem eindeutig einen fortschrittlichen Theorienwandel vor sich hat, zeigt, daß die Lücken in den späteren Theorien leicht ausgewogen werden durch das, worauf es offenbar

[86] Wie sieht ${T_2}^*{}_{T_1}$ aus? In T_2 folgt die Negation von T_1-1 aus T_2-1b und T_2-2, deshalb ist eine konsistente Eingliederung von T_1 in T_2 im Prinzip durch Aufgabe von T_2-1b oder durch Aufgabe von T_2-2 möglich. Als Gesetz von T_2 hat T_2-1b jedoch einen privilegierten Status (vgl. auch Fußnote 83) und soll, wenn irgend möglich, beibehalten werden. Deshalb ist es plausibel, daß ${T_2}^*{}_{T_1}$ die Axiome T_2-1, T_2-3 und T_1-1 hat, woraus zwar folgt, daß kein Mensch in Waterloo ein kleines Kind sein kann, aber nicht, daß in Waterloo jeder an der Tagung teilnimmt bzw. daß jeder ein normalwüchsiger Erwachsener ist.

noch mehr ankommt: auf die Erklärung von T_1-Anomalien[87] und auf die „kontrafaktische Erklärung" von T_1 durch die Angabe von Bedingungen, unter denen T_1 gelten würde.

[87]Dies möchte ich unterscheiden von der Erklärung des Scheiterns von T_1 durch T_2 (oder T_3), denn diese letztere Erklärung würde, jedenfalls in dem Sinn, wie ich sie verstehe (vgl. Kapitel 7), doch wieder $T_2 *_{T_1} \vdash A$ bzw. $T_3 *_{T_1} \vdash A$ erfordern.

Kapitel 2

Der klassische Begriff der Reduktion

2.1 Ableitbarkeit und Definierbarkeit: das Originalkonzept von Nagel und Hempel

In Kapitel 1 habe ich versucht, aus der Reduktionsdiskussion im Nonstatement view einen kohärenten Begriff der Reduktion herauszuarbeiten. Nach einer Untersuchung verschiedener logischer Kriterien und Typologien für Reduktionen habe ich mit (1.9.7) sogar so etwas wie einen eigenen Vorschlag in diesem Rahmen gemacht. Totzdem bleibt der Eindruck, daß die Vielzahl an Kriterien — von denen zumindest einige noch dazu auf verschiedenerlei Art und Weise strukturalistisch formalisiert werden können — zu unüberschaubar ist, um zu wirklicher Klarheit zu führen. Es erscheint sinnvoll, noch einmal zu den Wurzeln zurückzugehen, sich den ursprünglichen Reduktionsbegriff anzusehen und sein Schicksal im Verlauf der wissenschaftstheoretischen Debatte zu verfolgen. Der klassische Begriff der Reduktion war jedenfalls von betörender Klarheit. Wie schon in Kapitel 1, Abschnitt 1.2, erwähnt, bestand er aus zwei Teilen, der *Ableitbarkeits-* und der *Definierbarkeitsbedingung*. Eine Theorie T_1 ist nur dann auf eine andere Theorie T_2 reduzierbar, wenn gilt:

(2.1.1) Die Gesetze von T_1 sind aus den Gesetzen von T_2 ableitbar;

(2.1.2) Die Begriffe von T_1 sind durch die Begriffe von T_2 definierbar.

Da die entsprechenden englischen Schagwörter „derivability" und „definability" heißen, werde ich dieses klassische Reduktionskonzept auch das *D-Konzept der Reduktion* nennen. Allerdings findet man die Idee nur selten in diesen prägnanten Begriffen formuliert; meines Wissens drücken sich nur Adams (1959, S. 256, 260f)[1] und Hempel (1966, S. 102; 1969, S. 182) genau so aus. Ernest Nagel, auf den das D-Konzept zurückgeht,[2] scheut sich davor, von Definitionen zu sprechen. Hier ist eine richtungweisende Stelle aus Nagel (1949, S. 301):

> The objective of the reduction is to show that the laws or general principles of the secondary science are simply logical consequences of the assumptions of the primary science. However, if these laws contain expressions that do not occur in the assumptions of the primary science, a logical derivation is clearly impossible. Accordingly, a necessary condition for the derivation is the explicit formulation of suitable relations between such expressions in the secondary science and expressions occurring in the premises of the primary discipline.

Was kann man aus diesem Zitat lernen? Nach Nagel ist (2.1.2) nur eine notwendige Bedingung für (2.1.1): In den — häufigeren und interessanteren — Fällen, wo eine „inhomogene" Reduktion vorliegt, d.h. wo die sekundäre Theorie T_1 Terme enthält, die in der primären Theorie T_2 nicht mit „appoximativ derselben Bedeutung"[3] verwendet werden, sind zusätzliche Annahmen — „connecting laws", „connecting principles" (Hempel 1966, S. 105; 1969, S. 188), „rules of correspondence" oder „bridge laws" (Nagel

[1] Es entbehrt nicht einer gewissen Ironie, daß ausgerechnet Adams, der Begründer des strukturalistischen Reduktionsbegriffs, das D-Konzept besonders kurz und klar formuliert.

[2] Ab und zu wird der historische Zusammenhang verzerrt dargestellt. So erwecken sowohl Suppe (1974, S. 55, Fußnote 113; 1977, S. 624) als auch Nickles (1974, S. 589, Fußnote 21) den Eindruck, daß der klassische Artikel von Kemeny und Oppenheim (1956) stammt, welcher durch Nagel (1961) nur verbessert werde. Tatsächlich ist der 1956er Aufsatz von Kemeny und Oppenheim aber aus der Kritik u.a. an Nagels Konzept, wie es in Nagel (1951) dargestellt ist, entstanden. Die einschlägigen Stellen in Nagel (1961, S. 336–366) stellen eine nicht sehr einschneidende Erweiterung von Nagel (1949) dar und die Auseinandersetzung mit Kemeny und Oppenheim dort beschränkt sich auf eine Fußnote (Nagel 1961, S. 355, Fußnote 5).

[3] Von „approximativ derselben Bedeutung" — eine Ausdrucksweise, die ich nur bei Nagel (1961, S. 339) gefunden habe — zu sprechen, ist ohne weitere Kommentierung nutzlos. Nagel bietet keine Erläuterung, scheint aber etwas von den Einwänden und Auswegen, die Laufe dieses Abschnitts besprochen werden, vorausgeahnt zu haben.

1970, S. 125) — nötig, damit eine Ableitung zustandekommen kann.[4] Es ist Nagel ein großes Anliegen zu betonen, daß diese Annahmen im allgemeinen *keine* Definitionen sind, wenn man Definitionen als analytische, d.h. aufgrund der Bedeutung der in ihnen vorkommenden Begriffe gültige Sätze auffaßt. Vielmehr seien solche Brückenprinzipien materiale, kontingente Hypothesen, da zu ihrer Etablierung und Begründung empirische Forschung nötig sei (Nagel 1949, S. 302–304; 1961, S. 354–358; 1970, S. 125–128). Es ist klar, daß — Nagels Standardbeispiel — „Temperatur" (ein Begriff aus der Alltagssprache oder der phänomenologischen Thermodynamik) nicht sinngleich mit „mittlere kinetische Energie" (ein Begriff der statistischen Mechanik) sein kann. Brückengesetze können aber nach Nagel (1961, S. 354, 356f) als konventionell gesetzte „coordinating definitions" oder nach Hempel (1966, S. 103) als Koextensionalität anzeigende, bikonditionale[5] „extensional definitions" bezeichnet werden, weshalb die Rede von Definitionen eine gewisse Rechtfertigung behält. Ich hoffe, diese Andeutungen genügen, um Mißverständnisse auszuschließen, wenn ich im folgenden die griffigere Redeweise von Definitionen beibehalte.

Schließlich weist das Nagel-Zitat nachdrücklich darauf hin, daß die Ableitbarkeitsbedingung (2.1.1) elliptisch ist: Wenn überhaupt, dann sind die Gesetze von T_1 im allgemeinen nur unter Zuhilfenahme der Brückenprinzipien aus den Gesetzen von T_2 herleitbar. Darüber hinaus ist es meistens so, daß die reduzierte Theorie T_1 einen gegenüber T_2 eingeschränkten Geltungsbereich hat. Ganz analog zu den Antezedensbedingungen im H-O-Schema der Erklärung singulärer Sachverhalte (Hempel und Oppenheim 1948, in Hempel 1965, S. 249) muß also zu T_2 noch eine Beschreibung der Anwendungsbedingungen von T_1 hinzukommen, damit die speziellere reduzierte Theore T_1 abgeleitet werden kann. So besagt etwa eine der zahlreichen Anwendungsbedingungen für die Ableitung des idealen Gasgesetzes in der statistischen Mechanik — so wie Nagel sie beschreibt[6]

[4] Man beachte die Indizierung der Theorien. Während Nagel und manche andere ältere Arbeiten von „primären" und „sekundären" Theorien reden, soll meine Konvention, die reduzierte Theorie als T_1 und die reduzierende Theorie als T_2 zu bezeichnen, der Tatsache Rechnung tragen, daß die reduzierte im allgemeinen (zeitlich) vor der reduzierenden Theorie aufgestellt wird.

[5] Hempel (1969, S. 189) schreibt, daß nicht einmal eine strikte Koextensionalität der verknüpften Terme nötig sei, sondern daß die reduzierende Theorie unter Umständen extensionserweiternd wirken kann. Damit sind aber Brückenprinzipien nurmehr einfache materiale Konditionale (keine Bikonditionale), und eine „reduction of terms", wie Hempel sie an dieser Stelle fordert, liegt nicht mehr vor.

[6] Zur Problematik von Nagels Darstellung vgl. Feyerabends Bemerkungen über die Brownsche Bewegung; siehe Abschnitt 2.3. Vgl. auch Brush (1977).

—, daß die Gasmoleküle völlig elastische Kugeln mit gleichen Massen und gleichen, vernachlässigbar[7] kleinen Durchmessern seien. Wir erhalten also eine vervollständigte schematische Darstellung der Ableitbarkeitsbedingung (2.1.1), wenn wir die Brückenprinzipien und die Anwendungsbedingungen für T_1, die wir in einer Satzmenge A zusammenpacken wollen, mit anschreiben:

(2.1.3) $T_2, A \vdash T_1$.[8]

Damit ist das klassische D-Konzept der Reduktion in seinen Grundzügen vollständig dargestellt.

2.2 Relativierung auf Beobachtungsdaten: die Kritik von Kemeny und Oppenheim

Die erste gewichtige Kritik an diesem Modell kam aus dem eigenen, dem empiristischen Lager, und zwar von Kemeny und Oppenheim (1956). Die beiden Autoren argumentieren zunächst, daß (2.1.2) nicht notwendig (wie Nagel meinte), sondern hinreichend für (2.1.1) sei. Wenn nämlich die T_1-Terme tatsächlich in T_2 definierbar sind — eventuell nur in dem schwachen Sinn, daß man entsprechende gut bestätigte empirische Bikonditionale hat[9] —, dann ist die „Übersetzung" der wohlbegründeten Theorie T_1 durch diese Definitionen (empirischen Bikonditionale) eine wohlbegründete Theorie in der Sprache von T_2 und damit — bei plausiblen Vollständigkeitsannahmen für T_2 — bereits eine Teiltheorie von T_2.[10]

[7] Das heißt nach Nagel: vernachlässigbar gegenüber den Abständen zwischen den Kugeln. Zur Mehrdeutigkeit des Wortes „Vernachlässigung", das sowohl im Approximations- als auch im Idealisierungssinne verstanden werden kann, vgl. speziell für dieses Beispiel Kapitel 8.2 und für allgemeinere Betrachtungen Kapitel 9.

[8] Vgl. auch Feyerabend (1962, S. 46), wo A nur als die Menge der Anwendungsbedingungen für T_1 gedeutet wird. Das Schema kann so nicht funktionieren, wenn man — wie Hempel (vgl. Fußnote 10) — die Brückenprinzipien als Teile der reduzierten Theorie auffaßt.

[9] Kemeny und Oppenheim melden allerdings Zweifel daran an, daß Brückenprinzipien empirisch bestätigt werden können.

[10] Im Gegensatz zu Kemeny und Oppenheim plädiert Hempel (1966, Fußnote 6 und S. 189) dafür, daß die „Connecting principles" als Gesetze der reduzierten Theorie T_1 aufzufassen sind, und zwar deswegen, weil T_1 in der „Präsuppositionshierarchie" wissenschaftlicher Theorien „nach T_2" komme. Hempel (1966, S. 102; 1969, S. 181f, 189) hält (2.1.2) ohnehin für ein völlig eigenständiges Kriterium für Reduktionen. Intuitiv wird (2.1.2) („Reduktion von Termen") nach Hempel durch den Wunsch gerechtfertigt, daß T_2 alle *Beschreibungen* von T_1 repräsentieren kann, während (2.1.1) („Reduktion von Gesetzen") den entsprechenden Wunsch für *Erklärungen* wiedergibt.

Deshalb verlangen sie zusätzlich, daß die Güte der Systematisierung — eine vage Kombination aus Einfachheit und logischer Stärke — von T_2 größer sei als die von T_1. Nagel (1961, S. 355, Fußnote 5) betont aber, ebenso wie Hempel (1969, S. 188f), daß für das Bestehen von (2.1.1) die Brückenprinzipien nicht die logische Form von Bikonditionalen haben müssen, sondern daß hinreichende *oder* notwendige Bedingungen zur Verknüpfung der relevanten Terme im allgemeinen genügen. Dies mahnt zu einem noch größeren Vorbehalt gegenüber der Bezeichnung „Definierbarkeitsbedingung".[11]

Der entscheidende Einwand von Kemeny und Oppenheim richtet sich aber nicht nur gegen die Bikonditionale, auf die (2.1.2) anspielt (1956, S. 13), sondern überhaupt gegen die Idee einer Übersetzbarkeit von Theorien (1956, S. 16). Sie gehen von der (nicht problematisierten) Dichotomie „beobachtbar — theoretisch" aus und relativieren den Reduktionsbegriff auf eine vorgegebene Menge von Beobachtungsdaten.[12] Die zentrale Bedingung ihres endgültigen Vorschlags läuft dann darauf hinaus (1956, S. 15, Theorem 1), daß die reduzierende Theorie T_2 jeden Beobachtungssatz, der von der reduzierten Theorie T_1 impliziert wird, ebenfalls impliziert, oder, falls es einen stärksten von T_1 implizierten Beobachtungssatz T_1^B gibt:

(2.2.1) $T_2, A \vdash T_1^B$.

Wegen des Fehlens „theoretischer Terme" enthält A enthält hier keine Brückenprinzipien, sondern nur Anwendungsbedingungen für T_1.

Halten wir eine kurze Bewertung des Vorschlags von Kemeny und Oppenheim fest. (2.2.1) erhält die Grundidee von (2.1.3) aufrecht. Einerseits verlangen Kemeny und Oppenheim aber mehr als Nagel und Hempel, nämlich die klare Ausgrenzbarkeit einer theorieneutralen Beobachtungssprache, und diese naiv-empiristische Präsupposition wird man heute nicht mehr als erfüllt ansehen. Andererseits — und das ist wichtiger — ist (2.2.1) eine deutliche Liberalisierung gegenüber (2.1.3), indem die für (2.1.3) notwendige „Definierbarkeit" von theoretischen T_1-Termen bzw. Übersetzbarkeit von T_1-Sätzen nicht mehr vorausgesetzt wird. Dieses instrumentalistische Zurückspielen von Theorien auf ihren „rein empirischen" Gehalt soll

[11] Deshalb darf man auch nicht — wie Putnam (1965, S. 206, 208) — „Nagelian reduction" mit „reduction by means of biconditionals" gleichsetzen.

[12] Die Relativierung auf Beobachtungsmaterial findet sich einmal auch bei Feyerabend, und zwar in Feyerabend (1962, S. 206, 208). Dies scheint aber auch schon der einzige Aspekt der Feyerabendschen (Anti-) Reduktionstheorie zu sein, welcher von Kemeny und Oppenheim vorweggenommen wird — auch hier gibt Putnam (1965, S. 206) ein schiefes Bild.

ermöglichen, daß auch Theorien mit völlig unvergleichbaren begrifflichen Gerüsten in einer Reduktionsrelation stehen können. Ich werde (2.2.1) nicht weiter diskutieren, und zwar aus genau den Gründen, die in Sklar (1967, S. 114f) vortrefflich zusammengefaßt sind.[13] Insbesondere schließe ich mich Sklars Meinung an, daß in solchen Fällen, in denen (2.2.1) erfüllt ist, auch stärkere Reduktionsbeziehungen nachzuweisen sind. Darüber hinausgehend hoffe ich, in Fortsetzung der Sklarschen Argumente zeigen zu können, daß sogar in den Fällen, wo (2.2.1) *nicht* erfüllt ist, informative Bedingungen angegeben werden können.

2.3 Inkonsistenz und Inkommensurabilität, die beiden Arten der Nichtmonotonie: die fundamentale Kritik von Feyerabend

Die zweite, viel tiefgreifendere Kritik am Nagel-Hempelschen Modell kommt aus einer sozusagen entgegengesetzten Richtung: von Feyerabend. Ganz im Gegensatz zu Kemeny und Oppenheim kommt für Feyerabend nicht Beobachtungen, sondern Theorien das erkenntnistheoretische Primat zu.[14] Auch ohne vorbereitende Erklärungen dürfte sinnfällig sein, daß Feyerabends Bedingungen an einen Nachfolgkandidaten T_2 für T_1 den klassischen Bedingungen (2.1.1) und (2.1.2) geradezu hohnsprechen: „from the point of view of scientific method, T_2 will be most satisfactory if it is

(2.3.1) *inconsistent* with T_1 in the domain where they both overlap, and if it is

(2.3.2) *incommensurable* with T_1."

(Feyerabend 1962, S. 92f; Hervorhebungen im Original, neue Numerierung und Bezeichnung der Theorien von mir.) Um es mit den in der Einleitung

[13] Man beachte insbesondere folgende Stelle: „The authors [Kemeny und Oppenheim] seem to have hoped that the incompatibility of the reduced and reducing theories in these cases could be isolated at the theoretical level ... But this is not true." (Sklar 1967, S. 115)

[14] Siehe z.B. Feyerabend (1965a, S. 213): „Theories are meaningful independent of observations; observational statements are not meaningful unless they have been connected with theories. ... It is ... the *observation sentence* that is in need of interpretation and *not* the theory." Wenn sich ein Widerspruch zwischen Theorie und Beobachtung zeigen läßt, dann scheint für Theorie und Beobachtung nach Feyerabend (1965a, S. 152) aber eine Symmetrie zu bestehen bzgl. der Chancen, aus dem Konflikt als Sieger hervorzugehen. Vgl. auch Fußnote 31 unten.

eingeführten Begriffen zu formulieren: Feyerabend empfiehlt einen Theorienwandel, der nichtmonoton im ersten und im zweiten Sinne vonstatten geht.

Ehe wir uns der genaueren Erläuterung dieses Zitats zuwenden, sei eine kurze Einschätzung des Feyerabendschen Werks eingeschoben. Jener Art von Wissenschaftstheorie, die traditionell ein streng systematisches, an Logik und Mathematik orientiertes Vorgehen bevorzugt, ist die unter der Parole „Wider den Methodenzwang" provokant vorgetragene „anarchistische Erkenntnistheorie" Feyerabends meistens ein Dorn im Auge. Doch man sollte Feyerabends Beiträge keinesfalls unter den Tisch fallen lassen. Denn erstens ist er ein Kenner auch der neueren Geschichte der Wissenschaft (besonders der Physik), gibt wertvolle Anregungen in halbwegs detaillierten Beispielen und vermeidet damit die Nachteile einer bloß abstrakten Argumentation. Und zweitens — hier wichtiger — kann der Einfluß seines Werkes, das, jedenfalls soweit es für die Reduktionsdiskussion relevant ist, in den wesentlichen Punkten schon mit Feyerabend (1962) und (1963) veröffentlicht war, gar nicht überschätzt werden. Ich möchte sogar behaupten, daß Feyerabend der Hauptverantwortliche dafür ist, daß das klassische Reduktionsmodell so viel von seiner einstigen Bedeutung verloren hat. Und dies, obwohl zentrale Punkte seiner Kritik nicht nur von Nagel und Hempel, sondern auch von sehr vielen anderen Wissenschaftsphilosophen als unhaltbar zurückgewiesen werden.[15]

Die oben zitierte Stelle faßt Feyerabends vehemente, mit großem Nachdruck[16] unterbreitete Kritik an zwei Konsequenzen des Nagel-Hempelschen Modells zusammen, an der *Konsistenzbedingung* und an der *Bedingung der Bedeutungsinvarianz*:

(2.3.3) Only such theories are ... admissible in a given domain which either *contain* the theories used in this domain, or which are at least *consistent* with them inside the domain.

[15] Unten werden Nagel, Hempel und Shapere besprochen. Siehe aber auch Suppe (1974, S. 199–208; 1977, S. 623f).

[16] Der große Nachdruck bei Feyerabend besteht nicht zuletzt auch in der großen Redundanz seines Werkes. Sicher hat er sich aber nicht deswegen so oft wiederholt, weil er nichts anderes zu sagen gehabt hätte, sondern weil er — mit Erfolg — auf die propagandistische Wirksamkeit von Wiederholungen spekuliert hat. Übrigens führt Feyerabend (1963, S. 9; 1965a, Fußnote 78) das kritisierte D-Modell *nicht* auf Nagel oder Hempel, sondern auf Poppers *Logik der Forschung* (1934, Abschnitt 12) zurück. Interessant ist, daß Feyerabend andererseits in seinen Kritiken ebendieses Modells deutlich von Popper (1957) angeregt wurde. Vgl. Feyerabend (1962, Fußnote 113), aber auch Popper (1972, S. 205) und Feyerabend (1980, S. 77, Fußnote 8, S. 354, Fußnote 46); sowohl Feyerabend als auch Popper geben Duhem (1906/78) als Vorreiter an.

(2.3.4) Meanings will have to be invariant with respect to scientific progress; that is, all future theories will have to be phrased in such a manner that their use in explanations does not affect what is said by the theories, or factual reports to be explained.

(Feyerabend 1963, S. 10).[17] Bedingung (2.3.3) ist eine Abschwächung von (2.1.1) bzw. (2.1.3),[18] und Bedingung (2.3.4) ist eng mit (2.1.2) verknüpft.[19] Etwas vereinfacht dargestellt, ist (2.3.1) eine Antithese zu (2.1.1) (bzw. (2.1.3)), und (2.3.2) ist eine Antithese zu (2.1.2). Feyerabend attackiert, wie gesagt, beide Thesen des D-Konzepts heftig, aber es ist deutlich zu spüren, daß ihn das Thema „Inkommensurabilität" („Nichtmonotonie im ersten Sinn") viel mehr interessiert als das Thema „Inkonsistenz" („Nichtmonotonie im zweiten Sinn").[20] Viele werden dieses Interesse teilen, scheint doch die Inkommensurabilität oder die Bedeutungsänderung von wissenschaftlichen Begriffen ein ungleich tiefliegenderes Problem darzustellen als die altbekannte logische Inkonsistenz. Feyerabend macht nur vereinzelt erhellende Bemerkungen zum Zusammenhang zwischen seinen beiden Bedingungen (2.3.1) und (2.3.2). Wir wollen aber gerade den von ihm hergestellten Zusammenhang zwischen den beiden Arten von Nichtmonotonie ganz genau unter die Lupe nehmen, zumal wir ja bereits einige Feststellungen über den Zusammenhang der korrespondierenden „Antithesen" getroffen haben. Zu diesem Zweck lassen wir einige zentrale Verlaut-

[17] Es gibt viele ähnliche Stellen, z.B. Feyerabend (1962, S. 43f; 1965a, S. 164). Vgl. auch (1965b, Abschnitte[!] 4) und (1983, Kapitel 3 und 17).

[18] Vermutlich wechselt Feyerbend deswegen von der Deduzierbarkeits- auf die Konsistenzbedingung, weil er am Ende auf die Inkonsistenz von Theorien hinauswill. Er liefert aber keine Motivation dafür, warum er die beiden Bedingungen zusammenwirft.

[19] Auch der Übergang von (2.1.2) zu (2.3.4) ist nicht ganz überzeugend. In einer Hinsicht ist (2.3.4) *spezieller* als (2.1.2), da es zum Beispiel fordert, daß für ein Prädikat P, welches in T_1 und T_2 verwendet wird, $||P||_{T1}=||P||_{T2}$ gilt, während (2.1.2) nur fordert, daß $||P||_{T1}=||Q||_{T2}$ (bzw. nur $||P||_{T1}\subseteq||Q||_{T2}$ oder $||P||_{T1}\supseteq||Q||_{T2}$) für irgendein in T_2 definierbares Prädikat Q gelte („$||\ldots||_T$" bezeichne hier die T-abhängige Extensionsfunktion). In einer anderen Hinsicht ist (2.3.4) aber auch *allgemeiner* als (2.1.2), da es nichts hinsichtlich der T_1-Terme fordert, die nicht in T_2 vorkommen. Das Nagel-Zitat jedenfalls, welches Feyerabend (1962, S. 33; 1963, S. 10; 1965a, S. 164) in diesem Zusammenhang stets anführt, ist nicht einschlägig. Es besagt nämlich, daß die Bedeutung der T_2-Begriffe durch die „Prozeduren" von T_2 festgelegt und also über die „Verwendungsregeln" von T_2 verstanden würden, und Analoges für T_1. Das Zitat impliziert damit weder Koextensionalität noch die Existenz von Brückenprinzipien zwischen den Ausdrücken von T_1 und T_2.

[20] Ein gutes Abbild der Verhältnisse ist die Länge der entsprechenden Kapitel in Feyerbend (1983): Kapitel 3 geht über die Konsistenzbedingung und hat 16 Seiten, Kapitel 17 über Inkommensurabilität und hat — inklusive Anhang — 81 Seiten.

barungen Feyerabends Revue passieren.

In Feyerabend (1962, S. 81f) folgt auf die Bemerkung, daß (2.3.3) durch (2.3.4) impliziert werde, eine Zusammenfassung der Argumentation (Unterstreichungen i.f. stets von mir):

> Our *argument against meaning invariance* ... proceeds from the fact that usually some of the principles involved in the determination of the meanings of older theories or points of view are *inconsistent* with the new, and better theories. It points out that it is natural to resolve this *contradiction* by eliminating the troublesome and unsatisfactory older principles and to replace them by principles, or theorems, of the new and better theory. And it concludes by showing that such a procedure will also lead to the elimination of the old meanings and thereby to the *violation of meaning invariance*.

In demselben Aufsatz wird nach einer Fallstudie der Begriff der Inkommensurabilität erläutert (1962, S. 74):

> the „inertial law" ... of the impetus theory is *incommensurable* with Newtonian physics in the sense that the main concept of the former, viz., the concept of impetus, can neither be defined on the basis of the primitve descriptive terms of the latter, nor related to them via a correct empirical statement. The reason for this incommensurability was also exhibited: ... the „rules of usage" to which we must refer in order to explain the meanings of its main descriptive terms contain the law [„motion is a process arising from the continuous action of a source of motion, or a ‚motor', and a ‚thing moving.' "] ... and, more especially, the law that constant forces bring about constant velocities. Both of these laws are *inconsistent* with Newton's theory.[21]

In Feyerabend (1963, S. 30) — und fast wörtlich genauso in Feyerabend (1965a, S. 180) — wird Inkonsistenz zum Garant von Nichtreduzierbarkeit erklärt:

> if we consider two contexts with basic principles which either *contradict* each other, or which lead to *inconsistent* consequences in certain domains, it is to be expected that some terms of

[21] Ähnlich deutliche Stellen zur Inkonsistenz als Ursprung der Inkommensurabilität findet man in Feyerabend (1962) auf den Seiten 57, 59 und 75.

> the first context will not occur in the second context with exactly the same meaning. Moreover, if our methodology demands the use of mutually *inconsistent*, partly overlapping, and empirically adequate theories, then it thereby also demands the use of conceptual systems which are *mutually irreducible* (their primitives cannot be connected by bridge laws which are meaningful *and* factually correct).

Am prägnantesten ist Fußnote 19 von Feyerabend (1965a):

> Two theories will be called *incommensurable* when the meanings of their main descriptive terms depend on mutually *inconsistent* principles.

Zur Bekräftigung schließlich noch einige spätere Äußerungen Feyerabends zur Inkommensurabilität. Nachdem er in *Wider den Methodenzwang* (1983, S. 352–354) eine sehr allgemeine Formulierung („einige universelle Prinzipien außer Kraft setzen") verwandt und den Rückgriff auf Logik, Widersprüche und den Statement view überhaupt explizit abgelehnt hat,[22] interpretiert er sich selbst in einer späteren Fassung von Feyerabend (1970) — und genauso in Feyerabend (1977, S. 365) — folgendermaßen:

> Ich erklärte also Theorien für *deduktiv getrennt*, wenn die eine Theorie zusammen mit ihren ontologischen Konsequenzen die *Falschheit* der ontologischen Konsequenzen der anderen Theorie *impliziert*. ... Was dabei wichtig ist, ist, daß bei mir *Inkommensurabilität* nie etwas anderes bedeutet hat, als deduktive Trennung.

(Feyerabend 1978, S. 180; vgl. den englischen Ausdruck „deductive disjointedness" in Feyerabend (1977, S. 365), bei dem man sich wohl nicht durch die Lexikonübersetzung „Zusammenhanglosigkeit" irritieren lassen darf.) Schließlich ist auch die in Feyerabend (1981, S. 142) neu hinzugefügte Fußnote 27a zu Feyerabend (1965b) recht deutlich:

> Ein Begriff ist *inkommensurabel* mit einer Theorie, wenn seine Verwendung gewissen (expliziten, oder bloß impliziten) Grundsätzen der Theorie *widerspricht*.

[22] Ebenso explizit allerdings weist Feyerabend die Behauptung zurück, daß der Nonstatement view in diesem — oder auch in anderem — Zusammenhang Probleme lösen könne, an denen der Statement view scheitert. Im Grunde bevorzugt Feyerabend doch noch eher die letztere Sichtweise. Vgl. vor allem Feyerabend (1977, S. 361, 366).

Diese Parforcetour durch zentrale Stellen des Feyerabendschen Werkes war leider nötig, um eine überraschende und durchaus weitreichende Schlußfolgerung ausreichend abzustützen. Trotz der stets zur Schau gestellten Abneigung Feyerabends gegen Logik und Formalismen in der Wissenschaftstheorie und trotz des scheinbaren qualitativen Unterschieds von Inkommensurabilität und Inkonsistenz halten wir fest: Die hauptsächlichste, vermutlich[23] sogar die einzige Quelle der Inkommensurabilität zweier Theorien — so wie Feyerabend sie präsentiert — ist ihre simple logische Unverträglichkeit.

Dies ist heute wohl weitgehend in Vergessenheit geraten. Ich erhebe aber beileibe keinen Anspruch darauf, einen solchen Zusammenhang als erster herausgearbeitet zu haben. So erinnert zum Beispiel Giedymin (1973, S. 272) daran, daß Philipp Frank bereits 1938 (allerdings mit anderen Argumenten) ähnliche Gedanken formulierte: „To conclude, according to Frank, there is *conceptual disparity between NM [Newtonian Mechanics] and SRM [Special Relativity Mechanics]* which ... is *due to the mutual inconsistency* (*i.e.* logical comparability) *between the two theories.*" Fine (1967, S. 231) referiert das Standardargument für die Bedeutungsänderung wissenschaflicher Terme so, daß es über die Inkonsistenz wissenschaftlicher Theorien läuft.[24] Und Stegmüller (1985, S. 300) führt als die „Feyerabendsche Variante der Inkommensurabilitätsthese ... die Verwerfung der ‚Konsistenzbedingung', wonach nur miteinander verträgliche Theorien einander ablösen sowie seine These von der Theorienabhängigkeit der Bedeutung" an. Feyerabend selbst, der immerhin beansprucht, die Idee der Inkommensurabilität als erster vorgetragen zu haben,[25] thematisiert diese Abhängigkeit nicht — sie wäre für seinen Geschmack wohl zu abstrakt. Ohne zu bestreiten, daß man den Begriff der Inkommensurabilität noch mit weiteren

[23] Diese Einschränkung bezieht sich unter anderem auf folgende zwei Vermutungen: erstens, daß eine Explizitmachung des Feyerabendschen Beispiels „Masse in der klassischen vs. Masse in der speziell-relativistischen Physik (siehe Feyerabend 1963, S. 14–16; 1965a, S. 168–170) ebenfalls auf Inkonsistenzen aufbauen würde (was Feyerabend andeutet); zweitens, daß die oben genannten „ontologischen Konsequenzen" einer Theorie (vgl. auch Feyerabend 1965a, S. 170, 198), wenn man sie explizit macht, auch als logische Konsequenzen derselben zum Ausdruck kommen.

[24] Interessanterweise erwähnen sowohl Frank/Giedymin als auch Fine das unten so bezeichnete „Shaperesche Argument".

[25] Und zwar in Feyerabend (1958). Der Anspruch wird z.B. formuliert in Feyerabend (1977, S. 365), in (1978, S. 30, 179, 204) und in (1983, S. 374f). — Vergleiche aber die folgende Fußnote.

Facetten ausstatten kann,[26] ziehen wir für unseren Zweck, die Reduktionsdiskussion, folgende Zwischenbilanz: Feyerabends Bedingungen (2.3.1) und (2.3.2) deuten nicht auf zwei grundsätzlich verschiedene Eigenschaften von Theorien hin, sondern eher auf zwei Betrachtungsweisen einer Eigenschaft: der Nichtmonotonie im wissenschaftlichen Theorienwandel. Um Fortschritt zu erzielen, müssen Teile des alten Fundus akzeptierter Sätze und Begriffe aufgegeben werden.[27]

Ich will die Betrachtungsweise wählen, von der aus ich die besseren Einsichten zu gewinnen hoffe: die der Inkonsistenz. Diese Seite der Feyerabendschen Kritik ist, jedenfalls was ihre wissenschaftshistorische Adäquatheit angeht, auch ziemlich unumstritten. Man weiß, daß Galileis und Keplers Gesetze mit der Newtonschen Dynamik, daß die geometrische mit der Wellenoptik und daß die klassische mit der speziell-relativistischen Mechanik genaugenommen unverträglich sind.[28] Gleiches behauptet Feyerabend von der phänomenologischen Thermodynamik und der kinetischen Gastheorie (und trifft damit das Lieblingsbeispiel Nagel). Die Brownsche Bewegung eines Partikels stelle ein Perpetuum mobile der zweiten Art dar und widerlege damit den zweiten Hauptsatz der Thermodynamik (vgl. Feyerabend 1962, S. 65f; 1963, S. 23f; 1965a, S. 175f). Überdies zeige dieses

[26] Feyerabend (1977, S. 363–366; 1978, S. 178–182) selbst betont, daß der Kuhnsche Inkommensurabilitätsbegriff wesentlich reichhaltiger ist, indem dieser nicht nur auf die Verschiedenheit von Begriffen, sondern auch auf die Verschiedenheit von Wahrnehmungen und Methoden der Forschung und der Bewertung von Forschungsergebnissen abhebt. Entsprechend wenig genau beschreibt Kuhn das Verhältnis von Inkommensurabilität und (logischer?) Unvereinbarkeit. Wenn er schreibt: „The normal-scientific tradition that emerges from a scientific revolution is not only *incompatible* but often actually *incommensurable* with that which has gone before" (1970, S. 103), so ist Inkommensurabilität vermutlich als eine Art Steigerung von Inkompatibilität gemeint, während im nächsten Zitat die beiden Begriffe anscheinend synonym verwendet werden: „the argument to the problem of choice between two *incompatible* theories, urging in brief conclusion that men who hold *incommensurable* viewpoints be thought of as members of different language communities ..." (1970, S. 175) (Hervorhebungen von mir.) In neuerer Zeit ist Kuhn genauer und setzt Inkommensurabilität konsequent mit Unübersetzbarkeit gleich (s. z.B. Kuhn 1986, S. 4).

[27] Natürlich darf man die Feyerabend-Zitate nicht so verstehen, daß *jede* Inkonsistenz zur Inkommensurabilität von Theorien führt. Wenn sich T_1 und T_2 nur dadurch unterscheiden, daß in einem quantitativen Gesetz von T_1 eine Konstante k_1 durch eine neue Konstante k_2 erstetzt wird (mit $k_1 \neq k_2$, aber eventuell $k_1 \approx k_2$), dann sind T_1 und T_2 zwar inkonsistent, man wird sie aber wegen einer numerischen Feinheit sicher nicht inkommensurabel nennen wollen. Erst „wichtige" Inkonsistenzen wird man für Bedeutungsänderungen der Terme verantwortlich machen. Vgl. Feyerabend (1965b, Fußnote 27) mit dem in Kapitel 3 über „theoretische Wichtigkeit" Gesagten.

[28] Zum Fall Kepler-Newton vgl. Duhem (1906/78, S. 253–260) und Kapitel 8.1.

Beispiel nicht nur die faktische, sondern auch die normativ-methodologische Adäquatheit seines *Proliferationsprinzips*: „Invent, and elaborate theories which are inconsistent with the accepted point of view, even if the latter should happen to be highly confirmed and generally accepted." (Feyerabend 1965b, S. 223f) Ohne die vorgängige Existenz der kinetischen Theorie (und der darauf aufbauenden Berechnungen Einsteins) sei die experimentelle Widerlegung des zweiten Hauptsatzes gar nicht möglich gewesen; die Durchführbarkeit eines Experimentum crucis (Svedberg und Perrin) und damit einer „indirekten Widerlegung"[29] sei abhängig von der Verletzung der Konsistenzbedingung gewesen.

Ich sehe keinen Grund, das Feyerabendsche Kredo, daß einander ablösende Theorien inkompatibel sind und das auch sein sollen, in Frage zu stellen. Feyerabend (1965a, S. 165, 172) definiert sogar Begriffe wie „die Methode der theoretischen Wissenschaft" und „Fortschritt" unter wesentlicher Bezugnahme auf inkonsistente Theorien, und auch hier hat sich kaum Widerspruch geregt.

Die andere Seite der Feyerabendschen Kritik am „radikalen" (vgl. Feyerabend 1965a, S. 149) Empirismus, seine Inkommensurabilitätsthese, hat viel mehr und viel entschiedenere Gegenstimmen gefunden. Erstaunlich ist, daß bei dieser Debatte die Probleme der Inkommensurabilität und der Inkonsistenz weder von Feyerabend noch von seinen Kritikern konsequent als zwei Seiten einer Münze, nämlich der Nichtmonotonie, wahrgenommen werden. Tatsächlich formuliert Feyerabend seine weiterführenden Thesen zur Inkommensurabilität so provokant und mit so ungenauen Begriffen von Bedeutung und Theorie, daß der oben herausgearbeitete Zusammenhang völlig in den Hintergrund gerät und etwa Shapere (1966, besonders S. 53-65) kaum Mühe hat, die Inkommensurabilitätsdoktrin Feyerabends in ihren Grundfesten zu erschüttern.[30] Ich will mich hier darauf beschränken, eine Überlegung Shaperes zu untersuchen, die von Hempel (1969, S. 91) und Nagel (1970, S. 130) als ein Hauptargument gegen Feyerabend verwendet wurde und die Zweifel an meiner Deutung des Verhältnisses von Inkommensurabilität (im Feyerabendschen Sinne) und Inkonsistenz aufwerfen könnte:

[29] Über die Tragweite von „indirekten Widerlegungen" in den Wissenschaften vgl. Feyerabend (1965a, Fußnote 122).

[30] Die Bedeutung von Shaperes Arbeit zeigt sich schon darin, daß sie sowohl von Hempel (1969, Fußnoten 9, 12, 20) und Nagel (1970, Fußnote 12) als auch von Feyerabend (1965b, S. 231f) lobend erwähnt wird. Ähnlich gelagert wie Shapere (1966) ist Achinstein (1964), wo auch das nachfolgend skizzierte Argument schon angeführt ist (deshalb ist meine unten gebrauchte Bezeichnung „Shaperes Argument" kein Prioritätenetikett); siehe besonders die von Achinstein akzeptierte „Annahme B" (1964, S. 499).

> in order for two sentences to contradict one another (to be inconsistent with one another), one must be the denial of the other; and this is to say that what is denied by the one must be what the other asserts; and this in turn is to say that the theories must have some common meaning.

(Shapere 1966, S. 57; auch schon zitiert in Feyerabend 1965b, S. 231f.) Zusammen mit Feyerabends (1965b, S. 231) Behauptung, daß sich beim Theorienwandel in der Regel die Bedeutungen *aller* deskriptiven Terme (insbesondere auch die der sogenannten Beobachtungsterme[31]) ändern, scheint diese fraglos richtige Überlegung zu zeigen, daß die Inkommensurabilität zweier Theorien T_1 und T_2 nicht nur eine Ableitbarkeitsbeziehung und damit eine D-Reduktion, sondern auch eine Inkonsistenz zwischen T_1 und T_2 unmöglich macht. Nehmen wir dazu die oben belegte These Feyerabends, daß es gerade Inkonsistenzen sind, die Inkommensurabilitäten erzeugen, landen wir in einem regelrechten „Paradoxon" (Achinstein 1964, S. 508) — wenn man annimmt, daß es inkonsistente Theorien gibt.

Auf den ersten Blick sieht es so aus, als ob Feyerabend selbst von der Kritik beeindruckt ist, wenn er — ganz anders als in Feyerabend (1962, Fußnote 35) — schreibt: „Terms such as (1) consistent, (2) incompatible, (3) follow from, are applied to pairs of theories, [T,T'] and they mean that (1) there do not, (2) there do, exist pairs of predictions, P and P', one following from T, one from T' (via corresponding initial conditions) which are incompatible." (Feyerabend 1965b, Fußnote 5; Bedingung (3) wurde offenbar vergessen.) Es gibt hier zwei Möglichkeiten, von denen eigentlich keine in Feyerabends Sinn sein kann. Entweder gibt es keinen grundlegenden Unterschied zwischen „Beobachtungssprache" und „theoretischer" Sprache; dann heißen aber „to follow from" und „incompatible" im Definiens dasselbe wie im Definiendum, und die Definition ist offenbar zirkulär. Oder diese Ausdrücke bezeichnen im Definiens die üblichen logischen Relationen; dies setzt aber eine Bedeutungsinvarianz derjenigen Sprache voraus,

[31] Vgl. Feyerabend (1958), These 1: „Die Interpretation einer Beobachtungssprache ist durch die Theorie bestimmt, die wir verwenden, um das zu erklären, was wir beobachten, und sie ändert sich, sobald sich die Theorie ändert." (Zitiert nach Feyerabend 1978, S. 18). Und dazu: „... aus These 1 folgt, daß die Begriffe einer Theorie und die Begriffe einer Beobachtungssprache, mit denen man die Theorie prüfen will, dieselben logischen (ontologischen) Probleme aufwerfen." (1978, S. 19) Andere relevante Stellen sind Feyerabend (1965a, S. 214): „each theory will possess its own experience, and there will be no overlap between these experiences." und Feyerabend (1978, S. 32): „Beobachtungsbegriffe sind nicht *geladen mit* Theorien, sie sind *völlig theoretisch*." — Vgl. auch Fußnote 14.

in der P und P′ formuliert sind — im Widerspruch zu Feyerabends eigener Doktrin, wonach es keinen festen Kern von Theorien gebe (vgl. die Fußnoten 14 und 31). Ich halte die letztere Möglichkeit für die wahrscheinlichere. Deshalb ist die eben angezeigte Kritik an der ziterten „Definition" nur eine Spezialisierung der allgemeineren Kritik an Feyerabends „pragmatischer Theorie der Beobachtung", die schon von Shapere (1966, S. 59–61) geübt wurde.

Ich glaube jedoch, das Shaperesche Argument kann auf andere Weise entkräftet werden. Mit etwas gutem Willen kann man die Sachlage so darstellen: Sei T_2 die Nachfolgertheorie von T_1. Wir dürfen der Einfachheit halber annehmen, daß T_1 und T_2 die gleiche Sprache verwenden, denn inkommensurable Theorien müssen, rein oberflächensyntaktisch gesehen, nicht verschiedene Sprachen haben. Man denke etwa an die (das Fragment der) speziell-relativistischen Mechanik, welche(s) nur Begriffe der klassischen Mechanik benutzt, aber wohl allein schon wegen der Geschwindigkeitsabhängigkeit „ihrer" („seiner") Masse als inkommensurabel mit dieser bezeichnet werden muß. Nach Feyerabend ist T_2 in aller Regel mit T_1 inkonsistent — sofern man, sozusagen „von außen", nur die Syntax der Sätze von T_1 und T_2 betrachtet. Sei nun A ein Satz von T_2 derart, daß $\neg$A ein Satz von T_1 ist. Feyerabend scheint vorauszusetzen, daß sowohl T_1 als auch T_2, „von innen" betrachtet, (in ihrem Anwendungsbereich) korrekte oder „wahre" Bilder der Wirklichkeit liefern.[32] Das kann aber nur dann so sein, wenn mindestens einer der in A vorkommenden Terme in T_1 anders interpretiert wird (d.h. in T_1 eine andere Extension hat) als in T_2. In diesem Sinn sind T_1 und T_2 dann inkommensurabel. (Nicht rechtfertigen kann man hiermit Feyerabends weitergehende Behauptung, daß *alle* Terme uminterpretiert werden; vgl. Feyerabend 1965b, S. 231.) Wenn man die Terminterpretationen von T_1 und T_2 jetzt mitbetrachtet, dann wird man die beiden Theorien aber nicht mehr als inkonsistent bezeichnen. Diese einfache Unterscheidung der syntaktischen „Außen"- und der syntaktisch-und- semantischen „Innen"-Perspektive behebt das Paradoxon und gestattet uns, weiterhin zu sagen, Inkommensurabilität und Inkonsistenz seien bei Feyerabend nur zwei Seiten einer Münze.[33]

[32] Vgl. Hempels (1969, S. 191f) Einwände gegen die Auffassung, daß theoretische Gesetze analytisch sind und theoretische Terme implizit definieren. (Hempel schreibt diese Auffassung Feyerabend zu.)

[33] Man könnte sich hier vielleicht daran stören, daß nach dieser Argumentation die Frage der Bedeutungsänderung und insbesondere der Referenzänderung wissenschaftlicher Terme keine faktische, sondern eine eher konventionell zu entscheidende Frage ist. Genau dies ist aber wohl richtig. Vgl. hierzu die Quintessenz aus Fine (1975, S. 27):

Nagel (1970) und Hempel (1969) allerdings haben in ihren autoritativen, abschließenden Arbeiten zum (Original-)Konzept der Reduktion — nicht zuletzt aufgrund des Shapereschen Arguments — die Feyerabendschen Thesen zur Bedeutungsvarianz als erledigt betrachtet. Die Kritik an der Konsistenzbedingung, zumindest ihre historisch-deskriptive Seite, hingegen wird explizit akzeptiert (vgl. Nagel 1970, S. 120; Hempel 1969, S. 192). Aus diesem Grunde muß das D-Konzept der Reduktion modifiziert werden. Die Ende der 60er Jahre quasi offiziell gültige Antwort auf die Feyerabendsche Kritik sei wieder durch Zitate belegt:

(2.3.5) the arguments ... to show that a given theory is reducible to another ... aim to show that the general laws asserted by the old theory are — within a limited domain, to which its supporting data were restricted — approximations of what the new theory implies for that domain (Hempel 1969, S. 193)[34]

(2.3.6) in homogeneous reductions the reduced laws are either derivable from the explanatory premises, or are good approximations to the laws derivable from the latter ... . inhomogeneous reductions, ... like homogeneous reduction, and with similar qualifications referring to approximations, ... embody the pattern of deductive explanations. (Nagel 1970, S. 121, 125)

Schematisch ausgedrückt, wird (2.1.3) also abgeändert zu

(2.3.7) $T_2, A \vdash T_1^*$ und $T_1^* \approx T_1$,

wobei man sagen kann, T_1^* sei eine Approximation oder „Approximationstheorie" von T_1.

Wie wir sehen, zeigen sich die Verfechter des klassischen Konzepts der Reduktion von Feyerabends (1965b, S. 230) Verdikt „recourse ... to ‚approximate theories' is out" unbeeindruckt. Nun hat zwar der tatsächliche Verlauf der Reduktionsdiskussion gezeigt, daß angenäherte Theorien nicht aus der Mode gekommen sind; man bedenke nur die Literatur im Rah-

„whenever a case can be made for sameness of reference, an equally good case can be made for difference of reference, and conversely."

[34] In sympathischer Bescheidenheit schreibt Hempel im anschließenden Abschnitt das so liberalisierte D-Konzept der Reduktion Smart (1965) und Putnam (1965) zu. Man vergleiche aber Hempel (1965, S. 344): „the theory involved does not, strictly speaking, imply the presumptive laws to be explained; rather, it implies that those laws hold only within a limited range, and even there, only approximately." Hempel gilt auch schon der Dank von Kemeny und Oppenheim (1956, S. 13) für die Anregungen zu einem Absatz, der folgenden Satz enthält: „the old theory usually holds only within certain limits, and even then only approximately."

men des Non-statement view.[35] Doch können wir durchaus eine intuitive Kritik an der vorgeschlagenen Liberalisierung des D-Konzepts anbringen: Wenn wir Theorien als Mengen von Sätzen auffassen, wie verhält sich dann die „Approximation" T_1^* einer Theorie T_1 zu T_1? Nach der nächstliegenden Deutung ist intendiert, daß die „beobachtbaren" oder „nichttheoretischen" Konsequenzen von T_1 mit denen von T_1^* „fast genau" übereinstimmen sollen. Mit dieser Explikation handeln wir uns aber sowohl das grundsätzliche Problem der Beobachtbarkeit bzw. (T_1-)Theoretizität als auch die notorische Vagheit des „fast genau" ein. Außerdem sind wir auf eine eher instrumentalistische Interpretation von Theorien festgenagelt, welche die strukturellen Beziehungen zwischen den stark „theoriehaltigen" Gesetzen von T_1 und T_2 unberücksichtigt läßt. Dabei wäre es doch besonders wünschenswert, daß wir angeben könnten, „an welcher Stelle" denn genau die reduzierende Theorie T_2 besser, richtiger, informativer ist als die reduzierte Theorie T_1. Dazu liefert der blanke Verweis auf eine „Approximation" keine Einsichten. Man vergleiche die Stellungnahme Feyerabends (1962, S. 48):

> the remark that we explain ‚by approximation' is much too vague and general to be regarded as the statement of an alternative theory. As a matter of fact, ... the idea of approximation cannot any more be incorporated into a formal theory, since it contains elements which are essentially subjective.

An dieser Stelle können wir nichts anderes tun, als dem ersten Satz der Feyerabendschen Bewertung zustimmen. Für die weitergehenden Behauptungen, daß Approximationen stets wesentlich subjektiv seien und daß *deshalb* keine formale Theorie über Approximationen möglich sei, müßte man aber noch wirklich überzeugende Gründe beibringen. Wie dem auch sei, ganz sicher ist, daß die Aburteilung von Approximationen keine hinreichende Begründung für Feyerabends (1962, S. 28) forsche Schlußfolgerung „a formal account of reduction and explanation is impossible for general theories" ist. Denn wir werden bald sehen, daß Approximationen nicht die einzig mögliche Antwort auf Feyerabends Inkompatibilitätseinwand sind. In späteren Teilen dieses Buchs (ab Kapitel 7) werde ich den Alternativbegriff der Idealisierung vorstellen. Auch bei Idealisierungen kann man

[35] Siehe zum Beispiel Ludwig (1978), Moulines (1976; 1980; 1981) und Mayr (1981). Vgl. auch Scheibe (1973) und Ehlers (1986). Genaueres diskutiere ich am Beispiel der Beziehung zwischen den Keplerschen Gesetzen und der Newtonschen Gravitationstheorie in Kapitel 8.1.

— in gewissen Deutungen — subjektive Elemente ausmachen, was uns aber nicht davon abhalten wird, eine formale Grundlage für Idealisierungen zu entwickeln (in den Kapiteln 3–6). Zudem werde ich Argumente dafür anführen, daß die für Idealisierungen grundlegende Relation der „theoretischen Wichtigkeit" nicht subjektiv, sondern als Teil einer wissenschaftlichen Theorie selbst aufgefaßt werden sollte.

Fassen wir noch einmal das Ergebnis dieses Abschnitts zusammen. Das einfache D-Konzept der Reduktion wurde von Feyerabend erfolgreich angegriffen. Gewiß hatte er Recht mit seiner Kritik an der Konsistenzbedingung. Seine Attacke auf die Bedingung der Bedeutungsinvarianz jedoch erscheint entweder — wenn man Shapere, Hempel und Nagel folgen will — verfehlt oder im Kern nichts anderes als eine Attacke auf die Konsistenzbedingung — wenn man, wie ich, die angeführten Belegstellen als Nachweis dafür ansieht, daß Inkommensurabilitäten im Sinne von Feyerabend eine Folge von Inkonsistenzen sind. Einerlei, welcher dieser Argumentationen man sich anschließt, für das klassische Konzept der Reduktion bleibt das Problem der Inkonsistenz und nur dieses bestehen. Und dies ist nicht die Schlußfolgerung aus der Lektüre eines leichthin simplifizierenden „Formalisierers", sondern eines wissenschaftshistorisch sehr beschlagenen „erkenntnistheoretischen Anarchisten". Wir dürfen also hoffen, ohne Angst vor Inkommensurabilität die Idee des klassischen D-Konzepts mit formalen Mitteln retten können.[36]

Dazu dürfen wir natürlich nicht einfach wie Feyerabend sagen, daß inkonsistente Theorien gegenseitig irreduzibel sind. Vielmehr möchte ich behaupten, daß auftretende Inkonsistenzen analytisch bearbeitet und aufgelöst werden können, und zwar — dies ist die Schlüsselidee — nach einem formalen Modell, das im Rahmen eines ganz anderen Forschungszusammenhangs in der analytischen Philosophie entwickelt worden ist. Die Kernbedingung (2.1.3) wird dabei nicht zur beanstandeten Bedingung (2.3.7), sondern zu einer anderen, informativeren Bedingung abgewandelt werden. Doch der Weg dahin ist noch weit. Wer von keiner der beiden oben genannten Argumentationsstrategien überzeugt worden ist und auf einer tiefliegenden, eigenwertigen Bedeutungsverschiedenheit von sprachli-

[36] Daß das D-Konzept einer Rettung bedarf, ist nicht nur Feyerabends Ansicht, der meint, daß der Rückgriff auf approximative Theorien „nichts als ein verdecktes Eingeständnis der Niederlage" ist (Feyerabend 1965b, S. 229, zitiert nach 1981, S. 136). Man bedenke auch Hempels (1969, S. 197) Fazit: „the construal of theoretical reduction as a strictly deductive relation between the principles of two theories, based on general laws that connect the theoretical terms, is indeed an untenable oversimplification which has no strict application in science".

chen Ausdrücken verschiedener Theorien besteht, wird von diesem Buch nicht zufriedengestellt werden. Er mag sich mit der pragmatischen Entschuldigung abfinden, daß eine allgemein akzeptable Bedeutungs- und Referenztheorie der (Wissenschafts-)Sprache und ihre Anwendung auf aktuell vorliegende wissenschaftliche Theorien ganz sicher nicht nebenbei in einer Doktorarbeit geleistet werden kann. Statt eines solchen Versuchs, der dilettantisch bleiben müßte, werde ich also die Voraussetzung machen, daß eventuell zu diagnostizierende Bedeutungsverschiebungen keine prinzipielle Unübersetzbarkeit und keinen Kommunikationszusammenbruch zwischen Vertretern konkurrierender Theorien zur Folge haben würden. Insofern wird meine Analyse vielleicht idealisiert und in ihrer Reichweite eingeschränkt bleiben müssen.

2.4 Inkonsistenz, Approximation und Kontrafaktizität: Sklar, Schaffner, Fine, Glymour, Eberle und Nickles

Wie hat nun, von heute aus betrachtet, das klassische Konzept der Reduktion den Angriff Feyerabends überstanden? Zur Beantwortung dieser Frage wie auch zur Herausarbeitung der Umrisse einer systematischen Verbesserung des klassischen Konzepts möchte ich sechs einschlägige Aufsätze heranziehen: Sklar (1967), Schaffner (1967), Fine (1967), Glymour (1970), Eberle (1971) und Nickles (1973). (Seitenzahlen beziehen sich, wenn nicht anders gekennzeichnet, für den Rest des Kapitels immer auf diese Arbeiten der genannten Autoren.) Die Aufsätze sind einerseits noch unter dem frischen Eindruck der Attacke von Seiten der wissenschaftshistorisch orientierten „Antiempiristen" entstanden, andererseits sind sie unaufgeregt, kompetent und bis heute nicht veraltet, da m.E. auf diesem Gebiet aus noch auszumachenden Gründen seitdem keine wirklich entscheidenden Fortschritte erzielt worden sind. Zusammengenommen verschaffen die kaum 90 Seiten einen ausgezeichneten Überblick über verschiedene Aspekte und Begriffe der intertheoretischen Reduktion und können, bei aller am einfachen Nagel-Hempel-Modell geübten Kritik, als autoritative Darstellung dessen angesehen werden, was aus dem klassischen Modell geworden ist.[37]

[37] Anzumerken ist, daß in Fines Aufsatz das Wort „Reduktion" überhaupt nicht vorkommt. Er beschäftigt sich mit dem Übergang von einer Theorie T_1 zu einer anderen Theorie T_2, insbesondere mit dem — normalen — Fall, in dem T_1 und T_2 (theoretische)

Die genannten Arbeiten stimmen insofern mit der im vorigen Abschnitt abgegebenen Bewertung der Feyerabendschen Kritik überein, als sie allesamt den Inkommensurabilitätseinwand geringschätzen, den Inkompatibilitätseinwand jedoch sehr ernst nehmen. Zum Thema Kommensurabilität: Sklar (S. 113–121) hält die reduzierende Theorie T_2 und die reduzierte Theorie T_1 mindestens bezüglich ihrer in der gemeinsamen Teilsprache formulierten Voraussagen („(approximative) partielle Reduktion"), oft aber auch bezüglich der theoretischen Terme („identifikatorische Reduktion" über synthetische Identitäten) für vergleichbar; Schaffner (S. 140, 144) setzt voraus, daß T_2 Konsequenzen hat, die einerseits — sogar nach dem „Popper-Feyerabend-Kuhn-Paradigma" der Reduktion — T_1 „korrigieren" und andererseits große Ähnlichkeit„ mit und „starke Analogie" zu T_1 aufweisen; Fine (S. 232–236) weist die Diagnose einer Bedeutungsänderung mancher beim Theorienwandel beibehaltener Terme als (sprach-)philosophisch unhaltbar zurück; gemäß Glymour (S. 342, 345f) kann man aus T_2 typischerweise eine Satzmenge „generieren", die — nach einer definitorischen Erweiterung zur Herstellung eindeutiger Korrespondenzen von T_1-Termen mit T_2-Termen — T_1 impliziert; Eberle (S. 483–491) präsupponiert die Existenz einer „repräsentierenden Funktion", welche Formeln von T_1 in Formeln von T_2 übersetzt;[38] und schließlich mißt Nickles (S. 186–189) dem Einwand der Bedeutungsänderung von Termen beim Übergang von T_1 auf T_2 wenig Bedeutung zu, zumal ja die von ihm in den Vordergrund gestellte „Reduktion$_2$" (und, spezieller, die „bereichserhaltende Reduktion") wegen ihrer hauptsächlich „rechtfertigenden" und „heuristischen" Funktion relativ unempfindlich gegen diesen Einwand sei.[39]

Terme gemeinsam haben. Es besteht aber kein Zweifel, daß man Fines Arbeit als einen unmittelbar relevanten Beitrag zur Reduktionsdiskussion lesen kann und auch sollte.

[38] Repräsentationsfunktionen mit semantischen Forderungen werden von Schroeder-Heister und Schaefer (1989) diskutiert.

[39] Nickles' „Reduktion$_2$" liegt vor allem dann vor, wenn sich T_2 durch den Grenzübergang eines T_2-Parameters „auf T_1 reduziert". In diesem angeblich wesentlich nichtdeduktiven Fall müsse man die normale Richtung der Reduktion umkehren und von der „Reduktion einer Nachfolgertheorie T_2 auf ihren Vorgänger T_1" sprechen. Nickles' Zwei-Typen-Klassifikation leidet aber unter zwei argumentativen Mängeln: Erstens unterscheidet er nicht zwischen dem transitiven „*reduzieren auf*"/ „*reduziert werden auf*" („to reduce s.th. to"/„to be reduced to") und dem reflexiven „*sich reduzieren auf*" („to reduce to"), und zweitens vermengt er die (Nagelsche) Dichotomie „heterogene vs. homogene Reduktion" und die (in unserem Zusammenhang wichtige) Dichotomie „Konsistenz/Ableitbarkeit vs. Inkonsistenz zwischen T_1 und T_2" zu einer einzigen Dichotomie „Reduktion$_1$ vs. Reduktion$_2$". — Ein ähnliches Umkehrungsverhältnis zwischen „historischer" und „praktischer" Reduktion konstatieren Balzer, Moulines und Sneed (1987, S. 253f), die sich sonderbarerweise dazu entschließen, „T_1 reduziert sich auf T_2" dann

Hingegen wird dem Einwand, daß sich T_1 und T_2 typischerweise logisch widersprechen, große Aufmerksamkeit zuteil. Die Autoren gestehen in der Regel ohne Zögern zu, daß für „viele wichtige Fälle von Reduktion in der Wissenschaft" (Sklar, S. 111f) die reduzierte mit der reduzierenden Theorie unverträglich ist, daß „in wichtigen und typischen Fällen der Übergang von einer wissenschaftlichen Theorie zu einer anderen, welche mit ihr inkonsistent ist, erfolgt" (Fine, S. 231), daß „die paradigmatischen Fälle intertheoretischer Erklärung Theorien involvieren, die inkonsistent sind" (Glymour, S. 340) oder daß „reduzierte und reduzierende Theorie beinahe immer logisch unverträglich sind" (Nickles, S. 188). Ohne das Wort „Inkonsistenz" zu benutzen, scheinen Schaffner und Eberle doch das Phänomen Inkonsistenz anzuerkennen, wenn sie sagen, daß im Falle einer Reduktion T_2 Konsequenzen hat, die T_1 „korrigieren" und „anzeigen, warum T_1 inkorrekt war" (Schaffner, S. 144) bzw. daß „häufig alte Theorien durch neue ersetzt werden, die ... einen Irrtum der früheren Theorien korrigieren" (Eberle, S. 496).[40]

Nachdem wir nun die grundlegende Gemeinsamkeit der in Rede stehenden Arbeiten erkannt haben, wollen wir uns ihren Unterschieden zuwenden. Die entscheidende Frage, die sich stellt, ist nämlich: Wie wird die Inkonsistenz von T_1 und T_2 modelliert, so daß man trotz Inkonsistenz noch von einer Reduktion sprechen kann? Hier lassen sich, wenn ich es recht sehe, vier verschiedene Ansätze auseinanderhalten, die ganz grob durch die Schlagworte „Bereichseinschränkung", „Approximation im ersten/zweiten Sinn" und „Kontrafaktizität" zu kennzeichnen sind.

2.4.1 Bereichseinschränkung, Approximation und Kontrafaktizität

Die Idee der *Einschränkung des Geltungsbereiches* von T_1 durch Relativierung auf eine Substruktur wurde offenbar zuerst von Eberle (S. 496–498) vorgebracht. Eberle will damit die Ablösung „inakkurater" Theorien explizieren. Wir sind dieser Idee bereits in den Abschnitten 1.9 und 1.10

zu sagen, wenn T_2 die komplexere oder bessere Theorie ist.

[40] Eberles Beispiel hier (S. 496–498) ist die Ersetzung von Galileis Fallgesetz (G) durch „Newtons Gesetz" (N) für fallende Körper. In der von ihm präsentierten Form von (G) und (N) liegt keine Inkonsistenz vor, denn aus ihrer Konjunktion folgt einfach, daß die Höhe des betrachteten Körpers über dem Erdboden konstant ist. Dies ist aber für *fallende* Körper eine sicherlich zurückzuweisende Folgerung. Mit der Zusatzbedingung „h≠konstant" (als Teil von (G) und/oder (N)) kann man die intuitiv erwartete Inkonsistenz bekommen. Vgl. auch Kapitel 1, Fußnote 77, und Kapitel 8.1, Fußnote 42.

bei der Untersuchung der Vorschläge Erhard Scheibes begegnet, der eine Bereichseinschränkung generell, also auch für strikte Reduktionen vorsieht. So kann man etwa die Theorie T_1 im Miniaturbeispiel aus Abschnitt 1.10 kaum als inakkurat bezeichnen — sie ist ganz einfach falsch. Von daher ist uns schon bekannt, wie das Schema

(2.4.1) $T_2 \vdash T_1|_A$ $(\equiv (1.10.1))$

funktioniert und wir brauchen nicht noch einmal darauf einzugehen. Eberle spricht hier von einer „Reduktion von T_1 auf T_2 relativ zu den von A ausgedrückten Bedingungen". Genauer gesagt ist T_1 nach Eberle genau dann auf T_2 reduzierbar, wenn es eine effektive, die syntaktische Struktur erhaltende Übersetzung $\ddot{U}$ von Formeln der Sprache von T_1 in Formeln der Sprache von T_2 gibt, so daß logische Folgerungen aus T_1 auf logische Folgerungen von T_2 abgebildet werden. Im hiesigen strukturalistischen Rahmen entspricht Eberles Reduktionsbegriff sehr genau der Umkehrung von $(K8^{s\prime\prime})(a)$:

$$M_1 \subseteq B \Rightarrow M_2 \subseteq ||\ddot{U}(B)||$$

Wir untersuchen die Beziehung dieser Bedingung zu $(K2^s)$ $\mathbf{F}[M_2] \subseteq M_1$. Einerseits folgt $(K2^s)$ aus ihr, wenn M_1 in T_1 definierbar ist: Sei B so, daß $M_1 = ||B||$. Dann gilt nach nach der eben angegebenen Bedingung $M_2 \subseteq ||\ddot{U}(B)||$, also $M_2 \cap M_{p2}^{\circ} \subseteq \mathbf{F}^{-1}[||B||] = \mathbf{F}^{-1}[M_1]$, also $\mathbf{F}[M_2] = \mathbf{F}[M_2 \cap M_{p2}^{\circ}] \subseteq \mathbf{F}[\mathbf{F}^{-1}[M_1]] \subseteq M_1$. Man nehme andererseits an, daß $(K2^s)$ gelte und $M_1 \subseteq ||B||$. Dann gilt $M_2 \cap M_{p2}^{\circ} \subseteq \mathbf{F}^{-1}[\mathbf{F}[M_2]] \subseteq \mathbf{F}^{-1}[M_1] \subseteq \mathbf{F}^{-1}[||B||] = ||\ddot{U}(B)|| \cap M_{p2}^{\circ}$. Aber das gewünschte $M_2 \subseteq ||\ddot{U}(B)||$ bekommt man nur, wenn entweder $M_2 \subseteq M_{p2}^{\circ}$ ist oder die Übersetzung außerhalb von M_{p2}° „vernünftig" oder leer (d.h. $||\ddot{U}(C)|| \cap \mathcal{C}M_{p2}^{\circ} = \mathcal{C}M_{p2}^{\circ}$ für jedes C) ist.[41]

Die am meisten in Anspruch genommene, aber durchaus nicht ganz geklärte Möglichkeit, mit Inkonsistenzen zu Rande zu kommen ist wohl die Idee der Approximation. Zunächst einmal wollen wir festhalten, daß es zwei deutlich verschiedene Intuitionen gibt, die — meist ohne genau auseinander gehalten zu werden — beide unter dem Titel „Approximation" geführt werden. Zum ersten zielen Sklar (S. 111, 117) und Schaffner (S. 140, C(2), 144, (4)) auf ein Schema der folgenden Art ab:

[41] Eberle zeigt in seinen Theoremen 4 und 5, daß diese Bestimmung dem strukturalistischen Kriterium $(K2^s)$ mit der Zusatzbedingung $M_2 \subseteq M_{p2}^{\circ}$ entspricht, wenn man $\mathbf{F}(x_2)$ als dasjenige (im Rahmen von T_1 eindeutig bestimmte x_1 definiert, für welches gilt:

$\forall$ T_1-Sätze A ($x_1 \models A$ gdw. $x_2 \models \ddot{U}(A)$) .

Aber Eberle geht von $\ddot{U}$ aus und konstruiert $\mathbf{F}$, im Unterschied zum strukturalistischen Verfahren.

(2.4.2) $T_2 \vdash T_1^*$ und $T_1^* \approx T_1$,

T_1^* ist hier „approximativ gleich" T_1 in dem Sinn, daß es annähernd dieselben Voraussagen liefert wie T_1. (2.4.2) ist eine vergröberte Form des späten Nagel-Hempel-Kriteriums (2.3.7). Zum zweiten haben Fine (S. 238f), Glymour (S. 342-345) und Nickles (S. 194–199) ein Schema nach Art von

(2.4.3) $\lim_{p \to p^*} T_2 = T_2^*$ und $T_2^* \vdash T_1$

im Sinn. T_2^* wird hier durch Approximation aus T_2 gewonnen, i.a. dadurch, daß man einen (oder mehrere) „Parameter" p aus T_2 gegen einen Grenzwert p* (welcher selbst oft als Wert der Parameter ausgeschlossen ist) „gehen läßt".

Offensichtlich sind die beiden Begriffe von Approximation grundsätzlich verschieden, und entsprechend verschieden sind auch die kritischen Bemerkungen, die über sie gemacht werden können. Wenn nur eine *Approximation im ersten Sinn*, also im Sinn von (2.4.2), vorliegt, dann scheint man durchaus noch keinen hinreichenden Grund dafür zu haben, von einer Reduktion von T_1 auf T_2 zu sprechen. Denn Theorien von gänzlich verschiedener, eventuell unübersetzbarer Begrifflichkeit, von verschiedener Gesetzesstruktur und verschiedener Ontologie können zufälling zu ähnlichen, unter Umständen sogar zu gleichen Voraussagen führen, ohne daß man im entferntesten geneigt wäre, von einer Reduktion zu sprechen, weil T_1 ja „nichts Wesentliches" mit T_2 gemein hat. Für eine Reduktion muß T_2 auch eine strukturelle Verwandtschaft mit T_1 haben (vgl. dazu Sklar, S. 115–117, Schaffner, S. 144, (5), Fine, S. 237, (1), Glymour, S. 341, Nickles, S. 185, Nickles 1974, S. 587–589).

Anders liegt die Sache im — interessanteren — Fall der *Approximation im zweiten Sinn*, im Sinne von (2.4.3), wo durch den „Grenzübergang" eines Parameters (und anschließende Deduktion) eine enge strukturelle Beziehung zwischen T_2 und T_1 hergestellt wird. Die Frage ist, was *genau* bei diesem Grenzübergang passiert. Ich möchte nun die Vorstellungen der hier betrachteten Autoren anhand einiger Zitate zu Fallbeispielen veranschaulichen und dabei auf einige Verwirrungen hinweisen. (Die Hervorhebungen in den folgenden Zitaten stammen von mir.)

Ein nicht nur bei Feyerabend beliebtes Beispiel ist das Verhältnis (von Teilen) der speziell-relativistischen Mechanik zu (Teilen) der klassischen Mechanik. Für Sklar (S. 116) liegt hier ein klarer Fall von approximativer *partieller*[42] Reduktion vor:

[42] Das Attribut „partiell" bei Sklar zielt auf den Gedanken der Bereichseinschränkung ab, das Attribut „approximativ" auf Approximation im ersten Sinn. Die beiden At-

> If we first restrict our attention to sentences framable in purely kinematic concepts, and then further restrict our attention to the subset of sentences in this class *dealing with sufficiently low velocities*, we find for every sentence in this subset derivable from the relativistic theory there is a sentence approximate to it derivable from the Newtonian theory.

Kritische Fragen hierzu: Handeln kinematische Sätze von Geschwindigkeiten? Nein, sie handeln von physikalischen Objekten und Systemen. Gibt es *einen* genügend niedrigen Schwellenwert für Geschwindigkeiten, so daß man immer approximativ richtige Voraussagen erhält, wenn dieser Wert unterschritten wird? Nein. Denn der Schwellenwert ist abhängig vom gewünschten Genauigkeitsgrad und von der Länge der Zeitspanne, während der man ein Objekt oder System betrachten will. — Fine (S. 239) klassifiziert dieses Beispiel als einen typischen Fall, im dem Ableitbarkeit[43] trotz Inkonsistenz vorliegt:

> ... relativistic mechanics encompasses classical mechanics. The basic terms are shared ... — ‚mass', ‚force', ‚time', ‚distance', etc. — and, assuming that the velocity is *small relative to the velocity of light*, the „laws of mechanics" are readily *derived*.

Stimmt es aber wirklich, daß, wenn man von der relativistischen Mechanik ausgeht, bei (relativ) kleinen Geschwindigkeiten die Gesetze der klassischen Mechanik gelten? Nein, sie gelten auch dann nur ungefähr. Kann man die klassischen Gesetze also ableiten, wenn nur die Geschwindigkeiten (relativ) klein genug sind? Nein, denn strenggenommen widersprechen sie auch auch dann noch der Relativitätstheorie. „Ableiten" könnte man sie nur, wenn man in die relativistischen Gesetze als Voraussetzung $v/c=0$ einsetzen dürfte, was aber in nichttrivialen Fällen ($v\neq 0$) wegen der Endlichkeit der Lichtgeschwindigkeit ($c\neq\infty$) nicht angeht — jedenfalls nicht ohne Schwierigkeiten zu verstehen, was wir dadurch mit den relativistischen Gesetzen eigentlich anstellen. — Nickles (S. 184, 196) erwähnt dasselbe Beispiel als einen paradigmatischen Fall seiner „Reduktion$_2$" (die er von einer approximativen Reduktion im ersten Sinn abgrenzt):

---

tribute kann man, wie wir unten zu zeigen versuchen, so kombinieren, daß man eine Approximation im zweiten Sinn erhält.

[43] Fine benutzt fast immer die Wendung „to encompass" (dt. um-, einschließen), manchmal auch „to be derivative from", bei der Bearbeitung der Frage nach der Ableitbarkeit von T_1 aus T_2. Es ist nicht klar, ob und — wenn ja — warum er sich damit vom gewöhnlichen Ableitbarkeitsbegriff distanzieren will.

> STR reduces to CM in the limit of low velocities; i.e., the operation of *taking the limit on STR* transforms it (leads us back) to a basic version of CM. ... We do not get a corrected secondary theory analogous to the historical CM by logical derivation from STR. It is only by taking the limit of certain STR equations as *the velocity of the system goes to zero* (or as *the maximum signal velocity goes to infinity)* that we get the post-Einstein version of CM.

Was ist „die Operation des Grenzwertnehmens auf STR"? Ich weiß es nicht. Was heißt es, daß die Geschwindigkeit „des Systems" gegen Null geht? Die Objekte in *dem* System — das man zum Beispiel gerade vor sich hat — haben feste Geschwindigkeiten, die man nicht einfach „gegen Null gehen" lassen kann. Oder was heißt es, daß die maximale Signalgeschwindigkeit (Lichtgeschwindigkeit) gegen unendlich geht? Sie ist doch fest, und zwar endlich — eine Tatsache, die für die Relativitätstheorie bekanntlich von entscheidender Bedeutung ist. (Die Idee $c \to \infty$ wird später, auf S. 198–201, von Nickles als „mathematisch illegitim" und „physikalisch sinnlos" zurückgewiesen.[44])

Schwierigkeiten hat auch Glymour (S. 343) mit der Präsentation eines harmloseren Beispiels:

> ... the equation for the van der Waals' gas law
> $$P = RT/(V\text{-}b) - a/V^2$$
> *becomes identical in form* with the equation for the ideal gas law
> $$P = RT/V$$
> *when* $V \to \infty$, *a and b constant.*

Ist es hier wirklich angemessen, von einer „Gleichheit der Form" beider Gesetze zu sprechen, wenn das Molvolumen bei konstantem a und b gegen unendlich geht? Ich denke nicht. Glymours Erläuterung, wonach „Gleichheit der Form" als eine Identität zwischen (Kombinationen von) Symbolen von T_1 und T_2 aufzufassen sei, hilft hier auch nicht weiter.[45] Es

[44] Beachte aber Nickles, S. 201, Fußnote 36.

[45] Aufgrund dieser Erläuterung könnte man auf die Idee kommen, das ‚V-b' des van der Waalsschen mit dem ‚V' des idealen Gasgesetzes und das ‚$P+(a/V^2)$' bei van der Waals mit dem ‚P' im idealen Gasgesetz „gleichzusetzen". Dann aber wäre — ganz abgesehen von erheblichen Deutungsproblemen — der Approximationsprozeß $V \to \infty$ nicht mehr wesentlich.

stimmt zwar, daß bei größerem Molvolumen das ideale Gasgesetz bessere Voraussagen liefert, d.h. hier: Voraussagen, die näher an den Voraussagen des van der Waalsschen Gasgesetzes liegen. Dies ist aber kein verbindliches Kriterium für eine formale „Analogie" oder „Korrespondenz". Ebenso ist Glymours (S. 344) Formulierung „a and b ... disappear on taking limits" einfach zu lässig, um den Grenzübergang $V \to \infty$ korrekt und erhellend wiederzugeben.[46]

Als letztes zum wohl einfachsten Beispiel: dem Verhältnis zwischen dem Galileischen Fallgesetz und Newtons Gravitationstheorie (inklusive Mechanik). Lassen wir zunächst Schaffner (S. 139) zu Wort kommen:

> The Galilean law is not exactly derivable — rather a more complicated law is derivable which gives experimental results which are quite close to the predictions of the Galilean law. The sentences expressing these laws are still different, however, and could only be said to be *formally identical* if the earth's radius were *infinitely large*, which it is not. ... *Consequently* the reduced theory is only derivable approximately from the reducing theory ...

Wäre Galileis Fallgesetz

$$\exists k \in \mathbb{R}\ (g = k) \quad \text{(g Fallbeschleunigung)}$$

wirklich „formal identisch" mit der Newtonschen Gleichung

$$g = GM/(R+h)^2$$

(G universelle Gravitationskonstante, M Erdmasse, R Erdradius, h Höhe des betrachteten Körpers über der Erdoberfläche),[47]

wenn die Erde — was äußerst schwer vorstellbar ist — unendlich groß wäre? Die Antwort lautet „nein", denn für beliebig großes, aber festes R ist g nach der Newtonschen Gleichung eben nicht konstant, und für „$R=\infty$" ist g bei endlicher Erdmasse gleich null, bei unendlicher Erdmasse aber völlig unbestimmt. Ist dieses Beispiel einschlägig für das sog. Popper-Feyerabend-Kuhn-Paradigma der Reduktion, wo man nach Schaffner (S. 138) „T_1 deduktiv aus T_2 erhält: wenn man zu T_2 gewisse kontrafaktische Prämissen dazunimmt, die in gewissen experimentellen Kontexten (relativ zum Stand der Wissenschaft) nicht experimentell falsifizierbar wären"? Nein, denn die

[46] Viele für das Verständnis dieses Beispiels nützliche Informationen sowie eine alternative Interpretation als intertheoretische Idealisierung werden unten in Kapitel 8.2 gegeben.

[47] Diese Formulierungen der beiden Gesetze stammen von Eberle (S. 496).

endliche Größe des Erdradius war seit Eratosthenes quantitativ ziemlich genau bestimmt, so daß die Unendlichkeitsannahme schon mehr als 2200 Jahre lang in allen experimentellen Kontexten falsifizierbar ist. Wenn wir aber die *kontrafaktische Gültigkeit* des Galileischen Gesetzes (Wenn $R=\infty$ wäre, dann gälte g=konstant) einmal zugestehen, dann dürfen wir deshalb noch lange nicht — wie Schaffner anzudeuten scheint — auf seine *approximative Gültigkeit* ($g \approx$ konstant) schließen: Man stelle sich nur einmal vor, die Erde wäre in Wirklichkeit ein ganz winziger Planet.

Vergleichen wir dies schließlich mit Eberles Vorschlag, das Galileische Fallgesetz auf Körper mit $h=0$ zu relativieren, und seinem Kommentar dazu (Eberle, S. 498):

> ... the limit, according to Newton's law, of the magnitude of gravitational acceleration *as the height h(b) of a body b approaches zero*, is just what it should be according to Galileo's law. Thus, if *the appropriate term* in Newton's law is replaced by a term denoting a limit whose condition of convergence is just that expressed by the definiens in the definition of π [$\forall x(\pi x \leftrightarrow [x \text{ is a body} \rightarrow h(x)=0])$], then Galileo's law (as it stands) is reducible to this alteration of Newton's theory by a function which represents Galileo's constant k by the limit term.

Hat Eberle mit der Feststellung, daß der Newtonsche Grenzwert $\lim_{h \to 0} g = GM/R^2$ gleich der Galilei-Konstante k ist, wirklich den Punkt des Galileischen Fallgesetzes getroffen? Nein, denn die Existenz irgendeines beliebigen Grenzwerts sichert die Erfüllung des Galilei-Gesetzes, *wenn* man die Ersetzung der h-abhängigen Funktion g durch den Grenzwert $\lim_{h \to 0} g$ als legitim ansieht — aber eben dies ist sehr zweifelhaft. Kann man dann, wie Eberle weiter anregt, für „den zuständigen Term in Newtons Gesetz" einfach einen Term einsetzen, „der einen Grenzwert bezeichnet, dessen Konvergenzbedingung gerade durch die Definition von π ausgedrückt wird"? Auch das wohl nicht, und zwar unabhängig davon, ob mit „dem zuständigen Term" g oder h gemeint ist, denn wenn man in „Newtons Gesetz" g durch GM/R^2 ersetzt, dann hat man kein Gesetz mehr, und wenn man darin die Höhe h durch 0 ersetzt, dann hat man keine fallenden Körper mehr. Eberle (S. 497f) betont ausdrücklich, daß mit den relativierenden Bedingungen (hier speziell mit $h=0$) die „idealen oder kontrafaktischen Bedingungen" angegeben werden, „unter denen T_1 korrekt ist — wenn Korrektheit nach den Maßstäben von T_2 beurteilt wird" (oder speziell hier „unter der Galileis

Theorie nach Newton weiterhin gilt").

Mit dieser letzten Bemerkung — wie auch schon mit Schaffners Bemerkungen zum Galilei-Newton-Fall — könnten wir direkt zur letzten Modellierung der Reduktion von miteinander inkompatiblen Theorien T_1 und T_2, nämlich zur Methode der „kontrafaktischen Annahmen", überleiten. Bevor wir dies tun, wollen wir der Fairness halber aber noch darauf verweisen, daß der Approximationsbegriff im zweiten Sinn zwar offenkundig nicht ganz wohlverstanden ist, daß es in den betrachteten Aufsätzen aber *eine* präzise und allgemeine Formulierung gibt, die uns sagt, was genau gemeint sein könnte. Die Formulierung findet sich unter Arthur Fines (S. 237f) Forderungen für das Beibehalten eines Terms S beim Theorienwechsel von T_1 zu T_2. Er gibt für quantitative Funktionsterme folgendes extensionale Kriterium:

> There are conditions C that can be formulated in T_2, such that
> (a) objects of T_2 that satisfy C are suitable objects for T_1;
> (b) ... if S is a term for a magnitude and v is the value of S applied to object O in T_1 and if v' is the value of S applied to object O in T_2, then to each number $\varepsilon>0$ there corresponds conditions C_ε satisfying (a) such that $|v\text{-}v'|<\varepsilon$ whenever object O satisfies C_ε.

In Teil (b) ist von entscheidender Bedeutung, daß die Randbedingungen C_ε von einem vorgegebenen Genauigkeitsgrad ε abhängen (was den formalen Schönheitsfehler zur Folge hat, daß hier das ‚C' aus Teil (a) gar nicht mehr vorkommt). Etwas frei umformulierend, können wir Fines Idee als Grundlage verwenden zur Explikation des Approximationsbegriffs im zweiten Sinn durch eine Vielzahl von Approximationen im ersten Sinn. Das Schema (2.4.3), das — wie wir sahen — in seiner Anwendung auf Beispielsfälle allen Autoren Schwierigkeiten bereitete, kann durch die Hinzunahme von wechselnden Randbedingungen C_ε expliziert werden:

(2.4.4) $\forall\varepsilon>0\ \exists C_\varepsilon\ (\ T_2, C_\varepsilon \vdash T_1^*$ und $T_1^* \approx_\varepsilon T_1\ )$.

Man wird, um auf die Standardbeispiele zurückzukommen, zu *jedem* ε ein von ε abhängiges δ finden, so daß unter der Zusatzbedingung $v<\delta$ (oder $c>\delta$) aus der speziellen Relativitätstheorie (zeitlich begrenzte) Voraussagen folgen, die ε-nahe an Voraussagen von klassischen Mechanik liegen, und entsprechend ist die Idee für $V>\delta$ (oder $a,b<\delta$) im Fall des van der Waalsschen und idealen Gasgesetzes und für $h<\delta$ (oder $R>\delta$) im Fall des

Newtonschen und Galileischen Fallgesetzes.[48] Der Unterschied zum durch die Einbeziehung von Zusatzbedingungen verfeinerten Schema

(2.4.5) $T_2, C \vdash T_1^*$ und $T_1^* \approx T_1$[49]

der Approximation im ersten Sinn bleibt durchaus gewaltig, denn in (2.4.5) werden *wahre* Randbedingungen eingesetzt, die sich nicht nach einem beliebig (klein) vorgesetzten ε richten können. Im Gegensatz dazu entfernt man sich im Approximationsschema (2.4.4) mehr oder weniger zwangsläufig von der empirischen Realität. Reale Körper bewegen sich nicht beliebig langsam (im Vergleich zur charakteristischen, festen Lichtgeschwindigkeit c) und beliebig nahe an der Oberfläche der Erde (die einen charakteristischen, festen Radius R hat), und reale Gase finden sich nicht in beliebiger Verdünnung (und haben charakteristische, feste van der Waals-Konstanten a>0 und b>0). Die Zusatzbedingungen C_ε werden, wenn man ε nur klein genug macht, irgendwann *kontrafaktisch*. Alle Objekte, die im intendierten Anwendungsbereich von T_2 liegen, fallen einmal aus einem C_ε-Bereich heraus. Wenn nun T_2 als umfassende Theorie über die Welt auch kontingente Aussagen über Anfangswerte oder Randbedingungen enthält, dann wird C_ε irgendwann einmal, bei genügend kleinen ε's, mit T_2 in Konflikt kommen: C_ε ist nach dieser (diesem Buch zugrunde gelegten) Theorienauffassung *kontratheoretisch*. Aus T_2 und C_ε folgt dann aber nach dem *Ex falso quodlibet* Beliebiges, und das Schema (2.4.4) wird schließlich wieder entwertet.

Was also tun? Mit den kontrafaktischen und kontratheoretischen Annahmen C_ε sind wir mitten in der Diskussion des letzten der vier angekündigten Ansätze zur Ausräumung des Inkompatibilitätseinwands gegen die Reduzierbarkeit von T_1 und T_2 angelangt. Wir wollen sehen, inwiefern dieser Ansatz in den einschlägigen Arbeiten angelegt ist und ob er als Lösungsmöglichkeit für die Schwierigkeiten des Approximationsmodells taugt.[50]

[48] Ziemlich genau das Schema (2.4.4) wird von Scheibe (1973) in einer detaillierten Untersuchung auf das Kepler-Newton-Beispiel angewandt. Vgl. hierzu Kapitel 8.1.

[49] Dies ist das aufgeklärte Nagel-Hempel-Schema (2.3.7). Ich wüßte keinen Autor, der nicht ohne Umschweife anerkennen würde, daß Zusatzbedingungen zu T_2 — sei es in der Form von „Brückengesetzen", sei es in der Form von (bereichseinschränkenden) „Anfangs-" oder „Randbedingungen" — für die Ableitung von T_1 immer nötig sind.

[50] Damit soll nicht gesagt sein, daß die Schwierigkeiten des Approximationsmodells prinzipiell nicht zu beseitigen wären, sondern nur, daß ich dies im folgenden nicht versuchen will. In Kapitel 8.1 werden modernere Versionen des Approximationsmodells im dafür wohl geeigneteren strukturalistischen Rahmen wiedergegeben und kritisiert.

Schon Adams (1959) hatte bemerkt, daß sein paradigmatisches Beispiel, die Reduktion der Mechanik starrer Körper auf die Partikelmechanik, genaugenommen nur mit Hilfe kontrafaktischer Annahmen rekonstruierbar ist.[51] An einer (einzigen) Stelle zieht Feyerabend (1965c, S. 273) die folgende überraschende allgemeine Lehre aus der Analyse eines seiner Lieblingsbeispiele:

> when making a comparative evaluation of classical physics and of general relativity we do not compare meanings; we investigate the conditions under which a structural similarity can be obtained. If these conditions are contrary to fact, then the theory that does not contain them supersedes the theory whose structure can be mimicked only if the conditions hold (it is now quite irrelevant in what theory and, therefore, in what terms the conditions are framed).[52]

Schaffner und Eberle sind bei der Betrachtung des Galileischen Fallgesetzes auf kontrafaktische Bedingungen gestoßen: Nach Schaffner wäre es, von Newtons Gravitationstheorie aus gesehen, dann gültig, wenn der Erdradius unendlich wäre, nach Eberle hingegen, wenn der Abstand von der Erdoberfläche stets gleich Null wäre. Glymour (S. 344) verwendet eine grammatikalische Variante eines kontrafaktischen Konditionalsatzes, um die in seinem oben genannten Beispiel enthaltene Approximation zu erklären: „If van der Waals' is true, then the ideal gas law would only hold *were* a system to become arbitrarily dilute" (Hervorhebung von mir). Obgleich diese Erläuterung nicht ohne weiteres verständlich ist, hat Glymour — im Gegensatz zu Adams, Feyerabend, Schaffner und Eberle — doch versucht, kontrafaktische „special assumptions" und kontrafaktische Konditionalsätze ganz explizit und systematisch zur Analyse intertheoretischer Erklärungen in Anschlag zu bringen. Dies kommt zum Beispiel in den folgenden Textstellen Glymours (S. 341f) zum Ausdruck:

---

[51] „... the facts that molecules only approximate point-particles, and that they are not perfectly rigidly fixed within the bodies they compose, shows that the deduction of the laws of RBM from those of PM depends on a hypothesis which, taken exactly, is false." (Adams 1959, S. 264)

[52] Feyerabend (1965c, S. 272f) nennt zwei kontrafaktische Annahmen, die mit der allgemeinen Relativitätstheorie zu „kombinieren" seien: daß die globale Metrik des Raumes „beinahe Minkowskisch" und daß die Lichtgeschwindigkeit „beinahe unendlich" ist. Wegen des — unnötigen — „beinahe" kann man zumindest bei der zweiten Annahme bezweifeln, daß sie wirklich kontrafaktisch ist.

> Inter-theoretical explanation is an exercise in the presentation of counterfactuals. ... Explanation involves a contrast, usually implicit, between the contrary-to-fact special assumptions which do entail laws isomorphic to those of the secondary theory, and the true special assumptions for various cases — which generally entail the negations of the isomorphs of the laws of the secondary theory.

Glymour selbst hat diese Ideen auf eine Handvoll von Beispielen angewandt. Seine Betrachtungen mußten aber — wie jede en passant-Analyse von historisch realisiertem Theorienwandel — oberflächlich bleiben. Der Grund hierfür liegt nicht allein in der allzu knappen Darstellung der betrachteten konkreten Theorien und der fehlenden Rationalisierung der speziellen Übergänge von T_1 zu T_2, sondern auch schon auf begrifflich-abstakter Ebene. Wir brauchen Antworten auf die folgenden Fragen: Was ist eine intertheoretische Erklärung? Gibt es eine exakte Interpretation kontrafaktischer Konditionalsätze? Auf welche Weise ist es möglich, kontrafaktische „Spezialannahmen" zu einer umfassenden Theorie T_2 dazuzunehmen, d.h. wie muß man T_2 revidieren, wenn sie T_2 widersprechen? Kann man sich sicher sein, daß das Akzeptieren gewisser kontrafaktische Konditionalsätze in T_2 und das Revidieren von T_2 zum Zwecke der Eingliederung kontrafaktischer Zusatzannahmen auf dasselbe hinauslaufen? Nur zur ersten dieser Fragen nimmt Glymour — informell — Stellung; wir werden später (in Kapitel 7) auf seine Ideen zu sprechen kommen. Aber natürlich hängt der Erfolg der Idee, inkompatible Theorien mittels kontrafaktischer Annahmen und/oder Konditionalsätze in eine intertheoretische Relation zu bringen, von der präzisen und systematischen Beantwortung aller dieser Fragen ab. Wir werden in den Kapiteln 3–6 dieser Arbeit die einschlägigen Probleme ausführlich untersuchen und erst danach, in Kapitel 7, alternative Darstellungen zu den Schemata (2.4.1)–(2.4.5) vorschlagen.

Bevor wir uns aber dieser Aufgabe zuwenden, will ich noch einige Anmerkungen dazu machen, inwiefern sich diese vierte und letze Sichtweise inkonsistenter Nachfolgertheorien von der Approximationsperspektive unterscheidet. Dabei gehe ich auf drei Punkte ein: auf den Begriff der Idealisierung, auf die Frage, was denn eigentlich von einer reduzierenden Theorie T_2 erklärt wird, und zuletzt auf das intuitive Bild, welches der „kontrafaktischen Analyse" inkompatibler, aber in einer Art Reduktionsrelation stehender Theorien zugrunde liegt.

2.4.2 Reduktion, Idealisierung und Erklärung

Bei drei der sechs besprochenen Autoren finden wir Hinweise auf Gesichtspunkte, die ich in den Kapiteln 7–9 unter dem Stichwort „Idealisierung" diskutieren werde. Schaffner (S. 140, C(3), 144, (3)) fordert, daß T_2 erklären sollte, warum T_1 inkorrekt war, und er erwähnt, dies könne beispielsweise dadurch geschehen, daß T_2 auf die Tatsache verweist, daß T_1 eine „entscheidende Variable ignorierte". Anhand eines Beispiels (der Ableitbarkeit einer „korrigierten" Version der Fresnelschen Gleichungen aus der Maxwellschen Theorie des Elektromagnetismus) bemerkt Schaffner (S. 142), daß die Korrektur von T_1 durch T_2 zwar klein, aber bedeutsam sein kann, indem T_2 die bisher unbekannte Abhängigkeit einer Größe von einer anderen offenlegt. Dieser Aspekt der qualitativen Wichtigkeit bei quantitativer Unwichtigkeit wird von der Approximationsperspektive aus nicht deutlich.

In Fortsetzung der Schaffnerschen Andeutungen kommt man nun zu der Erkenntnis, daß die (bewußte oder unbewußte) Vernachlässigung einer „entscheidenden Variablen" in T_1 durchaus nicht nur kleine quantitative Fehler, sondern auch große Abweichungen gegenüber T_2 mit sich bringen kann. Die Theorie T_1 verliert damit keineswegs jeden Wert:[53] Sie sagt weiterhin, was der Fall ist oder wäre, wenn der vernachlässigte Faktor nicht wirksam ist oder wäre. Sie ist damit Auslöser eines eigenen kleinen Forschungsprogramms, welches zum Ziel hat, die genauen Auswirkungen des unterschlagenen Faktors (und sein Zusammenspiel mit anderen Faktoren) zu bestimmen — ein Ziel, das schon mit T_2 erreicht sein kann. Eberle (S. 496f) weist darauf hin, daß vernachlässigte Faktoren in Wirklichkeit oft überhaupt nicht ausgeschaltet werden können, d.h. daß es keine Objekte gibt, die die idealisierenden Voraussetzungen von T_1 erfüllen (Eberles Beispiel ist wieder die „Idealisierung" h=0 für Galileis Fallgesetz): „laws are employed in such contexts which would hold at best vacuously under actual fully described conditions." (S. 497) Diese Sichtweise steht im Kon-

[53] Vgl. dagegen Sklar (S. 117): „For what is it that makes us want to speak of some theoretical replacements as reductions ... and to speak of other replacements ... as the mere discarding of one theory in favour of a better? It is only the survival of the older theory, in the reductive cases, as a ‚useful instrument for prediction', despite its known falsity as a scientific theory. If reference to caloric, its presence and absence, its rate of flow, the capacity of bodies to absorb it, etc., were as useful to engineers — *despite the fact that there is no caloric* — as is the hydrodynamic theory to boat designers — *despite the fact that there are no microscopically continuous fluid media* — there is little doubt that we would speak of the reduction of the caloric theory to the energetic, rather than welcome its overthrow as we are now inclined to do." (Hervorhebungen von mir.)

trast zum Approximationsansatz, nach dem die Gesetze von T_1 — aus der Warte von T_2 beurteilt — im allgemeinen „at best approximations" (Sklar, S. 111) sind. Schließlich kann ich ohne weiteren Kommentar meine vorbehaltlose Zustimmung zu Nickles (S. 200) äußern, wenn er schreibt: „Not that a theory must be entirely successful to be a worthy precedent: it may be a mere ‚model,' deliberately oversimplified — the minimal realization of a promising theory program." (Nickles' Beispiele sind frühe Modelle von Festkörpern und Gasen in der statistischen Mechanik und ihre schrittweise Verbesserung durch die Einführung neuer Freiheitsgrade; siehe auch Nickles, S. 185, Fußnote 5.)

Ein zweiter interessanter Punkt ist das Verhältnis zwischen Reduktion und Erklärung im Falle einer mit der reduzierten Theorie inkonsistenten reduzierenden Theorie. Aus der Approximationsperspektive erscheint es klar, daß es die reduzierte Theorie T_1 ist, die — wenn auch bloß approximativ — von der reduzierenden Theorie T_2 erklärt wird. Daß mit einer Reduktion von T_1 auf T_2 auch eine Erklärung von T_1 durch T_2 vorliegt, ja sogar, daß „Reduktion" in etwa das Gleiche bedeutet wie „intertheoretische Erklärung", kann man auch als die Ansicht von Schaffner (S. 137) und Glymour (z.B. S. 340f, 352) betrachten. Aber das ist nicht die ganze Geschichte. Nach Glymour (S. 344) soll beispielsweise das Gasgesetz von van der Waals „erklären, warum das ideale Gasgesetz funktioniert, wo es funktioniert, und scheitert, wo es scheitert", und eine molekulare Theorie der Materie „sollte erklären, warum das Gesetz der korrespondierenden Zustände so gut funktioniert und wo seine Grenzen sind." Eine ähnliche Zweischneidigkeit der Erklärung bedingt sich Schaffner (S. 140, C(3), 144, (3)) aus; T_2 soll über das mittels (2.4.2) gewonnene $T_1{}^*$ anzeigen, warum T_1 inkorrekt war (z.B. weil es eine entscheidende Variable ignorierte) und warum es so gut funktionierte, wie es funktionierte." Gemäß diesen Bestimmungen erklärt T_2 also gewissermaßen beides: sowohl T_1 als auch das Scheitern von T_1. Noch etwas weiter von der Standardauffassung entfernt sich Nickles (S. 185). In Fußnote 4 sagt er zwar, daß approximative Reduktionen sozusagen erklären, warum die Vorgängertheorie so gut funktionierte, wie sie funktionierte; im dazugehörigen Haupttext jedoch wird klar, daß es sich hier nach Nickles' Auffassung gar nicht wirklich um eine („theoretische") Erklärung handelt: „Not all reduction is explanation!" Die meines Erachtens treffendste Charakterisierung der Erklärungsverhältnisse bei Reduktionen zwischen inkompatiblen T_1 und T_2 stammt von Sklar (S. 112). Da T_1 im Lichte von T_2 inkorrekt ist, so argumentiert er, kann T_1 gar nicht von T_2 erklärt werden. Was T_2 tatsächlich erklärt, ist, warum

T_1 korrekt gewesen zu sein schien oder warum es so großen scheinbaren Erfolg hatte.[54] Mit Sklars schönen Worten heißt das auch, daß T_2 T_1 nicht erklärt, sondern *wegerklärt*. Wieder anders gewendet: T_2 erklärt in Wirklichkeit nicht T_1, sondern im Gegenteil die Falschheit oder das Scheitern von T_1. An zwei Stellen verfällt Glymour ebenfalls in diese Redeweise,[55] und auch Nickles (S. 183) darf man in einem ähnlichen Sinn verstehen, wenn er schreibt, daß die Einsteinsche Theorie, auf welche die Newtonsche in paradigmatischer Weise reduzierbar ist, „gar nicht hätte erfolgreich sein können, ohne gleichzeitig ihre Vorgängerin zum Scheitern zu bringen." Diese letzten Zitate erscheinen mir als die genauesten, wenn auch immer noch ergänzungsfähigen und -bedürftigen Formulierungen. Statt wie die Vertreter des Approximationsansatzes zu fordern, daß T_2 T_1 approximativ erklären soll, werde ich im Verlauf dieser Arbeit (in Kapitel 7) eine Bedingung vorschlagen und präzisieren, nach der T_2 T_1 „kontrafaktisch erklärt" (während das, was es „faktisch" erklärt, das Scheitern von T_1 ist).

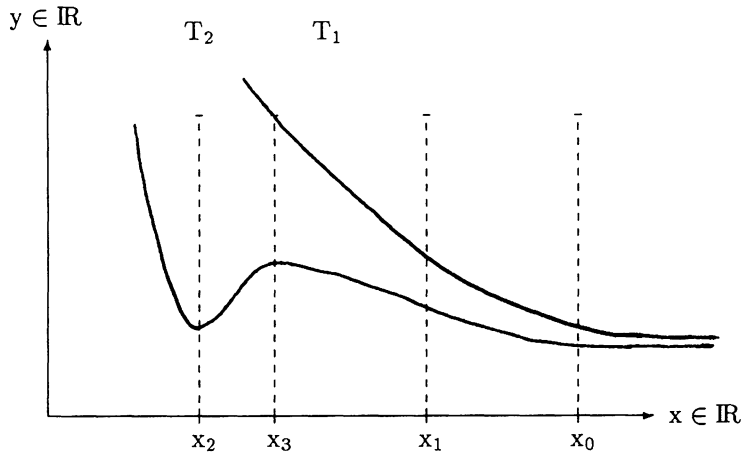
Zum Schluß dieses Kapitels über das klassische Konzept der Reduktion erscheint es mir nützlich, die Grundintuitionen hinter den hier unterschiedenen vier Ansätzen zur Reduzierbarmachung inkonsistenter Theorien noch einmal anhand eines Bildes zu veranschaulichen. Man betrachte folgendes typische Diagramm[56]

---

[54] Zur Antwort auf die naheliegende Frage, warum man sich nicht einfach Kriterien wie ($K4^{(s')}$) oder (1.9.5) zu eigen machen und sagen kann, T_2 erkläre *den Erfolg von* T_1 vgl. das Ende von Abschnitt 1.9 und den jetzt folgenden Absatz.

[55] Und zwar auf S. 342: „*That the secondary theory is generally false* is explained by the fact that the various parameters do not usually — and sometimes never — have their limiting values." und auf S. 345: „In the explanation of *why Galileo's law fails* one is ... implicitly contrasting a contrary-to-fact situation in which Galileo's law would hold with the real situation, in which Newton's laws entail the denial of Galileo's law". (Hervorhebungen von mir)

[56] Die Gestalt des Diagramms ist nicht fiktiv. Wenn die Variable x das Molvolumen eines Gases und die Variable y den Druck bezeichnet, dann beschreiben die Kurven von T_1 und T_2 in etwa die Isothermen von Kohlendioxid für 0°C gemäß dem idealen bzw. van der Waalsschen Gasgesetz.



SKIZZE 2.1

Die durch ihre Graphen repräsentierten quantitativen Theorien T_1 und T_2 widersprechen sich in dem Sinn, daß sie allen oder jedenfalls den meisten x-Werten verscheidene y-Werte zuordnen.[57] Zunächst einmal wird der Ansatz der Bereichseinschränkung zu zeigen versuchen, daß aus T_2 zum Beispiel $T_1|_{\{x:x>x_0\}}$ folgt. Aber genaugenommen ist der Bereich $x>x_0$ nicht viel anders geartet als andere Bereiche: T_1 und T_2 machen hier zwar sehr ähnliche, eventuell empirisch (heute noch) ununterscheidbare, aber eben doch verschiedene Aussagen. Nun kommt der Approximationsansatz im ersten Sinn. Dort muß ebenfalls eine Bereichseinschränkung vorgenommen werden, man ist aber, da man von vornherein nur auf eine ungefähre Übereinstimmung Wert legt, freier in der Abgrenzung dieses Geltungsbereiches und kann im allgemeinen einen besonders natürlichen oder ausgezeichneten Schwellenwert x_1 auswählen. Während der Approximationist im ersten Sinn also behauptet, daß im Bereich der Objekte oder Systeme mit $x>x_1$ T_1 annähernd aus T_2 ableitbar sei, interessieren die tatsächlichen x-Werte irgendwelcher Objekte oder Systeme den Vertreter des Approximationsansatzes im zweiten Sinn überhaupt nicht mehr. Er betont allein die Tatsache, daß für jedes $\varepsilon>0$ ein x_ε derart existiert, daß sich für alle $x>x_\varepsilon$ die y-Werte

[57] Dies muß nicht heißen, daß T_1 und T_2 streng logisch inkonsistent sind; die Kurven könnten sich z.B. für ein $x_4>x_0$ überschneiden und damit Systeme mit den Werten $\langle x_4, y_4\rangle$ als gemeinsame Modelle haben. Vgl. die Bemerkungen zu „Karussellmodellen" und zur „Idealkurve" in Kapitel 8.1 bzw. 8.2.

gemäß T_1 und T_2 um höchstens ε unterscheiden. Wer sich schließlich den kontrafaktischen Ansatz zu eigen macht, der betrachtet das Diagramm aus einem ganz anderen Blickwinkel. Er stellt zunächst einmal fest, daß sich die Kurven von T_1 und T_2 deutlich unterscheiden, wobei er insbesondere auch ein Auge auf die qualitative Verschiedenheit von T_1 und T_2 wirft, die hier darin besteht, daß die Kurve von T_2 (zwischen x_2 und x_3) ein Stück mit positiver Ableitung besitzt. Das Auseinanderdriften von T_1 und T_2, welches nur für große x klein (aber immer noch spürbar) und für kleine x groß ist, verlangt nach einer Erklärung. Eine besonders überzeugende Erklärung besteht im Aufweis durch T_2, daß T_1 einen oder mehrere der tatsächlich wirksamen Faktoren vernachlässigt, d.h. daß T_1 gültig *wäre*, *wenn* dieser Faktor keine Wirkung *hätte*. *Weil* er aber in Wirklichkeit eine solche hat (und weil dieses Wissen Bestandteil der umfassenden Theorie T_2 ist), deshalb erklärt T_2 eigentlich nicht T_1, sondern das Scheitern von T_1 (vorausgesetzt, man nimmt T_2 auch zum Maßstab der korrekten Wirklichkeitsbeschreibung).

Diese letzte Sichtweise ist zwar ungewöhnlich, mir kommt sie aber dennoch sehr vielversprechend vor. Jedenfalls ist ihr unter den vier betrachteten bisher am wenigsten wissenschaftstheoretische Aufmerksamkeit zuteil geworden. Dafür sehe ich zwei forschungsgeschichtliche Gründe. Zum einen ist die Kritik an der traditionell-empiristischen Wissenschaftstheorie und ihrem Bild eines monotonen Theorienwandels von Kuhn wohl noch überzeugender als von Feyerabend vorgetragen worden. Kuhn gründet seine Inkommensurabilitätsthese jedoch nicht auf die Inkonsistenzthese. Deshalb wurde in den Reaktionen auf Kuhn die in diesem Kapitel skizzierte Zurückspielung von Inkommensurabilität auf Inkonsistenz gar nicht erst versucht. Wie wir wissen, ist das Aufblühen der strukturalistischen Wissenschaftstheorie nicht zuletzt eine Folge der von Sneed und Stegmüller vorgelegten Kuhn-Rekonstuktion, und innerhalb dieser Schule liegt der approximative Ansatz (vor allem im zweiten Sinn) doch am nächsten. Zum anderen — und dies ist womöglich noch wichtiger — wurde die Idee des kontrafaktischen Ansatzes allein von Glymour ganz explizit formuliert, und dies zu einer Zeit, als die moderne Analyse kontrafaktischer Konditionalsätze noch in ihren Kinderschuhen steckte. Insofern ist es vielleicht verständlich, daß dieser Ansatz stecken geblieben ist. Doch heute ist die Situation hinsichtlich der formalen Beherrschung der Kontrafaktizität eine völlig andere. Sehen wir nun also, ob wir den kontrafaktischen Ansatz wieder in Schwung bringen können, und gehen zunächst an die Aufgabe, die abstrakten Grundlagen für seine Präzisierung zu erkunden. In Kapitel 3 wenden wir uns der Frage

zu, was es heißt, eine Theorie durch kontrafaktische Annahmen zu revidieren. Erst nachdem diese Formalia klargelegt sind, wollen wir in Kapitel 4 die Untersuchung kontrafaktischer Konditionalsätze beginnen und uns insbesondere über ihren (wie sich herausstellt, nichttrivialen) Zusammenhang mit kontrafaktischen Revisionen Gedanken machen.

Kapitel 3

Kontrafaktische Annahmen — Zum Modell der Theorienrevision nach Gärdenfors

3.1 Theorienrevisionen und -kontraktionen: die Gärdenfors-Postulate

Für die folgenden abstrakten, von wissenschaftstheoretischen Erfordernissen vorderhand unabhängigen Überlegungen dürfen wir ohne Skrupel den präzisen, aber auch idealisierenden Theorienbegriff des Logikers verwenden. Eine *Theorie* T ist demnach eine bezüglich einer gegebenen Logik abgeschlossene Menge von Formeln einer gegebenen Sprache. Wir wollen hier lediglich fordern, daß unsere Sprache die üblichen Junktoren $\neg$, $\wedge$, $\vee$, $\rightarrow$ und $\leftrightarrow$ der Aussagenlogik enthält, sowie die Satzkonstanten $\top$ („Verum", „Wahrheit") und $\bot$ („Falsum", „Falschheit"). Wenn wir eine Logik mit einer Konsequenzoperation Cn identifizieren, die einer beliebigen Formel-

menge S die Menge Cn(S) ihrer logischen Konsequenzen zuordnet,[1] so sind Theorien Fixpunkte von Cn:

3.1.1. Definition Eine Formelmenge T ist genau dann eine *Theorie*, wenn gilt: $T = Cn(T)$.

Statt $A \in Cn(T)$ für eine Formelmenge T und eine Formel A schreiben wir auch $T \vdash A$. Wir werden im folgenden präsupponieren, daß Cn die tautologische Folgerung der klassischen Aussagenlogik enthält, kompakt ist (d.h. falls $T \vdash A$, so existiert endliches $T_0 \subseteq T$ mit $T_0 \vdash A$) und die ebenfalls klassische Eigenschaft der Disjunktion der Prämissen (d.h. falls $T \cup \{B\} \vdash A$ und $T \cup \{C\} \vdash A$, so auch $T \cup \{B \vee C\} \vdash A$) besitzt. Das Deduktionstheorem für Cn folgt aus dem Enthalten der tautologischen Folgerung und der Disjunktion der Prämissen (vgl. Alchourrón und Makinson 1982, S. 18). $Cn(\emptyset) = Cn(\{\top\})$ ist die Menge aller logisch wahren Formeln, $Cn(\{\bot\})$ ist die Menge aller Formeln und wird i.f. als „die inkonsistente Theorie" $T_\bot$ bezeichnet.

Wie kann man eine solche Theorie durch kontrafaktische Annahmen revidieren? Zunächst sei noch einmal dargestellt, daß hier ein wirkliches Problem vorliegt. Wir legen in diesem Buch eine holistische Theorienauffassung zugrunde, nach der eine Theorie die Gesamtheit der Überzeugungen eines (stark idealisierten) epistemischen Subjekts, z.B. eines Wissenschaftlers, über die Welt ist. Eine Theorie ist, allgemein gesprochen, die „Wissensbasis", auf deren Grundlage ein Subjekt seine Planungen und Handlungen ausrichtet. Mithin ist eine Theorie nicht nur eine Ansammlung von „gesetzesartigen Aussagen" — zumal wir nicht von vornherein wissen, welche Aussagen denn als gesetzesartig gelten dürfen. (Später, auf der Grundlage eines detailliert entwickelten Revisionsmodells, werden wir „Gesetzesartigkeit" partiell durch „hohe theoretische Wichtigkeit" explizieren können.) Nachdem epistemische Subjekte aber nur die je momentan akzeptierte Theorie T als Maßstab für ihre Urteile über Wahrheit („Faktizität") und Falschheit („Kontrafaktizität") von Sätzen zugrundelegen können, wird klar, daß in diesem Kontext „kontrafaktisch" gleichbedeutend ist mit „kontratheoretisch". Eine Annahme A heißt (bzgl. T) genau dann *kontrafaktisch* (oder *kontratheoretisch*), wenn $\neg A$ in T enthal-

[1] Diese abstrakte Auffassung einer Logik geht zurück auf Tarski (1930). Normalerweise verlangt man zumindest, daß Cn ein Hüllenoperator ist, d.h. die Axiome (H1) $S \subseteq Cn(S)$, (H2) $S_1 \subseteq S_2 \Rightarrow Cn(S_1) \subseteq Cn(S_2)$ und (H3) $Cn(Cn(S)) = Cn(S)$ erfüllt. In den Kapiteln 4 und 5 werden wir jedoch Sprachen betrachten, für die das Axiom (H2) fragwürdig bzw. ungültig ist. Zur Theorie der Konsequenzrelationen vgl. Wójcicki (1988) und Makinson (1989a).

ten ist. Um eine kontrafaktische Annahme A hypothetisch — oder um eine ursprünglich für kontrafaktisch gehaltene Neuinformation A tatsächlich — ohne Widersprüche in die aktuelle Theorie T einzuverleiben, muß in T eingegriffen werden: Es müssen einige Sätze aus T entfernt werden, so daß die Ableitbarkeit von $\neg A$ verhindert wird. Das Ziel wird es hierbei sein, nur *minimale Änderungen* an T vorzunehmen, da man ja nichts unnötig vom Gehalt von T preisgeben will. Der Übergang zur *Expansion* von T durch A, definiert durch

3.1.2. Definition Die *Expansion* $T^+{}_A$ einer Theorie T durch einen Satz A ist die Menge der logischen Konsequenzen aus $T \cup \{A\}$, d.h. $T^+{}_A := Cn(T \cup \{A\}) = \{B: A \rightarrow B \in T\}$.[2]

ist im kontrafaktischen Fall offenbar nicht ratsam, denn das *Ex falso quodlibet* liefert sofort die inkonsistente Theorie $T_\perp$.

Wir müssen also versuchen herauszufinden, wie man zu einer vernünftigen *Revision* von T durch ein kontrafaktische Annahme A, in Zeichen: $T^*{}_A$, kommt.[3] Ein erster und wichtiger Schritt hierzu ist das Aufstellen von minimalen Bedingungen oder Rationalitätspostulaten, denen eine Revision genügen soll. Als erstes muß man sich vergegenwärtigen, *woran* solche Postulate eigentlich Anforderungen stellen sollen. Um möglichst flexibel zu sein, wollen wir nicht von vornherein davon ausgehen, daß alle logisch möglichen Theorien als akzeptierte Theorien in Frage kommen. Sei also **T** eine beliebige Menge von Theorien. Wir gehen dann von dem Begriff einer *Revisionsfunktion* * aus, die jeder Theorie T aus **T** und jedem[4] Satz A eine durch A revidierte Theorie *(T,A) — suggestiver als $T^*{}_A$ geschrieben — aus **T** zuordnet. Ein soches Paar $\langle \mathbf{T}, * \rangle$ nennen wir *Theorienrevisionsmodell*.[5] Wenn wir im weiteren etwas lax von Bedingungen an Revisionen

[2] Das letzte Gleichheitszeichen setzt voraus, daß Cn die Schlußregel des Modus Ponens beinhaltet und das Deduktionstheorem erfüllt. Der Begriff „Expansion von T durch A" rechtfertigt sich natürlich durch die Beziehungen $T \subseteq T^+{}_A$ und $A \in T^+{}_A$.

[3] Der Revisionsbegriff soll aber so weit gefaßt sein, daß auch nicht-kontrafaktische Annahmen erfaßt werden können. Im Rahmen dieses Kapitels geschieht das, wie aus dem folgenden klar wird, einfach durch die Verwendung von Expansionen. Daß dies für reichhaltige Sprachen aber problematisch werden kann, werden wir in den nächsten beiden Kapiteln sehen.

[4] Für manche Sprachen kann es jedoch sinnvoll oder erforderlich sein, den Bereich der Sätze, bezüglich derer man revidieren kann, einzuschränken. Vgl. Kapitel 4.

[5] Gärdenfors verwendet in seinen Arbeiten die Bezeichnungen „(belief) model" (1979), „belief change model" (1986; 1987) und „belief revision system" (1988, Kapitel 7). Von „Theorien" reden in diesem Zusammenhang Alchourrón und Makinson (1982), Alchourrón, Gärdenfors und Makinson (1985), Gärdenfors (1985), Makinson (1985; 1987), Grove (1988), Fuhrmann (1989) und, in einigen bislang unveröffentlichten Arbeiten,

sprechen, meinen wir also genaugenommen Bedingungen an Theorienrevisionsmodelle (und analog für Kontraktionen). Diese Präzisierung erscheint zunächst spitzfindig, wird aber im nächsten Kapitel wichtig werden.

Die folgende Kollektion von Bedingungen für Revisionen (genauer: für Revisionsmodelle $\langle \mathbf{T}, * \rangle$) stammt von Peter Gärdenfors (1979; 1982; 1988, Kapitel 3.3): Für alle Theorien T (aus $\mathbf{T}$) und alle Sätze A und B soll gelten

(T*1) T^*_A ist eine Theorie ;[6]
(T*2) $A \in T^*_A$;
(T*3) $T^*_A \subseteq T^+_A$;
(T*4) $\neg A \notin T \Rightarrow T^+_A \subseteq T^*_A$;
(T*5) $T^*_A = T_\perp \Rightarrow \vdash \neg A$;
(T*6) $\vdash A \leftrightarrow B \Rightarrow T^*_A = T^*_B$;
(T*7) $T^*_{A \wedge B} \subseteq (T^*_A)^+_B$;
(T*8) $\neg B \notin T^*_A \Rightarrow (T^*_A)^+_B \subseteq T^*_{A \wedge B}$.

Es würde zu weit führen, hier auf die Motivation dieser *Gärdenfors-Postulate für Revisionen* einzugehen; hierfür vergleiche man die angegebenen Arbeiten von Gärdenfors. Die Bedingungen (T*1) bis (T*6) heißen die *grundlegenden Postulate*. Für manche Zwecke ist es günstiger, anstelle der weitergehenden Postulate (T*7) und (T*8) eine Faktorisierungsbedingung für Revisionen durch Disjunktionen zu verwenden:

(T*7∧8) $T^*_{A \vee B} = T^*_A$ oder $T^*_{A \vee B} = T^*_B$ oder $T^*_{A \vee B} = T^*_A \cap T^*_B$.

Es gilt das

3.1.3. Lemma Erfülle * die grundlegenden Gärdenfors-Postulate (T*1)–(T*6). Dann erfüllt * genau dann (T*7∧8), wenn * sowohl (T*7) als auch (T*8) erfüllt.

(Die Beweise aller neuen Resultate dieses Kapitels sind als Anhang in Abschnitt 3.7 zusammengestellt.) (T*7∧8) kann man sogar dahingehend präzisieren, daß $T^*_{A \vee B}$ gleich T^*_A ist, falls $\neg B$ in $T^*_{A \vee B}$ ist, daß es gleich T^*_B ist, falls $\neg A$ in $T^*_{A \vee B}$ ist, und daß es ansonsten gleich $T^*_A \cap T^*_B$ ist.[7] Auch diese verstärkte Version von (T*7∧8) ist, wie aus dem Beweis

---

Segerberg.

[6] Dies folgt natürlich schon aus dem Begriff der Revisionsfunktion. Es wird hier wiederholt, um die Kontinuität zu den Gärdenforsschen Präsentationen, die nicht immer von Theorienrevisionsmodellen ausgehen, herzustellen.

[7] Schon aus den Lemmata 7 und 8 in Gärdenfors (1979, S. 393) geht hervor, daß (T*7) mit der Bedingung $T^*_A \cap T^*_B \subseteq T^*_{A \vee B}$ relativ äquivalent ist. Man überlege sich, daß auch die relative Äquivalenz von (T*8) mit der Bedingung $\neg A \notin T^*_{A \vee B} \Rightarrow T^*_{A \vee B} \subseteq T^*_A$ gilt. Vgl. Gärdenfors (1988), Bedingung (3.15).

des Lemmas ersichtlich, relativ zu (T*1)–(T*6) mit der Konjunktion von (T*7) und (T*8) äquivalent.

Eine weitere interessante Eigenschaft der kompletten Kollektion der Gärdenfors-Postulate ist die, daß man aus ihr eine hinreichende Bedingung der Identität von Revisionen durch verschiedene Sätze A und B erhält:

(T*I) $A \in T^*_B \wedge B \in T^*_A \Rightarrow T^*_A = T^*_B$.

Wegen (T*7) und (T*8) folgt aus dem Antezedens nämlich $T^*_A=(T^*_A)^+_B =T^*_{A\wedge B}=(T^*_B)^+_A=T^*_B$ (s. Gärdenfors 1982, S. 97; 1988, S. 57).

Die Gärdenfors-Postulate liefern nun zwar eine Anzahl guter Anhaltspunkte dafür, wie eine Revision auszusehen hat, aber natürlich noch kein Verfahren zur Konstruktion einer solchen. Eine genauere Vorstellung kann man sich machen, wenn man auf eine Idee von Isaac Levi (1977) zurückkommt. Nach Levi erhält man eine Revision T^*_A in zwei Schritten: zuerst eliminiert man $\neg A$ aus T und erhält eine *Kontraktion* $T^-_{\neg A}$; dann erweitert man diese Kontraktion um A, was jetzt konsistent möglich ist. Dies ist die *Levi-Identität*:

(L) $T^*_A = (T^-_{\neg A})^+_A$.

Die durch (L) aus einer Kontraktionsfunktion $^-$ gewonnene Revisionsfunktion * bezeichnen wir im folgenden mit $\mathbf{L}(^-)$. Levi läßt aus (nicht immer völlig durchsichtigen) philosophischen Gründen nur auf solche Weise dekomponierbare Revisionen gelten. Wir wollen uns im folgenden nur um die formale Seite von (L) kümmern. Man kann nun ziemlich leicht sehen, daß sich für Kontraktionen sehr ähnliche Probleme auftun wie für Revisionen (vgl. Gärdenfors 1982; 1988, Kapitel 3.4; Makinson 1985) und daß man auch für Kontraktionen vorerst nichts Besseres tun kann, als einschlägige Rationalitätspostulate aufzustellen:

(T–1) T^-_A ist eine Theorie ;
(T–2) $T^-_A \subseteq T$;
(T–3) $A \notin T \Rightarrow T^-_A = T$;
(T–4) $\nvdash A \Rightarrow A \notin T^-_A$;
(T–5) $T \subseteq (T^-_A)^+_A$;
(T–6) $\vdash A \leftrightarrow B \Rightarrow T^-_A = T^-_B$;
(T–7) $T^-_A \cap T^-_B \subseteq T^-_{A\wedge B}$;
(T–8) $A \notin T^-_{A\wedge B} \Rightarrow T^-_{A\wedge B} \subseteq T^-_A$.

Für die Motivation der *Gärdenfors-Postulate für Kontraktionen* vergleiche die angegebenen Arbeiten von Gärdenfors. Die Bedingungen (T–1) bis (T–6) heißen wieder die *grundlegenden Postulate*. Wie Makinson (1987) diskutiert hat, ist das „Wiedergewinnungspostulat“ (T–5) vielleicht das

problematischste dieser Postulate und hat einen gewissen Sonderstatus.[8] Ganz ähnlich wie bei Revisionen können die weitergehenden Postulate (T–7) und (T–8),[9] wie in Alchourrón, Gärdenfors und Makinson (1985, Observation 6.5) gezeigt wird, auch zu einer Faktorisierungsbedingung für Kontraktionen bezüglich Konjunktionen zusammengefaßt werden:

(T–7∧8) $T^-_{A\wedge B} = T^-_A$ oder $T^-_{A\wedge B} = T^-_B$ oder
$T^-_{A\wedge B} = T^-_A \cap T^-_B$.

(T–7∧8) ist, sofern die grundlegenden Postulate erfüllt sind, mit der Konjunktion von (T–7) und (T–8) äquivalent.

Es gibt auch eine hinreichende Bedingung der Identität von Kontraktionen bezüglich verschiedener Sätze A und B:

3.1.4. Lemma Erfülle $^-$ die Gärdenfors-Postulate (T–1)–(T–8). Dann gilt:

(a) $^-$ erfüllt auch

(T–I) $\quad A\rightarrow B \in T^-_B \wedge B\rightarrow A \in T^-_A \Rightarrow T^-_A = T^-_B$;

(b) wenn A und B in einer Theorie T sind, ist $T^-_A = T^-_B$ genau dann, wenn $A\leftrightarrow B \in T^-_A \cap T^-_B$.

Der Beweis von Teil (a) dieses Lemmas ist viel aufwendiger als der Beweis der „entsprechenden" Bedingung (T*I) unter der Voraussetzung von (T*1)–(T*8).[10]

Bei soviel Symmetrie zwischen Revisionen und Kontraktionen stellt sich die Frage, ob man nicht auch die umgekehrte Richtung von (L) beschreiten und Kontraktionen durch Revisionen definieren kann. Gärdenfors (1982; 1988, Kapitel 3.6) hat gezeigt, daß dies sinnvoll möglich ist:

(H) $\quad T^-_A = T \cap T^*_{\neg A}$.

Die durch (H) aus einer Revisionsfunktion * gewonnene Kontraktionsfunktion $^-$ bezeichnen wir im folgenden mit $\mathbf{H}(^*)$. Diese Gleichung wird von

[8] Vor allem Isaac Levi hat gegen (T–5) argumentiert. Ohne hier Gründe angeben zu können, möchte ich an dieser Stelle zum Ausdruck bringen, daß mir (T–5) intuitiv für hypothetische Revisionen als sehr zweifelhaft, aber für Revisionen aufgrund von faktischer Neuinformation als durchaus wohlbegründet erscheint. Die formale Rechtfertigung von (T–5) durch die unten angegebene Harper-Identität (H) hat etwas Triviales an sich: $A\rightarrow B$ ist für ein B in T genau deshalb in T^-A, weil B in T und $\neg A$ in $T^*_{\neg A}$ ist und sowohl T als auch $T^*_{\neg A}$ Theorien sind.

[9] Die den „entsprechenden" Postulaten (T*7) und (T*8) für Revisionen allerdings nicht sehr ähnlich sind. Vgl. Gärdenfors (1982, S. 98), wo ähnlichere, aber schwerer verständliche Postulate angegeben sind. Vgl. Fußnote 7.

[10] (T*7∧8) und (T–I) wurden inzwischen als Prinzipien (3.16) und (3.28) in Gärdenfors (1988) aufgenommen. Gärdenfors' kurzer Beweis von (3.28) läuft über Revisionen, verwendet die Levi- und die Harper-Identität (s.u.) und Theorem 3.1.5. Vgl. Gärdenfors (1988, S. 216 und 244, Fußnote 10).

Gärdenfors (1988, S. 70) nach Harper (1977) die *Harper-Identität* und von Alchourrón und Makinson (1982, S. 27) und Makinson (1987, S. 389) die *Gärdenfors-Identität* genannt. (H) kann durch die Levi-Identität und den Hinweis darauf motiviert werden, daß aus den grundlegenden Postulaten für Kontraktionen $T^-_A = T \cap (T^-_A)^+_{-A}$ folgt.[11] Wie gut die Postulate für Revisionen und Kontraktionen zusammenpassen, wird an folgendem Theorem deutlich:

3.1.5. Theorem (Gärdenfors) (a) Sei $^-$ eine Kontraktionsfunktion. Dann gilt: Erfüllt $^-$ die grundlegenden Postulate für Kontraktionen, so erfüllt $\mathbf{L}(^-)$ die grundlegenden Postulate für Revisionen; erfüllt $^-$ außerdem (T-7), so erfüllt $\mathbf{L}(^-)$ auch (T*7); erfüllt $^-$ außerdem (T-8), so erfüllt $\mathbf{L}(^-)$ auch (T*8).
(b) Sei * eine Revisionsfunktion. Dann gilt: Erfüllt * die grundlegenden Postulate für Revisionen, so erfüllt $\mathbf{H}(*)$ die grundlegenden Postulate für Kontraktionen; erfüllt * außerdem (T*7), so erfüllt $\mathbf{H}(*)$ auch (T-7); erfüllt * außerdem (T*8), so erfüllt $\mathbf{H}(*)$ auch (T-8).

Für den Beweis dieses Theorems siehe Gärdenfors (1982; 1988, S. 215f) oder Gärdenfors und Makinson (1988). Theorem 3.1.5 liefert eine starke Rechtfertigung sowohl der beiden Kollektionen von Gärdenfors-Postulaten als auch der Levi- und Harper-Gärdenfors-Identitäten. Schließlich folgt, wie leicht nachzurechnen, aus den grundlegenden Postulaten für Revisionen, daß $\mathbf{H}(\mathbf{L}(^-))=^-$ ist, und aus den grundlegenden Postulaten für Kontraktionen, daß $\mathbf{L}(\mathbf{H}(*))=*$ ist, womit das Ganze so perfekt zusammenstimmt, wie man es sich nur wünschen kann.

Eine allgemeine Beziehung, von der wir in Kapitel 6 Gebrauch machen werden, formuliert

3.1.6. Lemma Erfülle $^-$ die grundlegenden Gärdenfors-Postulate (T-1)-(T-6) und sei $*=\mathbf{L}(^-)$. Sei weiter $T \neq T_\perp$ und A so, daß $\nvdash A$ und $\nvdash \neg A$. Dann gilt:
$T^*_A \cap T^*_{\neg A} = T^-_A \cap T^-_{\neg A} \subseteq T = T^-_A \cup T^-_{\neg A} \subseteq T^*_A \cup T^*_{\neg A}$.

Derjenige Teil von Lemma 3.1.6, der später (in Kapitel 6) am wichtigsten sein wird, ist die Tatsache, daß aus $B \in T^*_A$ und $B \in T^*_{\neg A}$ schon $B \in T$ folgt.

[11] Genauer gilt: Unter der Voraussetzung von (T-1), (T-2), (T-4), (T-6) und (L) ist (H) äquivalent mit (T-5); siehe Gärdenfors (1982, S. 93f). Umgekehrt braucht man für die entsprechende Motivation der Levi-Identität durch die Harper-Identität nur (T*2) und (T*3): $T^*_A = T^+_A \cap T^*_A$ (nach (T*3)) $= T^+_A \cap (T^*_A)^+_A$ (nach (T*2)) $= (T \cap T^*_A)^+_A$.

3.2 Die Relation der theoretischen Wichtigkeit

So schön die Postulate für Revisionen und ihre Zusammenhänge mit den Postulaten für Kontraktionen auch sind, sie reichen im allgemeinen nicht hin, zu gegebenem T und A eine eindeutige Revision T^*_A festzulegen. Da man die Gärdenfors-Postulate nicht mehr gut um weitere Postulate erweitern kann, ohne unsere Intuitionen allzusehr zu strapazieren, braucht man zusätzliche Informationen, um aus T und A ein eindeutiges T^*_A konstruieren zu können. Der Vorschlag von Gärdenfors ist es nun, daß diese Zusatzinformationen von einer Ordnungsrelation[12] $\leq$ auf den in T enthaltenen Sätzen geliefert wird. In Gärdenfors (1984) und (1985) nennt er diese Relation *epistemische Wichtigkeit* („epistemic importance"), in Gärdenfors (1988) und Gärdenfors und Makinson (1988) heißt sie *epistemische Befestigung* („epistemic entrenchment"[13]). Im Kontext dieses Buchs kommt mir die Bezeichnung *theoretische Wichtigkeit* am treffendsten vor. Denn ich will für folgende These plädieren: Wenn ein Wissenschaftler eine Theorie vertritt, dann akzeptiert er nicht nur eine Satzmenge T, sondern schon gleich eine Satzmenge *mit einer aufgeprägten Ordnungsstruktur*, d.h. ein Paar $\langle T,\leq\rangle$. Es ist plausibel anzunehmen, daß diese Struktur $\leq$ in T Eigenschaften der von Gärdenfors vorgeschlagenen Art hat (worauf wir gleich zu sprechen kommen). Da sie aber nicht bloß vom epistemischen Zustand eines einzelnen abhängt, sondern — wie ich meine — am besten *als Bestandteil der Theorie* aufzufassen ist, habe ich mich für die besagte Namensgebung entschieden.[14]

Welche Eigenschaften hat nun diese Relation $\leq$, genauer gesagt: welche Eigenschaften haben die Paare $\langle T,\leq\rangle$? Hier ist zunächst zu beachten, daß Gärdenfors mit dem unwichtigen terminologischen Wechsel von epistemischer Wichtigkeit zu epistemischer Befestigung einen zwar unscheinbaren, aber nicht ganz so unwichtigen Wechsel seiner Bedingungen verbindet, ohne hierfür seine Gründe zu veröffentlichen. Beginnen wir aber mit der ersten, in Gärdenfors (1985, S. 350–352) präsentierten Version. Für jede

[12] Genauer gesagt, von einer totalen Quasiordnung; siehe unten (TW1) und (TW2).

[13] Englisch „entrenchment" heißt wörtlich „(Schützen-)Graben" oder „Verschanzung".

[14] Strenggenommen sollte ich nun den Theorienbegriff neu definieren, so daß eine *Theorie* ein Paar $\langle T,\leq\rangle$ ist, wobei T eine Theorie im alten Sinn und $\leq$ eine Relation mit den im folgenden beschriebenen Eigenschaften ist. Dies würde aber erstens eine etwas schwerfällige Sprechweise nach sich ziehen und zweitens das in Abschnitt 3.6 angesprochene Problem aufwerfen. Deshalb bleibe ich i.a. bei Definition 3.1.1.

Theorie, aufgefaßt als Paar $\langle T,\leq\rangle$, soll gelten:

(TW1) $\forall A,B,C\in T\ (A\leq B \wedge B\leq C \Rightarrow A\leq C)$ (Transitivität),
(TW2) $\forall A,B\in T\ (A\leq B \vee B\leq A)$ (Konnexität),
(TW3) $\forall A,B\in T\ (A\vdash B \Rightarrow A\leq B)$ (Dominanz),
(TW4^D) $\forall A,B\in T\ (A\vee B\leq A \vee A\vee B\leq B)$ (Disjunktivität).[15]

(TW1) und (TW2) besagen, daß $\leq$ eine totale Quasiordnung in T ist, und bedürfen als rein formale Bedingungen an die Relation der theoretischen Wichtigkeit kaum einer näheren Motivation. Anders verhält sich das mit (TW3) und (TW4^D). Aus (TW3) kann man ersehen, daß „B ist theoretisch wichtiger als A" eben *nicht* mit „B hat einen größeren Informationsgehalt als A" gleichgesetzt werden darf. Denn ein logisch stärkeres A ist nach (TW3) im allgemeinen weniger „wichtig". Wir bemerken, daß die Bezeichnung „Wichtigkeit" etwas irreführend ist, denn intuitiv wird man ein Gesetz in T wohl als wichtiger ansehen als alle seine Instantiierungen. Besser ist es, „$A\leq B$" als „die Aufgabe von B bringt einen größeren Informationsverlust mit sich als die Aufgabe von A" oder kurz „A ist leichter aufgebbar als B". Dies widerspricht nur scheinbar dem eben Gesagten, denn die Aufgabe von B würde im Falle von $A\vdash B$ ja, wie Gärdenfors unterstreicht, auch die Aufgabe von A implizieren ($A\notin T^-_B$, weil nach (T–4) für nicht logisch gültiges B $B\notin T^-_B$ gilt und T^-_B nach (T–1) eine Theorie ist), während das umgekehrte nicht der Fall zu sein braucht ($B\in T^-_A$ ist im allgemeinen völlig problemlos möglich).

Die Motivation von (TW4^D) macht jedoch Schwierigkeiten. Gärdenfors' (1985, S. 352) Bemerkung, daß „der Informationsgehalt von $A\vee B$ nicht größer sein könne als der maximale Gehalt der getrennt genommenen A und B", scheint auf der falschen Lesart von theoretischer Wichtigkeit zu basieren. Für die gemäß (TW3) richtige Interpretation von (TW4^D), welche besagt, daß der Informationsverlust durch die Aufgabe von $A\vee B$ (was die Aufgabe von A *und* von B impliziert!) nicht größer sein kann als der durch die Aufgabe von nur einem — dem wichtigeren — der beiden Sätze A und B entstandene Informationsverlust, gibt es aber vorderhand keine plausible

[15] In Gärdenfors (1984, S. 143) arbeitet Gärdenfors nur mit der Transitivität, Konnexität und Intensionalität von $\leq$, welch letztere Bedingung formalisierbar ist als

$\forall A,B \in T\ ((A\vdash B \wedge B\vdash A) \Rightarrow (A\leq B \wedge B\leq A))$.

Die Zusammenstellung von (TW1), (TW2), (TW3) und (TW4^D) ist etwas redundant, denn (TW2) folgt aus den anderen drei Bedingungen: Aus (TW3) folgt, daß $A\leq A\vee B$ und $B\leq A\vee B$, (TW4^D) besagt, daß andererseits $A\vee B\leq B$ oder $A\vee B\leq A$, was zusammen wegen (TW1) unmittelbar $A\leq B$ oder $B\leq A$, also (TW2) impliziert. Ich werde die Konnexität von $\leq$ als unabhängig motivierbare Bedingung im folgenden aber immer eigens aufführen.

Begründung. Richtig hingegen scheint es zu sein, daß die Aufgabe von $A \wedge B$ (was die Aufgabe von A *oder* von B impliziert) ein Informationsverlust in Kauf genommen werden muß, der nicht kleiner sein kann als der durch die Aufgabe von nur einem — dem unwichtigeren — der beiden Sätze A und B entstandene Informationsverlust. Mit der durch (TW3) nahegelegten Interpretation von $\leq$ heißt dies:

(TW4) $\forall A,B \in T\ (A \leq A \wedge B \vee B \leq A \wedge B)$ (Konjunktivität)

Ein zweites Argument für (TW4) anstelle von ($TW4^D$) kann man in der von Gärdenfors (1984, S. 148f) vorgetragenen Idee finden, wonach man in Umkehrung der hier diskutierten Strategie auch die Relation der theoretischen Wichtigkeit „herleiten" kann, wenn das Kontraktionsverhalten bekannt ist. Der Schlüssel dazu liegt in der ebenso interessanten wie einleuchtenden Beziehung

$$A \leq B \text{ genau dann, wenn } A \notin T^-_{A \wedge B},$$

auf die wir in Abschnitt 3.4 genauer zu sprechen kommen. Gärdenfors (1984, S. 149) weist darauf hin, daß ein so erhaltenes $\leq$ (TW1)-(TW3) erfüllt, sofern Kontraktionen gemäß (T-1)-(T-8) vonstatten gehen. Nun konnte Gärdenfors zwar nicht zeigen, daß ($TW4^D$) erfüllt wird. Aber man sieht — unter der Voraussetzung von (T-4) und (T-1) (und $\nvdash A \wedge B$) — sofort, daß A nicht in $T^-_{A \wedge B} = T^-_{A \wedge (A \wedge B)}$ oder daß B nicht in $T^-_{A \wedge B} = T^-_{B \wedge (A \wedge B)}$ ist, und dies heißt nach der obigen Beziehung gerade, daß $A \leq A \wedge B$ oder $B \leq A \wedge B$, d.h. daß (TW4) gilt. (Für $\vdash A \wedge B$ folgt (TW4) sofort aus (TW3).)

(TW4) ist denn auch in der Tat die in Gärdenfors (1988) und in Gärdenfors und Makinson (1988) anstelle von ($TW4^D$) verwendete Bedingung. Man sieht sich nun zunächst vor die Wahl gestellt, entweder auf die in Gärdenfors (1985) erbrachten, durchaus beeindruckenden Resultate (s.u.) ganz zu verzichten oder aber auf die Frage nach der Interpretierbarkeit von ($TW4^D$) eine Antwort schuldig bleiben zu müssen. Wir wollen erst im nächsten Abschnitt auf die Gärdenforsschen Resultate für ($TW4^D$) eingehen; im Rest dieses Abschnitts untersuchen wir, ob sich ein Weg finden läßt, ($TW4^D$) zu einer zufriedenstellenden Interpretation zu verhelfen.

Eine erste Idee wäre es, zu versuchen, die kritische Bedingung ($TW4^D$) aus (TW1)-(TW3) und dem wohlbegründeten (TW4) abzuleiten. ($TW4^D$) und (TW4) sehen ja, wie der Mathematiker sagen würde, völlig dual aus.[16]

[16] Zum Beispiel kann auch die Konnexität von $\leq$ aus (TW1), (TW3) und (TW4) ganz analog wie aus (TW1), (TW3) und ($TW4^D$) bewiesen werden. Vgl. die letzte Fußnote mit Gärdenfors (1988, S. 90, Text zu Fußnote 12.)

Leider aber stößt man hier auf ein negatives Resultat. Zur Vorbereitung seien ein paar abkürzende Definitionen vorausgeschickt: $A \not\leq B$ stehe für „nicht $A \leq B$"; $A < B$ stehe für $(A \leq B \wedge B \not\leq A)$, was wegen (TW2) gleichbedeutend mit $B \not\leq A$ ist; $A \doteq B$ stehe für $(A \leq B \wedge B \leq A)$; $A \neq B$ stehe für „nicht $A \doteq B$". Man beachte außerdem, daß wegen (TW3) für alle Sätze A in T $A \leq B$ gilt, sofern $\vdash B$. Falls $A \doteq B$ für ein B mit $\vdash B$, so nennen wir A einen *maximalwichtigen* Satz von T; andernfalls nennen wir A einen *normalwichtigen* Satz von T.

3.2.1. Theorem Sei $\langle T, \leq \rangle$ eine Theorie, welche (TW1)-(TW4) und ($TW4^D$) erfüllt. Dann ist $\langle T, \leq \rangle$ insofern trivial, als gilt: Es gibt keine zwei normalwichtigen Sätze A und B von T mit $A \neq B$.

Zum Beweis dieses Theorems verwenden wir folgendes

3.2.2. Lemma Sei $\langle T, \leq \rangle$ eine Theorie, welche (TW1), (TW2), (TW3) und ($TW4^D$) erfüllt. Dann gilt: Sind A und B normalwichtige Sätze von T, so ist die materiale Äquivalenz $A \leftrightarrow B$ ein maximalwichtiger Satz von T.

Erinnern wir uns daran, daß (TW4) eine sehr gute, ($TW4^D$) jedoch nur eine schlechte Interpretation gefunden hatte,[17] so weckt das Theorem 3.2.1 zusätzliche Zweifel an ($TW4^D$). Beinahe noch schlimmer ist aber die Aussage von Lemma 3.2.2. Die materiale Äquivalenz zweier zufällig ausgewählter Sätze in T, die nicht so wichtig sind wie die logischen Wahrheiten,[18] ist natürlich wieder in T, aber ihr Status hinsichtlich der theoretischen Wichtigkeit ist intuitiv völlig offen. Warum sollte dieser „dahergelaufene" komplexe Satz genauso wichtig sein wie eine logische Wahrheit?

Ein kleines Beispiel mag veranschaulichen, zu welch kontraintuitiven Konsequenzen ($TW4^D$) im Kontext von (TW1)-(TW3) tatsächlich führt.[19] Stellen wir uns erstens vor, ein „Wissenschaftler" habe die Theorie, daß alle Raben schwarz und alle Smargde grün sind. Natürlich sind diese beiden

[17] Eine formale Bewährung findet (TW4), aber nicht ($TW4^D$), wenn man $\leq$ über Lewis' „comparative possibility" (1973a, S. 52-56), über Spohns (1988a; 1988b) ordinale Konditionalfunktionen oder über die namenlos gebliebene Ordnungsrelation von Grove (1988, S. 163–167) definiert. Eine Wiedergabe der jeweiligen Modelle und ihrer Verknüpfung mit der Relation der theoretischen Wichtigkeit würde hier zu weit führen, die Grundidee kann aber ganz grob formuliert werden: A ist genau dann theoretisch wichtiger als B, wenn $\neg A$ „weiter hergeholt" oder „theoretisch schlechter denkbar" ist als $\neg B$. (Vgl. Gärdenfors 1984, S. 153f; 1988, S. 90). Eine weitere Bewährung von (TW4) wird in Abschnitt 3.4 besprochen.

[18] Es ist nicht unplausibel zu verlangen, daß *nur* die logischen Wahrheiten maximalwichtige Sätze von T sein sollen; vgl. unten (TW6).

[19] Hier setze ich voraus, daß wir die zweite Gärdenforssche (1985, S. 347; 1988, S. 87) Schlüsselidee zur theoretischen Wichtigkeit so verstehen dürfen, daß Sätze mit höchster theoretischer Wichtigkeit, wann immer dies möglich ist, beibehalten werden.

„Gesetze" mit einer gewissen Unsicherheit behaftet, weshalb man an ihnen nicht so unnachgiebig wie an logischen Wahrheiten festhalten darf. Kommt unser Wissenschaftler im Laufe seiner Forschungen zu dem Resultat, es sei doch möglich, daß manche Raben Albinos (also weiß) sind, so möchte er vielleicht das Raben-Gesetz seiner Theorie streichen. Da nach Lemma 3.2.2 aber die materiale Äquivalenz zwischen dem Raben- und dem Smaragd-Gesetz von höchster theoretischer Wichtigkeit ist und damit unbedingt beibehalten werden muß, ist der arme Wissenschaftler gezwungen, mit dem Raben- auch sein Smaragd-Gesetz aufzugeben — sicher eine absurde Konsequenz.

Etwas allgemeiner argumentiert, sei T eine beliebige endlich axiomatisierte Theorie und A_T die Konjunktion der Axiome von T. Aus Lemma 3.2.2 folgt, daß für jedes normalwichtige A in T das materiale Konditional $A \rightarrow A_T$ maximalwichtig ist (denn wegen $A_T \vdash A$ und (TW3) ist mit A auch A_T normalwichtig, und wieder wegen (TW3) ist mit $A \leftrightarrow A_T$ auch $A \rightarrow A_T$ maximalwichtig). Sei $\mathbf{A}_0$ eine endliche, aber große Menge normalwichtiger A's in T, die zusammengenommen einen wesentlichen Teil von T ausmachen. Aus ($TW4^D$) folgt, daß die Disjunktion A_0 der Elemente von $\mathbf{A}_0$ normalwichtig ist. Gemäß Lemma 3.2.2 hat $A_0 \rightarrow A_T$ höchste theoretische Wichtigkeit und muß deshalb in den Kontraktionen T^-_B bezüglich der meisten, wenn nicht aller B aus T beibehalten werden.[20] Interessieren wir uns aber für Revisionen von T, dann wissen wir — unter Zugrundelegunge der Levi-Identität —, daß auch in den meisten, wenn nicht gar in allen Revisionen $T^*_{\neg B}$ (wobei B aus T) $A_0 \rightarrow A_T$ enthalten ist. Wegen (T*2) ist $\neg B$ in $T^*_{\neg B}$, wegen $A_T \vdash B$ ist also auch $\neg A_T$ in $T^*_{\neg B}$. Da $T^*_{\neg B}$ andererseits $A_0 \rightarrow A_T$ enthält, folgt schließlich per Kontraposition, daß $\neg A_0$ in $T^*_{\neg B}$ ist. A_0 war aber die Disjunktion der Elemente von $\mathbf{A}_0$. Wir erhalten also als Ergebnis, daß in den meisten, wenn nicht in allen Revisionen von T die Negationen *sämtlicher* Elemente der für die Identität von T gewichtigen Menge $\mathbf{A}_0$ enthalten sein müssen. Dies kann sicher nicht mit dem Ziel einer minimalen Revision von T übereinstimmen. Besonders skandalös wird die Sachlage, wenn sämtliche Axiome von T normalwichtig sind: Dann wird in $T^*_{\neg B}$ *jedes* Axiom von T explizit geleugnet.

Damit scheint klar zu sein, daß ($TW4^D$) nicht zu retten ist, zumindest dann nicht, wenn man (TW1)-(TW3) akzeptiert. Die „duale" Bedingung (TW4) dagegen ist gut zu interpretieren und paßt formal zu (TW1)-(TW3).

[20] Diese Aussage muß vage bleiben, solange wir nur eine vage Idee haben, wie genau die Relation der theoretischen Wichtigkeit bei Kontraktionsbildungen zur Anwendung kommt. Präzise Modelle hierfür werden in den nächsten Abschnitten entwickelt.

Gehen wir nun an die Begutachtung dessen, was mit der Relation der theoretischen Wichtigkeit bewerkstelligt werden kann: die Konstruktion eindeutiger Kontraktionen und — vermittels der Levi-Identität — eindeutiger Revisionen. In den Arbeiten von Gärdenfors, Alchourrón und Makinson finden sich dazu zwei verschiedene Ansätze, denen wir die nächsten beiden Abschnitte widmen wollen.

3.3 Der erste Weg zu eindeutigen Kontraktionen und Revisionen: Durchschnitte maximalkonsistenter Teilmengen

Wie kommt man von T zu T^-_A? Eine erste naheliegende Idee ist es, einfach eine mit $\neg A$ maximalkonsistente Teilmenge von T zu nehmen, d.h. eine Teilmenge M mit $M \not\vdash A$ und $M' \vdash A$ für alle von M verschiedenen M' mit $M \subseteq M' \subseteq T$. Ein solches M darf man als existent ansehen, wenn man — wie ich hier — das Auswahlaxiom unproblematisiert läßt. Interessanter ist die Feststellung, daß es normalerweise nicht nur ein, sondern sehr viele solche M's gibt. Aus Symmetriegründen kann man nun die zweite naheliegende Ansicht vertreten, man sollte alle mit $\neg A$ maximalkonsistenten Teilmengen von T gleich behandeln und die Kontraktion T^-_A als den Durchschnitt all dieser Mengen bestimmen. Für die erste Methode haben Alchourrón, Makinson und später Gärdenfors den Namen „*(maxi-)choice contraction*" und für die zweite Methode den Namen „*(full) meet contraction*" geprägt. Alchourrón und Makinson (1982, S. 18–21) haben nachgewiesen, daß beide Methoden — angewandt auf Theorien — zu unbefriedigenden Resultaten führen. Besonders plastisch wird dies deutlich, wenn man die Revisionen betrachtet, die über die Levi-Identität aus den entsprechenden Kontraktionen gewonnen werden können. Ein per sogenannter „(maxi-)choice revision" gewonnenes T^*_A enthält (für $\neg A \in T$) intuitiv viel zu viel: für jeden Satz B ist entweder B oder $\neg B$ in T^*_A (einerlei, wieviel in T bekannt war). Ein per sogenannter „(full) meet revision" gewonnenes T^*_A enthält (für $\neg A \in T$) intuitiv viel zu wenig: nämlich nur die logischen Konsequenzen von A.

Offensichtlich muß ein Mittelweg gefunden werden. Es ist eine natürliche Idee, als T^-_A den Durchschnitt *einiger* mit $\neg A$ maximalkonsistenter Teilmengen herzunehmen. Eine Kontraktionsoperation (über einer festen Theorie T), die nach dieser Methode konstruiert wird, heißt „*partial meet*

contraction function" („*PMCF*") (über T). Daß man durch diese Idee schon sehr viel von dem, was man haben will, erfaßt hat, zeigt folgendes *Repräsentationstheorem für die grundlegenden Gärdenfors-Postulate* (siehe Alchourrón, Gärdenfors und Makinson 1985, Observation 2.5):

3.3.1. Theorem Sei T eine Theorie. Eine Kontraktionsoperation $^-$ (über T) erfüllt die grundlegenden Gärdenfors-Postulate (T–1)–(T–6) genau dann, wenn sie eine PMCF (über T) ist.

Eine Weiterführung dieser Idee ist es, daß die für den Durchschnitt ausgewählten mit ¬A maximalkonsistenten Teilmengen solche sind, die in irgendeinem noch zu präzisierenden Sinn „theoretisch bevorzugt" sind. Auch die Idee einer solchen *theoretischen Bevorzugung* läßt sich mit Hilfe einer Relation $\preceq$ darstellen. Bezeichne nun T⊥A die Klasse der mit ¬A maximalkonsistenten Teilmengen von T. Dann kann man schreiben (vgl. Alchourrón, Gärdenfors und Makinson 1985, S. 517–524):

(RPM) $T^-_A = \bigcap \{M \in T\perp A: \forall M' \in T\perp A\ (M' \preceq M)\}$.

Falls T⊥A leer ist, d.h. falls $\vdash A$, wird T^-_A gleich T gesetzt. Um vernünftige Kontraktionen zu erhalten, müssen wir folgende Voraussetzung machen:

(TB0) $\{M \in T\perp A: \forall M' \in T\perp A\ (M' \preceq M)\} \neq \emptyset$, falls $T\perp A \neq \emptyset$.

Ist $\preceq$ eine (TB0) erfüllende Relation und kann die Kontraktionsoperation $^-$ (über T) vermittels (RPM) aus $\preceq$ konstruiert werden, so nennen wir $^-$ eine *$\preceq$-relationale PMCF*. Gibt es ein $\preceq$, so daß $^-$ eine $\preceq$-relationale PMCF ist, so heißt $^-$ eine *relationale PMCF*. Relationale PMCFs erfüllen schon (T–7) (siehe Alchourrón, Gärdenfors und Makinson 1985, Observations 4.2 und 4.3). Bezeichne weiter M(T) die Klasse $\bigcup \{T\perp A: \nvdash A\}$.[21] Wir führen uns noch einmal vor Augen, daß $\preceq$ eine Relation *auf Teilmengen von T* oder zumindest *auf den Elementen von M(T)* ist, anders als die Relation $\leq$ der theoretischen Wichtigkeit, die *Sätze in T* ordnet. Es fragt sich nun wieder, welche Eigenschaften $\preceq$ haben soll. Wir entscheiden uns dafür, im folgenden zu verlangen, daß $\preceq$ eine totale Quasiordnung in der Potenzmenge von T bzw. in M(T) sei:

(TB1) $\forall M,M',M''\ (M \preceq M' \wedge M' \preceq M'' \Rightarrow M \preceq M'')$ (Transitivität)

(TB2) $\forall M,M'\ (M \preceq M' \vee M' \preceq M)$ (Konnexität)

Gibt es eine transitive (bzw. transitive und konnexe) Relation $\preceq$, aus der die Kontraktionsoperation $^-$ (über T) konstruierbar ist, so nennen wir

[21] Diese Definition unterscheidet sich von der bei Gärdenfors (1985, S. 357; 1988, S. 81), wo $M(T) = \bigcup \{T\perp A: A \in T \wedge \nvdash A\}$, und von der bei Alchourrón, Gärdenfors und Makinson (1985, S. 523), wo $M(T) = \bigcup \{T\perp A: A \in T\}$ gesetzt ist.

$^-$ eine *transitiv relationale* (bzw. *transitiv-konnex relationale*) *PMCF*. Mit der Konnexität sind wir eigentlich sogar schon etwas über unser Ziel hinausgeschossen, wie das folgende *Repräsentationstheorem für die (vollständigen) Gärdenfors-Postulate* zeigt (siehe Alchourrón, Gärdenfors und Makinson 1985, Corollary 4.5, und Gärdenfors 1988, Theorem 4.16):

3.3.2. Theorem Sei T eine Theorie. Eine Kontraktionsoperation $^-$ (über T) erfüllt die Gärdenfors-Postulate (T–1)–(T–8) genau dann, wenn sie eine transitiv relationale PMCF (über T) ist.

Wie Alchourrón, Gärdenfors und Makinson (1985, Observation 5.1) nachweisen, schadet die Forderung der Konnexität aber auch nichts, da jede transitiv relationale PMCF auch schon transitiv-konnex relational ist. Weil wir die Konnexität von $\preceq$ aber im folgenden brauchen werden, gehen wir von den Bedingungen (TB1) und (TB2) für die Relation der theoretischen Bevorzugung aus.

Gärdenfors (1985) hat die intuitiv natürlichere Relation der theoretischen Wichtigkeit $\leq$ mit der hier für den Zweck der Theorienkontraktion und -revision benötigten Relation der theoretischen Bevorzugung $\preceq$ in Zusammenhang gebracht. Dabei hat er eine Interdefinierbarkeit von $\leq$ und $\preceq$ vorgeschlagen, vermittels der man nicht nur $\preceq$ als Funktion $\mathbf{TB}(\leq)$ von $\leq$, sondern auch $\leq$ als Funktion $\mathbf{TW}(\preceq)$ von $\preceq$ auffassen kann. (Genaueres dazu siehe unten.) Die Verbindung zwischen $\leq$ und $\preceq$ hat er so geschickt eingerichtet, daß er folgende Ergebnisse vorweisen kann: Wenn $\preceq$ (auf M(T)) die Bedingungen (TB1) und (TB2) erfüllt, dann erfüllt $\mathbf{TW}(\preceq)$ die Bedingungen (TW1), (TW2), (TW3) und (TW4^D); wenn umgekehrt $\leq$ die Bedingungen (TW1), (TW2), (TW3) und (TW4^D) erfüllt, dann erfüllt $\mathbf{TB}(\leq)$) (auf M(T)) die Bedingungen (TB1) und (TB2), sofern T finit ist. Eine *finite* Theorie ist dabei eine solche, die in endlich viele Äquivalenzklassen bezüglich Cn zerfällt.[22] Für finite Theorien T kann Gärdenfors des

[22] In der ansonsten mit Gärdenfors (1985) wesentlich identischen Entwurfsversion von Gärdenfors (1988, Kapitel 4) arbeitete Gärdenfors etwas allgemeiner mit vollständigen Theorien. Eine Theorie heißt dabei *vollständig*, wenn sie erstens in einer Sprache formuliert ist, die Disjunktionen (großes Disjunktionssymbol '$\bigvee$') und Konjunktionen (großes Konjunktionssymbol '$\bigwedge$') der Elemente beliebiger Satzmengen S erlaubt, und zweitens gegenüber einer Logik abgeschlossen ist, welche die beiden folgenden Schlußregeln enthält:

$\bigwedge S \vdash A$ für alle $A \in S$.

Falls $S' \vdash A$ für alle $A \in S$, so $S' \vdash \bigwedge S$.

(Es genügt offenbar nicht,

Falls $B \vdash A$ für alle $A \in S$, so $B \vdash \bigwedge S$

zu fordern.) Der entscheidende Punkt an finiten und vollständigen Theorien T ist, daß beide stets *Repräsentanten* $\bigwedge T$ beinhalten mit der Eigenschaft, daß A genau dann

weiteren zeigen: Wenn $\preceq$ (auf M(T)) (TB1) und (TB2) erfüllt ist, dann ist $\mathbf{TB}(\mathbf{TW}(\preceq))$ identisch mit $\preceq$; wenn andererseits $\leq$ die Bedingungen (TW1), (TW2), (TW3) und ($TW4^D$) erfüllt, dann ist $\mathbf{TW}(\mathbf{TB}(\leq))$ identisch mit $\leq$.

Das von Gärdenfors (1985) auf diese Art und Weise errichtete formale Gebäude ist natürlich sehr schön, und es wäre schade, wenn man auf die erwähnten Resultate verzichten müßte. Doch wir haben im vorigen Abschnitt gesehen, daß das in diesen Zusammenhang wesentlich eingehende ($TW4^D$) als Bedingung für $\leq$ unhaltbar ist und durch (TW4) ersetzt werden muß. Kann man dann noch analoge Ergebnisse erzielen? Ich werde in diesem Abschnitt zeigen, daß dies möglich ist.

Als erstes ist es notwendig, sich einmal genau anzusehen, wie Gärdenfors (1985, S. 359f) die Brücke zwischen $\leq$ und $\preceq$ schlägt. Seine Definitionen lauten so:

3.3.3. Definition (a) Sei die Relation $\leq$ der theoretischen Wichtigkeit in T gegeben. Dann ist $\preceq = \mathbf{TB}(\leq)$ nach Gärdenfors durch die folgende Bestimmung definiert: Für alle M und M′ in M(T) gilt

$$M \preceq M' \Leftrightarrow \exists A{\in}M\ \forall B{\in}M'\ (A{\leq}B)\ .$$

(b) Sei die Relation $\preceq$ der theoretischen Bevorzugung in M(T) gegeben. Dann ist $\leq = \mathbf{TW}(\preceq)$ nach Gärdenfors durch die folgende Bestimmung definiert: Für alle A und B in T gilt

$$A \leq B \Leftrightarrow \forall M{\in}M(T)\ (A{\in}M \Rightarrow \exists M'{\in}M(T)\ (B{\in}M' \wedge M{\preceq}M'))\ .$$

Nach Teil (a) dieser Definition kann M nicht den Vorzug vor M′ erhalten, wenn die unwichtigsten Sätze in M höchstens so wichtig sind wie die unwichtigsten Sätze in M′. Nach Teil (b) kann A nicht wichtiger sein als B, wenn man keine maximale, A enthaltende Teilmenge von T finden kann, die allen maximalen, B enthaltenden Teilmengen von T vorzuziehen ist.[23] Diese Bedingungen sind prima facie nicht unplausibel.[24] Aber wie wir in den intuitiven Betrachtungen zur Interpretation von $\leq$ schon gesehen haben, scheint es mehr darauf anzukommen, was bei Kontraktionen verloren wird,

---

in T ist, wenn $\bigwedge T \vdash A$ gilt. Insofern ist es einerlei, ob man mit finiten oder vollständigen Theorien arbeitet.

[23] Die Redeweise von „maximalen Teilmengen von T" ist schlampig; ich meine damit natürlich mit $\neg A$ maximalkonsistente Teilmengen von T, wobei A ein Satz aus T sei. Übrigens ist es kein Fehler, wenn man nicht erwähnt, *mit welchem* $\neg A$ ein M aus M(T) maximalkonsistent ist. Wie nämlich Alchourrón, Gärdenfors und Makinson (1985, Lemma 2.4) zeigen, ist ein M aus M(T) für *jedes* $A{\in}T{\setminus}M$ in $T{\perp}A$ (man beachte, daß M eine Theorie ist).

[24] Allerdings wirkt es in Teil (a) schon etwas sonderbar, daß auf die unwichtigsten und nicht auf die wichtigsten Sätze in M und M′ Bezug genommen wird.

als darauf, was in den Kontraktionen erhalten bleibt. Natürlich kann man bei gegebenem T das eine aus dem anderen erhalten. Die Pointe ist aber, daß als Bewertungsmaßstab für $\mathbf{TB}(\leq)$ und $\mathbf{TW}(\preceq)$ nicht die $\preceq$-Positionen bzw. $\leq$-Positionen der in den relevanten maximalkonsistenten Teilmengen enthaltenen, sondern die der herausgefallenen Sätze entscheidend sind. Ich schlage folgende Verbesserung der Gärdenforsschen Definitionen vor:

3.3.4. Definition (a) Sei die Relation $\leq$ der theoretischen Wichtigkeit in T gegeben. Dann sei $\preceq = \mathbf{TB}(\leq)$ definiert durch die folgende Bestimmung: Für alle M und M′ in M(T) gilt

$$M \preceq M' \Leftrightarrow \forall A{\in}T\backslash M' \; \exists B{\in}T\backslash M \; (A{\leq}B) \; .$$

(b) Sei die Relation $\preceq$ der theoretischen Bevorzugung in M(T) gegeben. Dann sei $\leq = \mathbf{TW}(\preceq)$ definiert durch die folgende Bestimmung: Für alle A und B in T gilt

$$A \leq B \Leftrightarrow \forall M{\in}M(T) \; (B{\in}T\backslash M \Rightarrow \exists M'{\in}M(T) \; (A{\in}T\backslash M' \wedge M{\preceq}M')).$$

Nach Teil (a) dieser neuen Definition kann M nicht den Vorzug vor M′ erhalten, wenn die wichtigsten in M′ fehlenden Sätze höchstens so wichtig sind wie die wichtigsten in M fehlenden Sätze. Insbesondere kann kein $M{\in}M(T)$ wichtiger als $T \in T{\perp\!\!\perp} \subseteq M(T)$ sein. Nach Teil (b) kann A nicht wichtiger sein als B, wenn man keine maximale, B nicht enthaltende Teilmenge von T finden kann, die allen maximalen, A nicht enthaltenden Teilmengen von T vorzuziehen ist. Diese intuitiven Paraphrasierungen sind sicher zu wenig überschaubar, als daß sie es erlaubten, Definition 3.3.4 vor Definition 3.3.3 eindeutig als besser auszuzeichnen. Man kann aber zeigen, daß das ganze Programm von Gärdenfors (1985) mit der neuen Definition durchzuführen ist, und zwar so, daß (TW4) an die Stelle von $(TW4^D)$ tritt:

3.3.5. Lemma (a) Sei $\leq$ (auf T) gegeben und $\preceq = \mathbf{TB}(\leq)$ nach Definition 3.3.4(a) bestimmt; wenn $\leq$ die Bedingungen (TW1) und (TW2) erfüllt, dann erfüllt $\preceq$ auf M(T) (TB1) und (TB2);[25]
(b) sei $\preceq$ (auf M(T)) gegeben und $\leq = \mathbf{TW}(\preceq)$ nach Definition 3.3.4(b) bestimmt; wenn $\preceq$ (TB1) und (TB2) erfüllt, dann erfüllt $\leq$ auf T die Bedingungen (TW1)-(TW4).

3.3.6. Theorem Es gilt bei Zugrundelegung von Definition 3.3.4:
(a) Wenn $\leq$ auf T die Bedingungen (TW1)-(TW4) erfüllt, dann ist $\mathbf{TW}(\mathbf{TB}(\leq))$ identisch mit $\leq$;
(b) wenn T eine finite (oder vollständige[26]) Theorie und wenn $\preceq$ eine Re-

[25] Den Nachweis, daß $\preceq$ auf M(T) (TB0) erfüllt, verschieben wir auf Abschnitt 3.5; s. dort den Beweis von Theorem 3.5.8.
[26] Siehe Fußnote 22.

lation auf M(T) ist, dann ist $\mathbf{TB}(\mathbf{TW}(\preceq))$ identisch mit $\preceq$.

Man beachte, daß die Eigenschaften (TB1) und (TB2) im Beweis von 3.3.6(b) nicht gebraucht werden (aber auch mit (TB1) und (TB2) würde man, soweit ich sehe, auf die Finitheits- oder Vollständigkeitsvoraussetzung für T nicht verzichten können). Die Einschränkung auf finite (oder vollständige) Theorien ist weniger gravierend als man denkt, wenn man sich vor Augen hält, daß wir an Relationen $\preceq$ auf M(T) interessiert sind, die mit Relationen $\leq$ auf T in Zusammenhang stehen. Wenn etwa das $\preceq$ in (b) aus einer Relation der theoretischen Wichtigkeit $\leq$ (welche (TW1)–(TW4) erfüllt) vermittels Definition 3.3.4(a) konstruierbar ist, d.h. wenn es ein $\leq$ gemäß (TW1)–(TW4) gibt mit $\preceq = \mathbf{TB}(\leq)$, dann gilt auch im allgemeinen Fall $\mathbf{TB}(\mathbf{TW}(\preceq)) = \preceq$, da nach Teil (a) von Theorem 3.3.6 $\mathbf{TB}(\mathbf{TW}(\mathbf{TB}(\leq))) = \mathbf{TB}(\leq)$. Wenn aber für ein $\preceq$, welches die Bedingungen (TB1) und (TB2) erfüllt, $\mathbf{TB}(\mathbf{TW}(\preceq)) \neq \preceq$ gelten sollte, dann sind immerhin die aus $\mathbf{TB}(\mathbf{TW}(\preceq))$ und aus $\preceq$ vermittels Definition 3.3.4(b) erhältlichen Relationen der theoretischen Wichtigkeit identisch, da $\mathbf{TW}(\preceq)$ nach Lemma 3.3.5(b) (TW1)–(TW4) erfüllt und $\mathbf{TW}(\mathbf{TB}(\mathbf{TW}(\preceq))) = \mathbf{TW}(\preceq)$ wieder nach Teil (a) von Theorem 3.3.6 richtig ist.

Zum Schluß dieses Abschnitts noch ein paar abschließende Bemerkungen. Die Richtung, die uns von unserer Fragestellung her mehr interessiert, ist die Richtung von $\leq$ zu $\preceq$. Es ist einleuchtend, von einer Ordnungsstruktur in T auszugehen und hinzuarbeiten auf eine Ordnungsstruktur auf M(T), die wir für die relationale PMCF-Konstruktion von Kontraktionen und Revisionen benötigen. Zwei Punkte zu dieser Richtung sollen noch angesprochen werden.

Erstens fällt auf, daß im Definiens der Definition 3.3.4(a) die Reihenfolge der Quantoren gegenüber der Gärdenforsschen Definition 3.3.3(a) umgedreht wurde: statt einer $\exists$–$\forall$-Bedingung haben wir jetzt eine schwächere $\forall$–$\exists$-Bedingung. Intuitiv sollte diese Abschwächung keine Schwierigkeiten bereiten, denn auch die schwächere Bedingung kann offenbar Grundlage einer Bevorzugung sein. Formal würden, wie man sich überzeugt, die im Anhang präsentierten Beweise auch mit der $\exists$–$\forall$-Bedingung durchgehen — mit einer wichtigen Einschränkung: Die Finitheitsvoraussetzung (bzw. Vollständigkeitsvoraussetzung) für T wäre ähnlich wie bei Gärdenfors in Lemma 3.3.5(a) und Theorem 3.3.6(a) nötig. Der einzige Teil, wo es mir nicht gelungen (und vermutlich unmöglich) ist, diese Voraussetzung loszuwerden, ist Theorem 3.3.6(b). Damit ist der erste Vorteil der $\forall$–$\exists$-Bedingung genannt. Weitere Vorteile werden wir in Abschnitt 3.5 beim Aufzeigen der Verbindung zwischen dem ersten und dem zwei-

ten Weg zu eindeutigen Kontraktionen und Revisionen kennenlernen (s. Lemma 3.5.5).[27]

Zweitens ist es interessant zu wissen, daß mit Definition 3.3.4(a) die Relation $\preceq = \mathbf{TB}(\leq)$ nicht nur auf M(T), sondern auch auf der ganzen Potenzmenge von T definiert werden kann. Den Teil (a) von Lemma 3.3.5 entsprechend erweiternd, kann man dann beweisen:

3.3.7. Lemma Sei $\leq$ (auf T) gegeben und $\preceq = \mathbf{TB}(\leq)$ nach Definition 3.3.4(a) bestimmt; wenn $\leq$ die Bedingungen (TW1) und (TW2) erfüllt, dann erfüllt $\preceq$ auf Pot(T) (TB1), (TB2)[28],

(TB3) $\forall$ M,M′ (M$\subseteq$M′ $\Rightarrow$ M$\preceq$M′) (Dominanz) und

(TB4) $\forall$ M,M′ (M$\preceq$M$\cap$M′ $\vee$ M′$\preceq$M$\cap$M′) (Durchschnitt).

Wenn man $\mathbf{TB}(\leq)$ nur in M(T) definiert haben will, ist weder (TB3) noch (TB4) relevant, insofern als für alle M und M′ aus M(T)\{T} M$\not\subseteq$M′ gilt und M$\cap$M′ nicht mehr in M(T) liegt. Man beachte, daß in (TB3) die Verhältnisse von (TW3) gewissermaßen umgekehrt werden: während (TW3) ungefähr besagt, daß ein Satz A desto $\leq$-größer ist, je schwächer, also je *weniger informativ* er ist, besagt (TB3), daß eine Satzmenge M desto $\preceq$-größer ist, je größer, also je *informativer* sie ist.[29]

3.4 Der zweite Weg zu eindeutigen Kontraktionen und Revisionen: eine direkte Konstruktion

Gärdenfors hatte im wesentlichen drei Gründe für die Aufgabe seines in Gärdenfors (1985) entwickelten Modells.[30] Zunächst überzeugten ihn die in Abschnitt 3.2 vorgebrachten formalen Argumente davon, daß die intuitive Kritik an ($TW4^D$) — auch David Makinson und Isaac Levi hatten den intuitiven Gehalt von ($TW4^D$) kritisiert — berechtigt ist, und besiegelten definitiv das Schicksal dieser Bedingung. Wir haben im letzten Abschnitt gezeigt, daß deswegen nicht gleich der ganze Ansatz rettungslos verloren

[27] Eine Quantorenvertauschung in Definition 3.3.4(b) bliebe wegen (TB0) dagegen ohne Wirkung.

[28] Die Konnexität von $\preceq$ kann nun auf dieselbe Weise als redundant nachgewiesen werden wie oben die Konnexität von $\leq$.

[29] Vergleiche auch Ginsbergs (1986, z.B. S. 46) Idee, eine komparative Relation auf Satzmengen einzuspannen, welche „die der Mengeninklusion entsprechende partielle Ordnung erweitert."

[30] Laut Brief vom 18. Februar 1987.

ist. Zweitens hatte er seine Idee der „Herleitbarkeit" der theoretischen Wichtigkeit aus dem Kontraktionsverhalten durch die Definition

$$A \leq B \text{ genau dann, wenn } A \notin T^{-}_{A \wedge B}$$

nicht mit dem 1985er Modell in Einklang bringen können. Wie wir in Abschnitt 3.2 angedeutet haben, schafft hier die Ersetzung von (TW4^{D}) durch (TW4) eine gewisse Abhilfe. Allerdings hat die Definition von $\leq$ über Kontraktionen bis jetzt noch keinen festen systematischen Wert. Drittens war die verschiedentlich notwendige Endlichkeits- bzw. Vollständigkeitsforderung für T eine zumindest unästhetische, unter Umständen aber wirklich einschneidende Voraussetzung. In einer 1988 veröffentlichten Arbeit war es Adam Grove inzwischen gelungen, das Gärdenforssche Revisionsmodell eng an Sphärensysteme à la David Lewis (1973a) anzubinden und den Weg zu einer direkten Konstruktion von Kontraktionen zu weisen. Grove vermied dabei erstens den „Umweg" über maximalkonsistente Teilmengen und zweitens die Endlichkeits- bzw. Vollständigkeitsforderung für T. Diese drei Gründe veranlaßten Gärdenfors, sein altes Modell ganz über Bord zu werfen und statt dessen ein neues Modell auszuarbeiten. Während in Gärdenfors (1988, Abschnitt 4.8) die Verpflichtung dieses neuen Modells gegenüber Grove deutlich wird, liegt es in Gärdenfors und Makinson (1988) nun in einer vereinfachten und sehr eleganten Form vor, die sich nicht mehr auf Grove bezieht. Wir wollen den zweiten Weg zu eindeutigen Kontraktionen und Revisionen nun in dieser letzteren Form wiedergeben.[31]

Zunächst kann man von der Ordnung der theoretischen Wichtigkeit ohne Umschweife zu Kontraktionen kommen:

3.4.1. Definition Sei T eine Theorie und seien A und B beliebige Sätze. Dann ist die *$\leq$-ordnungsgemäße* Kontraktionsoperation $^{-} = \mathbf{Kon}(\leq)$ definiert durch

$$B \in T^{-}_{A} \Leftrightarrow B \in T \wedge (A < A \vee B \vee \vdash A) .$$

Umgekehrt kann nun die bereits oben angesprochene Charakterisierung von $\leq$ über das Kontraktionsverhalten bezüglich T als systematischer Angelpunkt dienen:

3.4.2. Definition Sei T eine Theorie und seien A und B beliebige Sätze. Dann ist die Relation $\leq = \mathbf{TW}(^{-})$ definiert durch

$$A \leq B \Leftrightarrow A \notin T^{-}_{A \wedge B} \vee \vdash A \wedge B .$$

Definition 3.4.2 leuchtet sehr gut ein: A kann nur dann wichtiger als B sein, wenn es bei der Kontraktionsbildung bezüglich $A \wedge B$, die ja durch Aufgeben

[31] Doch möchte ich hervorheben, daß einige der in diesem Kapitel präsentierten Gedanken aus der Auseinandersetzung mit Grove (1988) hervorgegangen sind.

von A oder durch Aufgeben von B zu erledigen ist, nicht aufgegeben wird. Definition 3.4.1 hingegen ist nicht unmittelbar verständlich. Gärdenfors und Makinson (1988, S. 89f) haben ihr aber zu einer indirekten Rechtfertigung über Definition 3.4.2 verholfen. Aus dieser erhält man nämlich, wenn man 'B' durch 'A∨B' ersetzt, die Beziehung

$$A < A \vee B \Leftrightarrow A \vee B \in T^{-}_{A \wedge (A \vee B)} \;\&\; \nvdash A.$$

Durch beiderseitige Adjunktion von $\vdash A$ erhält man unter Zuhilfenahme von (T–6) und (T–1)

$$A < A \vee B \vee \vdash A \Leftrightarrow A \vee B \in T^{-}_{A}.$$

Nun schreibe man als Konjunktionsglied $B \in T$ auf beide Seiten:

$$(A < A \vee B \vee \vdash A) \;\&\; B \in T \Leftrightarrow A \vee B \in T^{-}_{A} \;\&\; B \in T.$$

Mit der Bedingung (T–5) ist nun aber für ein B in T der Satz $A \rightarrow B$, oder äquivalent $\neg A \vee B$, in T^{-}_{A}. Deshalb ist $A \vee B$ genau dann in T^{-}_{A}, wenn B in T^{-}_{A} ist. Mit (T–2) gewinnt man schließlich aus der letzten Beziehung die Definition 3.4.1.

Wie zwingend Definition 3.4.1 aber tatsächlich ist, sofern man den Rest des Rahmens akzeptiert, zeigt sich erst an den in Gärdenfors und Makinson bewiesenen Theoremen. In der neuen Arbeit ist $\leq$ eine Relation in der Menge aller Sätze (nicht nur in T), weshalb man sich bei den Kriterien (TW1)–(TW4) die Einschränkung „$\in T$“ von nun an gestrichen denke. Außerdem gesellen sich hier noch zwei weitere Kriterien für die Relation der theoretischen Wichtigkeit hinzu, nämlich die Extremalbedingungen

(TW5) $T \neq T_{\perp} \Rightarrow \forall A\, (A \notin T \Leftrightarrow \forall B\, (A \leq B))$ (Minimalität),
(TW6) $\forall A\, (\forall B\, (B \leq A) \Rightarrow \vdash A)$) (Maximalität).

Man beachte, daß (TW5) jetzt die einzige Bedingung ist, die explizit auf T Bezug nimmt. Sie ermöglicht eine nichttriviale Charakterisierung von T ($\neq T_{\perp}$) durch $\leq$ vermittels der Gleichung $T = \{A: \exists B(B < A)\}$ oder — wegen (TW3) und (TW1) äquivalent — vermittels $T = \{A: \perp < A\}$. — Um eine reibungslose Vergleichbarkeit mit dem Repräsentationstheorem des vorigen Abschnitts zu gewährleisten, wollen wir zuerst noch definieren, daß eine Kontraktionsoperation $^{-}$ (über T) genau dann *ordnungsgemäß* heißt, wenn es eine Relation $\leq$ der theoretischen Wichtigkeit gemäß (TW1)–(TW6) gibt, so daß $^{-}$ $\leq$-ordnungsgemäß ist. Dann bekommt man das zum vorigen Abschnitt analoge

3.4.3. Theorem (Gärdenfors und Makinson) Sei T eine Theorie. Eine Kontraktionsoperation $^{-}$ (über T) erfüllt die Gärdenfors-Postulate (T–1)–(T–8) genau dann, wenn sie eine ordnungsgemäße Kontraktionsoperation (über T) ist.

Der Unterschied zum 1985er Modell von Gärdenfors ist nun der, daß die Relation $\leq$ zu einer bestimmten Kontraktionsoperation sofort explizit und eindeutig angegeben werden kann: über Definition 3.4.2.[32] Insofern sind die Ergebnisse von Gärdenfors und Makinson (1988) viel befriedigender. Wir wollen sie noch einmal förmlich zusammenfassen:

3.4.4. Theorem Sei T eine Theorie. Dann gilt:
(a) Erfüllt die Relation $\leq$ der theoretischen Wichtigkeit die Bedingungen (TW1)-(TW6), so erfüllt $^- = \mathbf{Kon}(\leq)$ die Gärdenfors-Postulate (T-1)-(T-8), und es gilt

$$A \leq B \Leftrightarrow A \notin T^-_{A \wedge B} \vee \vdash A \wedge B .$$

(b) Erfüllt die Kontraktionsoperation $^-$ die Gärdenfors-Postulate (T-1)-(T-8), so erfüllt $\leq = \mathbf{TW}(^-)$ die Bedingungen (TW1)-(TW6), und es gilt

$$B \in T^-_A \Leftrightarrow B \in T \wedge (A < A \vee B \vee \vdash A) .$$

Als Korollar zu diesem Theorem führen Gärdenfors und Makinson an, daß $\mathbf{Kon}(\mathbf{TW}(^-) = ^-$ und $\mathbf{TW}(\mathbf{Kon}(\leq) = \leq$ gilt. Damit haben wir den zweiten Weg zu eindeutigen Kontraktionen und Revisionen in seiner ganzen Kraft und Eleganz kennengelernt.

3.5 Der Zusammenhang zwischen dem ersten und dem zweiten Weg

Für die Diskussion des Zusammenhangs von erstem und zweitem Weg zur Konstruktion eindeutiger Kontraktionen und Revisionen führen wir einen Begriff ein, der sich als sehr leistungsfähig erweisen wird:

3.5.1. Definition Sei $\langle T, \leq \rangle$ eine Theorie. Dann heißt eine Satzmenge S ein *TW-Schnitt* (*relativ zur theoretischen Wichtigkeit* $\leq$) genau dann, wenn für alle Sätze A und B gilt: Falls A in S ist und $A \leq B$, dann ist auch B in S.

TW-Schnitte haben folgende einfache Eigenschaften:

3.5.2. Lemma Sei $\langle T, \leq \rangle$ eine Theorie, die (TW1)-(TW6) erfüllt. Dann gilt:
(a) Für alle TW-Schnitte S und S' ist $S \subseteq S'$ oder $S' \subseteq S$.
(b) Satzmengen der Form $\{B\colon A < B\}$ und der Form $\{B\colon A \leq B\}$ sind TW-Schnitte.

[32] Im Beweis ihres Repräsentationstheorems geben Alchourrón, Gärdenfors und Makinson (1985, S. 519) eine komplizierte Konstruktion der Relation $\preceq$ zu einer bestimmten Kontraktionsoperation an.

(c) Der größte TW-Schnitt ist $Cn(\{\bot\}) = \{B: \bot \leq B\}$, der zweitgrößte TW-Schnitt ist $T = \{B: \bot < B\}$, dann folgen Teilmengen von T, der zweitkleinste TW-Schnitt ist $Cn(\emptyset) = \{B: \top \leq B\}$ und der kleinste TW-Schnitt ist $\emptyset = \{B: \top < B\}$.
(d) Die Klasse der TW-Schnitte ist unter Durchschnitts- und Vereinigungsbildung abgeschlossen.
(e) Ein TW-Schnitt $S \neq \emptyset$ ist eine Theorie, sofern Cn kompakt ist.

Wir gehen im folgenden stets davon aus, daß Cn kompakt ist und daß $\langle T, \leq \rangle$ die Bedingungen (TW1)–(TW6) erfüllt. Als besonders nützlich werden sich die in Teil (b) von Lemma 3.5.2 erwähnten Schnitte herausstellen, weshalb wir ihnen einen eigenen Namen geben wollen:

3.5.3. Definition Ein TW-Schnitt S, für den es ein A gibt mit $S = S_A := \{B: A < B\}$ heißt ein *offen abgrenzbarer TW-Schnitt*; ein TW-Schnitt S, für den es ein A gibt mit $S = S_{\underline{A}} := \{B: A \leq B\}$ heißt ein *geschlossen abgrenzbarer TW-Schnitt.*

Man beachte, daß im allgemeinen Fall nicht alle TW-Schnitte (offen oder geschlossen) abgrenzbar sein müssen. Dies ist jedoch dann der Fall, wenn T finit ist (und zwar gilt für $S \neq Cn(\bot)$ dann $S = \{B: \wedge S \leq B\}$).

Nun wollen wir Definition 3.3.4(a) unter Verwendung des Schnittbegriffs umformulieren. Zunächst überlegt man sich leicht, daß unter der Voraussetzung von (TW5) das Definiens von Definition 3.3.4(a) äquivalent ist mit der Bedingung

$$\forall A \notin M' \; \exists B \notin M \; (A \leq B) \; .$$

Dann betrachte man die folgende Definition:

3.5.4. Definition Sei die Relation $\leq$ der theoretischen Wichtigkeit in T gegeben. Dann sei $\preceq = \mathbf{TB}(\leq)$ definiert durch die folgende Bestimmung: Für alle Satzmengen M und M′ gilt

$$M \preceq M' \Leftrightarrow \forall \text{ TW-Schnitte S } (S \subseteq M \Rightarrow S \subseteq M') \; .$$

Diese handliche Definition ist mit Definition 3.3.4(a) äquivalent:

3.5.5. Lemma Erfülle $\leq$ die Bedingungen (TW1) und (TW2). Dann gilt für alle Satzmengen M und M′:

$$\forall A \notin M' \; \exists B \notin M \; (A \leq B) \Leftrightarrow \forall \text{ TW-Schnitte S } (S \subseteq M \Rightarrow S \subseteq M') \; .$$

Wenn wir die Kontraktion von T bezüglich A (mit $\not\vdash A$) nach der alten Methode und bei Zugrundelegung der oben befürworteten Definition 3.3.4(a) konstruieren, sind wir nun also bei der Gleichung

$$T^-_A = \bigcap \{M \in T \bot A: \forall M' \in T \bot A \; \forall \text{ TW-Schnitte S } (S \subseteq M' \Rightarrow S \subseteq M)\}$$

angelangt. Diese Gleichung kann weiter vereinfacht werden, indem wir das

folgende Ergebnis zu Hilfe nehmen:

3.5.6. Lemma Erfülle $\leq$ die Bedingungen (TW1)–(TW6) und sei A ein beliebiger Satz. Dann gilt für jedes M in $T\perp A$:

$$(\forall M' \in T\perp A\ \forall \text{ TW-Schnitte } S\ (S \subseteq M' \Rightarrow S \subseteq M)) \Leftrightarrow S_A \subseteq M\ .$$

Damit reduziert sich der Vorschlag aus Abschnitt 3.3 zu

$$T^-_A = \bigcap \{M \in T\perp A\colon S_A \subseteq M\}\ .$$

Das letzte Teilchen des Puzzles wird durch das folgende Resultat geliefert:

3.5.7. Lemma Erfülle $\leq$ die Bedingungen (TW1)–(TW6) und sei A ein Satz derart, daß $\nvdash A$. Dann gilt:

$$\bigcap \{M \in T\perp A\colon S_A \subseteq M\} = T \cap \{B\colon A < A \vee B\}\ .$$

Die rechte Seite in der Gleichung dieses Lemmas gibt, wie man sich erinnert, genau das Definiens der Gärdenforsschen Definition 3.4.1 aus Abschnitt 3.4 wieder. Wir haben gezeigt, daß die Kontraktionen, die ausgehend von einer Relation der theoretischen Wichtigkeit auf dem ersten Weg (über die in Definition 3.3.4(a) vorgeschlagene Brücke zwischen $\leq$ und $\preceq$) und auf dem zweiten Weg konstruiert werden, identisch sind. Nach Zusammensetzung dieser Lemmata haben wir die gewünschte Antwort auf den offen gebliebenen Zusammenhang zwischen den beiden von Gärdenfors und seinen Mitautoren eingeführten Kontraktionsmethoden gefunden:

3.5.8. Theorem Sei $\langle T, \leq\rangle$ eine Theorie, welche (TW1)–(TW6) erfüllt, und sei $\mathbf{TB}(\leq)$ gemäß Definition 3.3.4(a) bestimmt. Dann gilt: Die $\mathbf{TB}(\leq)$-relationale PMCF (über T) ist identisch mit der $\leq$-ordnungsgemäßen Kontraktionsoperation.

Auch die „Umkehrung" von Theorem 3.5.8 ist gültig. Zur Vorbereitung benötigen wir in Fortsetzung der Lemmata 3.3.5 und 3.3.7 Eigenschaften von $\preceq$, die den Minimalitäts- und Maximalitätseigenschaften (TW5) und (TW6) von $\leq$ entsprechen. Die passenden Bedingungen sind

(TB5) $\forall M\ (\forall M'\ (M \preceq M') \Leftrightarrow \exists A \notin M\ (\vdash A))$ (Minimalität),

(TB6) $T \neq T_\perp \Rightarrow \forall M\ (\forall M' \neq T_\perp\ (M' \preceq M) \Leftrightarrow T \subseteq M)$ (Maximalität).

Indem wir $\preceq$ auf die ganze Klasse aller Satzmengen erweitern wollen, denken wir uns nun in Definition 3.3.4(b) die Einschränkung „$\in M(T)$" gestrichen und „$\in T \backslash M$" durch „$\notin M$" ersetzt. Dann gilt:

3.5.9. Lemma (a) Sei $\leq$ auf der Menge aller Sätze gegeben und $\preceq = \mathbf{TB}(\leq)$ auf der Klasse aller Satzmengen nach Definition 3.3.4(a) bestimmt. Erfülle $\leq$ die Bedingungen (TW1)–(TW3); wenn $\leq$ zusätzlich die Bedingung (TW5) erfüllt, so erfüllt $\preceq$ (TB6); wenn $\leq$ zusätzlich die Bedingung (TW6) erfüllt, so erfüllt $\preceq$ (TB5).

(b) Sei $\preceq$ auf der Klasse aller Satzmengen gegeben und $\leq = \mathbf{TW}(\preceq)$ auf

der Menge aller Sätze nach Definition 3.3.4(b) bestimmt. Erfülle $\preceq$ die Bedingungen (TB1) und (TB2); wenn $\preceq$ zusätzlich die Bedingung (TB5) erfüllt, so erfüllt $\leq$ (TW6); wenn $\preceq$ zusätzlich die Bedingung (TB6) erfüllt, so erfüllt $\leq$ (TW5).

Während wir die Minimalitätsbedingung (TB5) im folgenden nicht brauchen werden, ist die Maximalitätsbedingung (TB6) wichtig für den Beweis der „Umkehrung" von Theorem 3.5.8:

3.5.10. Theorem Sei $T \neq T_\perp$ eine Theorie und $\preceq$ eine Ordnung auf der Klasse aller Satzmengen, die (TB0)–(TB2) und (TB6) erfüllt, und sei $\mathbf{TW}(\preceq)$ gemäß Definition 3.3.4(b) bestimmt. Dann gilt: Die $\mathbf{TW}(\preceq)$-ordnungsgemäße Kontraktionsoperation (über T) ist identisch mit der $\preceq$-relationalen PMCF.

Es bleibt uns noch zu bemerken, daß TW-Schnitte auch bei der direkten Kontraktionskonstruktion aus Abschnitt 3.4 eine größere Rolle spielen, als es auf den ersten Blick offenbar wird. In der Definition 3.4.1, nach der T^-_A mit $T \cap \{B: A < A \vee B\}$ gleichzusetzen ist, haben wir ja keinen Schnitt vor uns liegen. Ein gewisser Nachteil haftet dieser Charakterisierung deshalb an, weil es auch zumindest intuitiv etwas undurchsichtig und schwierig ist, für alle Sätze B die Disjunktionen $A \vee B$ auf ihre theoretische Wichtigkeit hin zu prüfen. Man besitzt keine unmittelbare Vorstellung davon, welche Sätze dies wohl sein könnten. Auch hier erweisen sich (offen abgrenzbare) Schnitte als außerordentlich hilfreich. Wir definieren, daß bei vorgegebener theoretischer Wichtigkeit $\leq$ eine „Kontraktionsoperation" $\backslash$ (über T) genau dann $\leq$-*schnittweise* heißen soll, wenn $\backslash$ für alle A mit $\not\vdash A$ durch die Definition $T\backslash_A := S_A = \{B: A < B\}$ aus $\leq$ gewonnen wird und für A mit $\vdash A$ $T\backslash_A = T$ gesetzt ist. Eine Kontraktionsoperation heiße genau dann *schnittweise*, wenn es eine Relation $\leq$ der theoretischen Wichtigkeit, die (TW1)–(TW6) erfüllt, gibt, so daß $\backslash$ $\leq$-schnittweise ist. Es gilt dann das folgende

3.5.11. Lemma Sei T eine Theorie. Eine schnittweise Kontraktionsoperation $\backslash$ (über T) erfüllt die Gärdenfors-Postulate (T–1)–(T–4) und (T–6)–(T–8), aber im allgemeinen nicht (T–5).

Über die Gärdenfors-Postulate hinausgehend, haben schnittweise Kontraktionen folgende interessante Eigenschaften: Sie sind durch Inklusion geordnet, es gilt $T\backslash_{A \wedge B} = T\backslash_A$ oder $T\backslash_{A \wedge B} = T\backslash_B$ (der Nachweis über (TW1)–(TW4) ist einfach), schließlich gilt noch $T\backslash_{A \wedge B} = T\backslash_A \cup T\backslash_B$.[33] Lemma 3.5.11 macht deutlich, daß schnittweise Kontraktionsoperationen beinahe,

[33] Ich danke hier David Makinson, der mir geraten hat, diese Eigenschaften herauszustellen.

aber nicht ganz Kontraktionsoperationen im Sinne von Gärdenfors sind.[34] Das Verletzen des „Wiedergewinnungspostulats" (T–5) bedeutet ein Verletzen der Idee der minimalen Änderung bei Kontraktionen.[35] Bei schnittweisen Kontraktionen wird im allgemeinen zu viel aufgegeben (vgl. das Gegenbeispiel zu (T–5) im Beweis von Lemma 3.5.11). Es gilt das

3.5.12. Lemma Sei eine Relation $\leq$ der theoretischen Wichtigkeit gegeben, die (TW1)–(TW6) erfüllt. Dann gilt für alle Sätze A

$$S_A \subseteq T \cap \{B\colon A < A \vee B\} .$$

Mit Lemma 3.5.12 und dem Gegenbeispiel zu (T–5) ist klar, daß schnittweise Kontraktionsoperationen im allgemeinen weniger liefern als ordnungsgemäße. Besonders interessant ist jedoch aus unserer Perspektive, aus welcher Kontraktionen ja eigentlich nur Mittel zum Zwecke der Konstruktion von Revisionen durch die Levi-Identität sind, die folgende Beobachtung:

3.5.13. Theorem Sei T eine Theorie und $\leq$ eine Relation der theoretischen Wichtigkeit, die (TW1)–(TW6) erfüllt. Sei weiter $^-$ die $\leq$-ordnungsgemäße und sei $\backslash$ die $\leq$-schnittweise Kontraktionsoperation. Dann sind $^-$ und $\backslash$ revisionsäquivalent in dem Sinn, daß gilt: $\mathbf{L}(^-) = \mathbf{L}(\backslash)$.[36]

Wenn man also an Konstruktionen gemäß Definition 3.4.1 nur insoweit interessiert ist, als sie über (L) zur Konstruktion von Revisionen führen, so kann man ohne jede Verletzung der Idee der minimalen Änderung genauso gut (offen abgrenzbare) TW-Schnitte verwenden.

3.5.14. Definition Sei $\langle T, \leq \rangle$ eine Theorie. Dann ist die $\leq$-*ordnungsgemäße* Revisionsoperation $* = \mathbf{Rev}(\leq)$ definiert durch $* = \mathbf{L}(^-)$, wobei $^- = \mathbf{Kon}(\leq)$ die $\leq$-ordnungsgemäße Kontraktionsoperation ist.

Die eindeutige $\leq$-ordnungsgemäße Revision T^*_A von T, so stellt sich durch Theorem 3.5.13 heraus, ist gerade die Satzmenge $(\{B\colon \neg A < B\})^+_A$. Man braucht also keineswegs irgendwelche wahrheitsfunktionalen Zusammensetzungen zu vergleichen, sondern nur den offen abgrenzbaren TW-Schnitt aller Sätze, welche theoretisch wichtiger als $\neg A$ sind, herzuneh-

[34] Wohl aber ist eine schnittweise Kontraktionsoperation eine „withdrawal operation" im Sinne von Makinson (1987, S. 388).

[35] Hier muß man jedoch unterscheiden zwischen *hypothetischen* Kontraktionen und Kontraktionen, die auf neuer *faktischer Evidenz* beruhen. Isaac Levi hat mich davon überzeugt, daß seine vehemente Kritik an (T–5) in gewissen Fällen für hypothetische Kontraktionen berechtigt ist. Kontraktionen aufgrund neuer Evidenzen halte ich aber weiterhin für (T–5)-erfüllend. Dies ist ein Thema, das ausführlicherer Behandlung bedarf. Vgl. a. Kapitel 4, Fußnote 11, und Kapitel 6, Fußnote 19.

[36] Der Begriff der Revisionsäquivalenz und der Gedanke, daß es im allgemeinen verschiedene revisionsäquivalente Kontraktionsoperationen gibt, findet sich in Makinson (1987).

men, um nach anschließender Expansionsbildung durch A die gewünschte Revision T^*_A zu erhalten. Damit haben wir eine Möglichkeit gefunden, Revisionen auf praktikable und leicht faßliche Weise zu erzeugen.[37]

3.6 Ein Ansatz zu einer Dynamik der theoretischen Wichtigkeit

Für die wissenschaftstheoretische Fragestellung dieses Buchs genügt es, einmalige Revisionen in Betracht zu ziehen. Wir haben bis jetzt ein Modell betrachtet, mit dem der Übergang von einer Theorie T, als deren Bestandteil eine Ordnungsrelation $\leq$ der theoretischen Wichtigkeit aufgefaßt wird, zu einer revidierten Theorie T^*_A (auf zweifache, doch gleichwertige Art und Weise) expliziert werden konnte. Kurz gesagt, sind wir den Weg von $\langle T,\leq\rangle$ zu T^*_A gegangen. Doch für weitere hypothetische Ausflüge von T^*_A aus wird man auch für T^*_A eine Relation $\leq^*_A$ benötigen. Es stellt sich die Frage, ob und — wenn ja — wie man iterierte Revisionen bewerkstelligen kann oder, anders gesagt, wie man von $\langle T,\leq\rangle$ zu $\langle T^*_A,\leq^*_A\rangle$ kommmt. Wir skizzieren nun einen Weg in diese Richtung, der allerdings sein Ziel nicht ganz und gar erreicht. Die grundlegende Definition ist

3.6.1. Definition Sei $\langle T,\leq\rangle$ eine Theorie. Dann sei die Relation $\leq^*_A$ der theoretischen Wichtigkeit in T^*_A gegeben durch

$$B \leq^*_A C \Leftrightarrow A{\rightarrow}B \leq A{\rightarrow}C\ .$$

Ein Satz B ist nach der Revision durch A gemäß dieser Definition genau dann theoretisch wichtiger als ein Satz C, wenn das materiale Konditional $A{\rightarrow}C$ vor der Revision wichtiger war als das materiale Konditional $A{\rightarrow}B$. Wir stellen nun die wesentlichen Eigenschaften dieser Definition zusammen:

3.6.2. Theorem Sei $\langle T,\leq\rangle$ eine Theorie und sei $\leq^*_A$ durch Definition 3.6.1 gegeben. Dann gilt:
(a) Erfüllt $\leq$ (TW1), so erfüllt auch $\leq^*_A$ (TW1); erfüllt $\leq$ (TW2), so erfüllt auch $\leq^*_A$ (TW2); erfüllt $\leq$ (TW3), so erfüllt auch $\leq^*_A$ (TW3); erfüllt $\leq$ (TW4), so erfüllt auch $\leq^*_A$ (TW4).
(b) Ist A ein Satz derart, daß $\nvdash\neg A$, und erfüllt $\leq$ (TW1)-(TW4) und (TW6), so erfüllt $\langle T^*_A,\leq^*_A\rangle$ genau dann (TW5), wenn T^*_A die $\leq$-ordnungsgemäße Revision von T durch A ist.
(c) Ist A ein Satz derart, daß $\nvdash A$, und erfüllt $\leq$ (TW3), so erfüllt $\leq^*_A$ nicht

[37] Eine ebenfalls sehr übersichtliche Schreibweise von T^*_A für ein A mit $\nvdash\neg A$ geht aus dem Beweis von Theorem 3.5.13 hervor: $T^*_A=\{B:\neg A<\neg A\vee B\}$.

(TW6).

Die durch Definition 3.6.1 erhaltene Relation der theoretischen Wichtigkeit in T^*_A hat fast alle erwünschten Eigenschaften. Man beachte besonders Teil (b), der in einer Richtung besagt, daß die Vorschrift, ordnungsgemäße Revisionen durchzuführen, im wesentlichen aus der Definition 3.6.1 folgt, wenn $\leq^*_A$ die Bedingung der Minimalität erfüllt. Der einzige Punkt, der nicht ins Bild paßt, kommt im Teil (c) zum Ausdruck. Nun ist einerseits die Verletzung der Maximalität nicht so schlimm. (TW6) ist eine Grenzwert-Bedingung, die durchaus weggelassen werden könnte, wenn man an anderen Stellen des Revisionsmodells geeignete technische Änderungen anbringt.[38] Aus Symmetriegründen könnte man vielleicht argumentieren, daß es kontingente A's mit $A \leq \bot$ gibt, nämlich die nicht in T enthaltenen Sätze, warum also keine kontingenten A's mit $\top \leq A$? Kandidaten für Sätze mit maximaler theoretischer Wichtigkeit wären etwa analytische Sätze — was auch immer das sein mag — und zentrale, per Konvention als unfalsifizierbar gesetzte Grundaxiome einer wissenschaftlichen Theorie oder, besser, eines wissenschaftlichen Forschungsprogramms oder „Paradigmas" — infrage käme hier wohl das zweite Newtonsche Gesetz.

Andererseits ist die Art und Weise, wie in Definition 3.6.1 die Bedingung der Maximalität unterlaufen wird, doch sehr unbefriedigend.[39] Denn man sieht sofort, daß A bezüglich $\leq^*_A$ ein größtes Element von T^*_A ist, da ja $\vdash A \rightarrow A$ gilt und $\leq$ (TW3) erfüllen soll. Damit hat man irgendein A, durch welches eine Revision ausgelöst wird, in T^*_A auf eine Stufe mit logischen Wahrheiten gestellt (auch wenn A schon in T enthalten war). Dies mag für manche Zwecke — etwa für die Analyse von kontrafaktischen Konditionalsätzen — angehen, ist aber sicher nicht allgemein brauchbar. Eine besonders unangenehme Eigenschaft ist es dabei, daß A diesen Sonderstatus durch spätere Revisionen nie wieder los werden kann, was ebenfalls an Definition 3.6.1 liegt. Es kann zum Beispiel keine sehr plausible Aussage über $(T^*_A)^*_{\neg A}$ gemacht werden. Denn wenn A in T^*_A ein größtes Element bezüglich $\leq^*_A$ ist, dann ist die $\leq^*_A$-

[38] Man vergleiche etwa die „leere Wahrheit" („vacuous truth") von Konditionalsätzen der Form Wenn ¬A, dann B, die bei Lewis (1973a) auch dann gegeben sein kann, wenn $\nvdash A$. Vgl. Kapitel 4, Fußnote 1. Damit korrespondiert die mögliche Maximalität von solchen Sätzen A. Zur „Korrespondenz" von Lewis- und Gärdenfors-Modellen vgl. Gärdenfors (1979) und Grove (1988).

[39] Ein weiterer Punkt. Für $A \in T$, $A \neq \top$, gilt $\leq^* \neq \leq$, also insbesondere $\langle T^*_A, \leq^*_A \rangle \neq \langle T, \leq \rangle$, d.h. Theorien sind nicht „stabil" gegenüber Informationen, die sie bereits enthalten. Generell wird bei konsistenten Revisionen („Additionen") das $\leq$-Verhältnis von schon in T enthaltenen Sätzen verändert.

ordnungsgemäße Revision $(\{B: A <^*_A B\})^+_{\neg A}$ von T^*_A durch $\neg A$ einfach gleich $(\emptyset)^+_{\neg A} = Cn(\{\neg A\})$, während man intuitiv wohl eher $(T^*_A)^*_{\neg A} = T^*_{\neg A}$ erwarten wird.[40] Diese Schwierigkeit liegt nach meiner Überzeugung aber nicht an der ungeschickten Wahl von Definition 3.6.1, sondern an prinzipiellen Beschränkungen des rein relationalen Modells der Revision via theoretischer Wichtigkeit. Ganz analoge Schwierigkeiten gibt es übrigens in der Wahrscheinlichkeitstheorie bei der üblichen Konditionalisierung: Das Wahrscheinlichkeitsmaß P(.|A), definiert durch $P(B|A) = P(B \wedge A)/P(A)$ für alle B (Voraussetzung $P(A) \neq 0$!), weist A die maximale Wahrscheinlichkeit 1 zu, die durch weitere Konditionalisierungen nicht mehr erniedrigt werden kann.[41]

Eine Lösung bzw. Umgehung dieses Problems besteht in Richard Jeffreys (1965) „verallgemeinerter Konditionalisierung". Die Lösung des analogen Problems für das Revisionsmodell muß, so scheint es, die bloß relationale durch eine ordinale Struktur von Theorien ersetzen und ist in der Spohnschen (1988a; 1988b) Theorie der ordinalen Konditionalfunktionen bereits gefunden. Auf eine Darlegung dieser Theorie soll hier jedoch verzichtet werden, zumal die in diesem Abschnitt aufgezeigte Unzulänglichkeit des Gärdenforsschen Revisionsmodells für unsere wissenschaftstheoretischen Zwecke nicht relevant ist und durch seine Einfachheit und Eleganz durchaus wettgemacht wird.

3.7 Anhang: Beweise

3.1.3. Lemma Erfülle * die grundlegenden Gärdenfors-Postulate (T*1)–(T*6). Dann erfüllt * genau dann (T*7∧8), wenn * sowohl (T*7) als auch (T*8) erfüllt.

Beweis: Erfülle * die Gärdenfors-Postulate (T*1)–(T*6).

Die Behauptung ist erfüllt, wenn sowohl $\vdash \neg A$ als auch $\vdash \neg B$ gilt, da dann wegen (T*2) alle beteiligten Theorien inkonsistent sind. Gelte also $\nvdash \neg A$ oder $\nvdash \neg B$.

Um zu zeigen, daß (T*7) und (T*8) zusammen (T*7∧8) implizieren, machen wir eine Fallunterscheidung:

Sei im ersten Fall $\neg A \in T^*_{A \vee B}$, dann ist wegen (T*2) und (T*1) $\neg A \vee B \in T^*_{A \vee B}$, also wegen (T*5) $\neg(\neg A \vee B) \notin T^*_{A \vee B}$. Nach Definition

[40] Noch schlimmer steht es mit $(\leq^*_A)^*_{\neg A}$. Laut Definition 3.6.1 muß es die triviale Relation sein, die alle Sätze als gleich wichtig einstuft!

[41] Zur Dynamik von Wahrscheinlichkeitsmaßen vgl. Gärdenfors (1988, Kapitel 5).

3.1.2 und (T*7), (T*8) und (T*6) gilt dann $T^*_{A\vee B} = (T^*_{A\vee B})^+_{\neg A\vee B} = T^*_{(A\vee B)\wedge(\neg A\vee B)} = T^*_B$.
Sei im zweiten Fall $\neg B\in T^*_{A\vee B}$, dann bekommt man völlig analog $T^*_{A\vee B} = T^*_A$.
Sei im dritten Fall $\neg A,\neg B\notin T^*_{A\vee B}$. Nach Definition 3.1.2 folgt $T^*_{A\vee B} = (T^*_{A\vee B})^+\top = (T^*_{A\vee B})^+_{(A\vee\neg B)\vee(\neg A\vee B)}$ =(Disjunktion der Prämissen) $(T^*_{A\vee B})^+_{A\vee\neg B}\cap (T^*_{A\vee B})^+_{\neg A\vee B}$. Nach der Voraussetzung dieses Falls kann man nun (T*7) und (T*8) anwenden, und es folgt $T^*_{A\vee B} = T^*_{(A\vee B)\wedge(A\vee\neg B)}\cap T^*_{(A\vee B)\wedge(\neg A\vee B)}$, was nach (T*6) gleich $T^*_A\cap T^*_B$ ist.
Damit haben wir alle möglichen Fälle abgedeckt und gezeigt, daß $T^*_{A\vee B}$ gleich T^*_A oder T^*_B oder $T^*_A\cap T^*_B$ ist, d.h. daß (T*7∧8) gilt.
Es bleibt umgekehrt zu zeigen, daß (T*7∧8) sowohl (T*7) als auch (T*8) impliziert. Nach (T*6) ist $T^*_A = T^*_{(A\wedge B)\vee(A\wedge\neg B)}$, also folgt nach (T*7∧8), daß T^*_A gleich $T^*_{A\wedge B}$ oder $T^*_{A\wedge\neg B}$ oder $T^*_{A\wedge B}\cap T^*_{A\wedge\neg B}$ ist. Wegen (T*2) und Definition 3.1.2 ist dann $(T^*_A)^+_B$ gleich $(T^*_{A\wedge B})^+_B = T^*_{A\wedge B}$ oder gleich $(T^*_{A\wedge\neg B})^+_B = T_\perp$ oder gleich $(T^*_{A\wedge B}\cap T^*_{A\wedge\neg B})^+_B$= (Disjunktion der Prämissen) $(T^*_{A\wedge B})^+_B\cap(T^*_{A\wedge\neg B})^+_B = T^*_{A\wedge B}\cap T_\perp = T^*_{A\wedge B}$. Damit ist bereits gezeigt, daß $T^*_{A\wedge B}\subseteq(T^*_A)^+_B$, d.h. daß (T*7) gilt. Für (T*8) nehme man an, daß $\neg B\notin T^*_A$; dann ist nach Definition 3.1.2 $(T^*_A)^+_B\neq T_\perp$, also bleibt nur die Möglichkeit $(T^*_A)^+_B = T^*_{A\wedge B}$, und wir sind fertig. □

3.1.4. Lemma Erfülle $^-$ die Gärdenfors-Postulate (T-1)-(T-8). Dann gilt:
(a) $^-$ erfüllt auch
(T-I) $A\rightarrow B\in T^-_B \wedge B\rightarrow A\in T^-_A \Rightarrow T^-_A = T^-_B$;
(b) wenn A und B in einer Theorie T sind, ist $T^-_A = T^-_B$ genau dann, wenn $A\leftrightarrow B\in T^-_A\cap T^-_B$.

Beweis: Erfülle $^-$ die Gärdenfors-Postulate (T-1)-(T-8).
(a) Sei $B\rightarrow A\in T^-_A$ und $A\rightarrow B\in T^-_B$. Zu zeigen ist $T^-_A = T^-_B$.
Für den Fall $\vdash A$ und $\vdash B$ folgt die Behauptung aus (T-6).
Der Fall $\vdash A$ und $\nvdash B$ kann nicht eintreten, daß sonst aus $A\rightarrow B\in T^-_B$ mit (T-1) auch $B\in T^-_B$ folgen würde, im Widerspruch zu (T-4). Für den Fall $\nvdash A$ und $\vdash B$ gilt das Analoge.
Liege von nun an also der Fall $\nvdash A$ und $\nvdash B$ vor.
Wenn $A\notin T$, dann ist nach (T-3) $T^-_A = T$, also nach Voraussetzung $B\rightarrow A\in T$, also wegen (T-1) auch $B\notin T$, also wieder nach (T-3) $T^-_B = T$, also $T^-_A = T = T^-_B$. Für den Fall $B\notin T$ gilt das Analoge.

Sei von nun an also $A \in T$ und $B \in T$.
Wegen $B \rightarrow A \in T^-_A$ gilt $A \vee B \notin T^-_A$, da sonst nach (T-1) $A \in T^-_A$, im Widerspruch zu (T-4). Also folgt gemäß (T-6) und (T-8) $T^-_A = T^-_{(A \vee B) \wedge (A \vee \neg B)} \subseteq T^-_{A \vee B}$.
Um zu zeigen, daß andererseits $T^-_{A \vee B} \subseteq T^-_A$, bemerken wir zunächst, daß wegen (T-5) $A \rightarrow B \in T^-_A$, was zusammen mit der Voraussetzung $A \leftrightarrow B \in T^-_A$ ergibt, womit wegen dem schon gezeigten $T^-_A \subseteq T^-_{A \vee B}$ auch $A \leftrightarrow B \in T^-_{A \vee B}$ gilt. Wegen (T-6) gilt $T^-_A = T^-_{(A \vee B) \wedge (A \leftrightarrow B)}$, und es folgt nach (T-7) $T^-_{A \vee B} \cap T^-_{A \leftrightarrow B} \subseteq T^-_A$. Also auch $(T^-_{A \vee B} \cap T^-_{A \leftrightarrow B})^+_{A \leftrightarrow B} = (T^-_{A \vee B})^+_{A \leftrightarrow B} \cap (T^-_{A \leftrightarrow B})^+_{A \leftrightarrow B} \subseteq (T^-_A)^+_{A \leftrightarrow B}$. Da $A \leftrightarrow B \in T^-_{A \vee B}$, ist erstens $(T^-_{A \vee B})^+_{A \leftrightarrow B} = T^-_{A \vee B}$; da nach (T-2) und (T-5) $T^-_{A \vee B} \subseteq T \subseteq (T^-_{A \leftrightarrow B})^+_{A \leftrightarrow B}$, gilt zweitens $T^-_{A \vee B} \cap (T^-_{A \leftrightarrow B})^+_{A \leftrightarrow B} = T^-_{A \vee B}$; und da $A \leftrightarrow B \in T^-_A$ und somit drittens $(T^-_A)^+_{A \leftrightarrow B} = T^-_A$, ergibt alles zusammen $T^-_{A \vee B} \subseteq T^-_A$.
Damit haben wir gezeigt, daß T^-_A mit $T^-_{A \vee B}$ identisch ist. Aufgrund der totalen Symmetrie der Situation bzgl. A und B kann man ganz genauso zeigen, daß T^-_B mit $T^-_{A \vee B}$ identisch ist. Deshalb gilt T^-_A gleich T^-_B, und wir sind fertig.
(b) Seien A und B in T. Daß aus $A \leftrightarrow B \in T^-_A \cap T^-_B$ schon $T^-_A = T^-_B$ folgt, haben wir im Beweis von Teil (a) gezeigt. Umgekehrt folgt für $A, B \in T$ aus (T-5), daß $A \rightarrow B \in T^-_A$ und $B \rightarrow A \in T^-_B$, das heißt mit $T^-_A = T^-_B$ gerade $A \leftrightarrow B \in T^-_A \cap T^-_B$. □

3.1.6. Lemma Erfülle $^-$ die grundlegenden Gärdenfors-Postulate (T-1)-(T-6) und sei $^* = \mathbf{L}(^-)$. Sei weiter $T \neq T_\perp$ und A so, daß $\not\vdash A$ und $\not\vdash \neg A$. Dann gilt:
$T^*_A \cap T^*_{\neg A} = T^-_A \cap T^-_{\neg A} \subseteq T = T^-_A \cup T^-_{\neg A} \subseteq T^*_A \cup T^*_{\neg A}$.

Beweis: Wegen (T-2), (T-3) und der Levi-Identität (L) sind die drei hinteren Beziehungen trivial. Zu zeigen bleibt die Inklusion $T^*_A \cap T^*_{\neg A} \subseteq T^-_A \cap T^-_{\neg A}$. Sei hierzu $B \in T^*_A \cap T^*_{\neg A}$. Mit (L) folgt $B \in (T^-_{\neg A})^+_A \cap (T^-_A)^+_{\neg A}$, d.h. nach Definition 3.1.2 (und unseren Voraussetzungen an Cn) gilt $A \rightarrow B \in T^-_{\neg A}$ und $\neg A \rightarrow B \in T^-_A$. Da nach (T-5) aber außerdem $\neg A \rightarrow B \in T^-_A$, folgt mit (T-1) sofort $B \in T^-_{\neg A}$ und $B \in T^-_A$. □

3.2.1. Theorem Sei $\langle T, \leq \rangle$ eine Theorie, welche (TW1)-(TW4) und ($TW4^D$) erfüllt. Dann ist $\langle T, \leq \rangle$ insofern trivial, als gilt: Es gibt keine zwei normalwichtigen Sätze A und B von T mit $A \neq B$.

Beweis: Erfülle $\langle T, \leq \rangle$ (TW1)-(TW4) und ($TW4^D$), und seien A und B Sätze in T mit $A < C$ und $B < C$ für ein C mit $\vdash C$.

(TW3) sagt, daß $A \doteq (A \vee B) \wedge (A \vee \neg B)$, und wegen (TW4) und (TW1) folgt $A \vee B \leq A$ oder $A \vee \neg B \leq A$.
Nach Lemma 3.2.2 ist aber $A \leftrightarrow B$ maximalwichtig, also ist wegen $A \leftrightarrow B \vdash A \vee \neg B$ und (TW3) auch $A \vee \neg B$ maximalwichtig, also gilt $A < A \vee \neg B$.
Deshalb muß $A \vee B \leq A$ richtig sein.
Völlig analog bekommt man auch $A \vee B \leq B$.
Andererseits gilt nach (TW3) $A \leq A \vee B$ und $B \leq A \vee B$.
Also gilt $A \doteq A \vee B \doteq B$. □

3.2.2. Lemma Sei $\langle T, \leq \rangle$ eine Theorie, welche (TW1), (TW2), (TW3) und ($TW4^D$) erfüllt. Dann gilt: Sind A und B normalwichtige Sätze von T, so ist die materiale Äquivalenz $A \leftrightarrow B$ ein maximalwichtiger Satz von T.

Beweis: Erfülle $\langle T, \leq \rangle$ (TW1), (TW2), (TW3) und ($TW4^D$), und seien A und B Sätze in T mit $A < C$ und $B < C$ für ein C mit $\vdash C$.
Es gilt $\vdash (A \vee B) \vee (A \leftrightarrow B)$, und nach ($TW4^D$) gilt $(A \vee B) \vee (A \leftrightarrow B) \leq A \vee B$ oder $(A \vee B) \vee (A \leftrightarrow B) \leq A \leftrightarrow B$.
Der erste Fall kann aber nicht eintreten, denn nach Voraussetzung und (TW3) gilt $A < (A \vee B) \vee (A \leftrightarrow B)$ und $B < (A \vee B) \vee (A \leftrightarrow B)$, nach (TW1) gälte dann also $A < A \vee B$ und $B < A \vee B$, im Widerspruch zu ($TW4^D$).
Also gilt der zweite Fall, und damit gibt es ein C mit $\vdash C$ und $C \leq A \leftrightarrow B$, das heißt $A \leftrightarrow B$ ist maximalwichtig. □

3.3.5. Lemma (a) Sei $\leq$ (auf T) gegeben und $\preceq = \mathbf{TB}(\leq)$ nach Definition 3.3.4(a) bestimmt; wenn $\leq$ die Bedingungen (TW1) und (TW2) erfüllt, dann erfüllt $\preceq$ auf M(T) (TB1) und (TB2);
(b) sei $\preceq$ (auf M(T)) gegeben und $\leq = \mathbf{TW}(\preceq)$ nach Definition 3.3.4(b) bestimmt; wenn $\preceq$ (TB1) und (TB2) erfüllt, dann erfüllt $\leq$ die Bedingungen (TW1)-(TW4).

Beweis: (a) Sei $\preceq = \mathbf{TB}(\leq)$ nach Definition 3.3.4(a), und erfülle $\leq$ die Bedingungen (TW1)-(TW2). Zu zeigen: $\preceq$ ist transitiv und konnex.
Transitivität: Seien M,M',M'' in M(T) und gelte $M \preceq M'$ und $M' \preceq M''$, d.h. nach Definition 3.3.4(a) $\forall A \in T \backslash M'\ \exists B \in T \backslash M\ (A \leq B)$ und $\forall C \in T \backslash M''\ \exists D \in T \backslash M'\ (C \leq D)$. Durch Einsetzen von D für A und Anwenden von (TW1) sieht man, daß $\forall C \in T \backslash M''\ \exists B \in T \backslash M\ (C \leq B)$, d.h. nach Definition 3.3.4(a) $M \preceq M''$.
Konnexität: Seien M und M' in M(T) und gelte $M \not\preceq M'$, d.h. nach Definition 3.3.4(a) $\exists A \in T \backslash M'\ \forall B \in T \backslash M\ (A \not\leq B)$, woraus wegen (TW2) $\exists A \in T \backslash M'\ \forall B \in T \backslash M\ (B \leq A)$ folgt. Dies impliziert quantorenlogisch $\forall B \in T \backslash M\ \exists A \in T \backslash M'\ (B \leq A)$, d.h. nach Definition 3.3.4(a) $M' \preceq M$.

(b) Sei $\leq = \mathbf{TW}(\preceq)$ nach Definition 3.3.4(b), und erfülle $\preceq$ die Bedingungen (TB1) und (TB2). Zu zeigen: $\leq$ erfüllt (TW1)-(TW4).
(TW1): Seien A,B,C$\in$T und gelte A$\leq$B und B$\leq$C, d.h. nach Definition 3.3.4(b) $\forall M \not\ni B\ \exists M' \not\ni A$ ($M \preceq M'$) und $\forall M'' \not\ni C\ \exists M''' \not\ni B$ ($M'' \preceq M'''$). Durch Einsetzen von M''' für M und Anwenden von (TB1) sieht man, daß $\forall M'' \not\ni C\ \exists M' \not\ni A$ ($M'' \preceq M'$) , d.h. nach Definition 3.3.4(b) A$\leq$C.
(TW2): Folgt aus (TW1), (TW3) und (TW4).
(TW3): Seien A und B in T mit A$\vdash$B. Da alle Elemente von M(T) abgeschlossen unter Cn sind, gilt für alle M$\in$M(T) mit B$\notin$M auch A$\notin$M. Da $\preceq$ wegen (TB2) reflexiv ist, gibt es also für jedes M$\in$M(T) mit B$\notin$T ein M'$\in$M(T) mit A$\notin$M' und M$\preceq$M', nämlich M'=M. Also folgt nach Definition 3.3.4(b) A$\leq$B.
(TW4): Seien A und B in T. Angenommen, es gilt A$\not\leq$A$\wedge$B und B$\not\leq$A$\wedge$B, d.h. nach Definition 3.3.4(b) $\exists M_1 \not\ni A \wedge B\ \forall M' \not\ni A$ ($M_1 \not\preceq M'$) und $\exists M_2 \not\ni A \wedge B$ $\forall M'' \not\ni B$ ($M_2 \not\preceq M''$) . Da $\preceq$ wegen (TB2) reflexiv ist, muß A$\in M_1$ und B$\in M_2$ gelten. Da aber A$\wedge$B weder in M_1 noch in M_2 ist und da M_1 und M_2 Theorien sind, gilt B$\notin M_1$ und A$\notin M_2$. Also folgt $M_1 \not\preceq M_2$ und $M_2 \not\preceq M_1$, im Widerspruch zu (TB2). Also gilt (TW4). □

3.3.6. Theorem Es gilt bei Zugrundelegung von Definition 3.3.4:
(a) Wenn $\leq$ auf T die Bedingungen (TW1)-(TW4) erfüllt, dann ist $\mathbf{TW}(\mathbf{TB}(\leq))$ identisch mit $\leq$;
(b) wenn T eine finite (oder vollständige) Theorie und wenn $\preceq$ eine Relation auf M(T) ist, dann ist $\mathbf{TB}(\mathbf{TW}(\preceq))$ identisch mit $\preceq$.

Beweis: (a) Erfülle $\leq$ (TW1)-(TW4), sei $\preceq = \mathbf{TB}(\leq)$ und $\leq^* = \mathbf{TW}(\mathbf{TB}(\leq))$ und seien A und B in T. Wir zeigen: A$\leq^*$B $\Leftrightarrow$ A$\leq$B .
„$\Rightarrow$“: Sei A$\leq^*$B, d.h. nach Definition 3.3.4(b)

$\forall M \not\ni B\ \exists M' \not\ni A$ ($M \preceq M'$) , d.h. nach Definition 3.3.4(a)

$\forall M \not\ni B\ \exists M' \not\ni A\ \forall C \in T \setminus M'\ \exists D \in T \setminus M$ (C$\leq$D) .

Durch Einsetzen von A für C erhält man

$\forall M \not\ni B\ \exists D \in T \setminus M$ (A$\leq$D) . (*)

Betrachte nun die Menge $S_B := \{E: B<E\}$. Es gilt $S_B \not\vdash B$: Denn sonst gäbe es eine endliche Menge von Sätzen $E_1, \ldots, E_n \in S_B$ mit $E_1 \wedge \ldots \wedge E_n \vdash B$, was nach (TW3) $E_1 \wedge \ldots \wedge E_n \leq B$ implizieren würde; mehrfache Anwendung von (TW4) und (TW1) ergibt $E_i \leq E_1 \wedge \ldots \wedge E_n$ für ein $i \in \{1, \ldots, n\}$, also folgt mit (TW1) $E_i \leq B$ für ein $i \in \{1, \ldots, n\}$, im Widerspruch zu $E_i \in S_B$.
Da nun $S_B \not\vdash B$, gibt es (mit Auswahlaxiom) ein $M_0 \in T \perp B \subseteq M(T)$ derart, daß $S_B \subseteq M_0$. Für M_0 gilt: $\forall D \in T \setminus M_0 (D \leq B)$.
Da B$\notin M_0$, ergibt dies zusammengenommen mit (*) unter Verwendung von (TW1)

$A \leq B$.

„$\Leftarrow$“: Sei $A \not\leq^* B$, d.h. nach Definition 3.3.4(b)

$\exists M \not\supseteq B\ \forall M' \not\supseteq A\ (M \not\preceq M')$, d.h. nach Definition 3.3.4(a)

$\exists M \not\supseteq B\ \forall M' \not\supseteq A\ \exists C \in T \backslash M'\ \forall D \in T \backslash M\ (C \not\leq D)$.

Wegen (TW2) ist dies äquivalent mit

$\exists M \not\supseteq B\ \forall M' \not\supseteq A\ \exists C \in T \backslash M'\ \forall D \in T \backslash M\ (D < C)$.

Betrachte wieder die o.g. Menge S_B und ein $M_0 \in M(T)$ mit $B \notin M_0$ und $S_B \subseteq M_0$. Für M_0 gilt: $\forall D \in T \backslash M_0 (D \leq B)$.

Offenbar gilt für jedes M mit $B \notin M$ $\exists D \in T \backslash M\ (D = B)$ — nimm einfach B für D ($\leq$ ist wegen (TW2) reflexiv); somit hat M_0 unter allen $M \not\supseteq B$ die unwichtigsten fehlenden Sätze (d.h. $\forall M \not\supseteq B\ \exists D \in T \backslash M\ \forall E \in T \backslash M_0\ (E \leq D)$), und es gilt

$\forall M' \not\supseteq A\ \exists C \in T \backslash M'\ \forall D \in T \backslash M_0\ (D < C)$.

Setze nun B für D ein:

$\forall M' \not\supseteq A\ \exists C \in T \backslash M'\ (B < C)$.

Betrachte nun $S_A := \{E: A < E\}$ und ein $M_0' \in T \perp A \subseteq M(T)$ mit $S_A \subseteq M_0'$ (so ein M_0' existiert, wie im Teil „$\Rightarrow$“ gezeigt). Da $A \notin M_0'$, gilt

$\exists C \in T \backslash M_0'\ (B < C)$,

und da $\forall C \in T \backslash M_0'\ (C \leq A)$, folgt unter Verwendung von (TW1)

$B < A$, d.h. $A \not\leq B$.

(b) Sei T eine finite (oder vollständige) Theorie.

Wir benötigen für den Beweis eine kleine Vorbereitung. Sei für $M \in M(T)$ $F_M = \bigvee \{A: A \in T \backslash M\}$ die Disjunktion der ein M fehlenden Sätze (bzw. im finiten Fall eigentlich genauer: die Disjunktion von Repräsentanten der entsprechenden Äquivalenzklassen bzgl. Cn). $F_T = \bigvee \emptyset$ definieren wir als $\perp$. F_M ist für finite (und vollständige) Theorien T ein Satz der Objektsprache und ist für $M \neq T$ auch in T enthalten. Es gilt das

Hilfslemma Sei $T \neq T_\perp$ eine finite (oder vollständige) Theorie und sei $M \in M(T)$ mit $M \neq T$. Dann gilt:

(i)$\forall\ A \in T\ (A \in T \backslash M \Leftrightarrow A \vdash F_M)$.

(ii)$F_M \in T \backslash M$.

(iii)$\forall\ M' \in M(T)\ (F_M \in T \backslash M' \Rightarrow M' = M)$.

Beweis des Hilfslemmas: (i) „$\Rightarrow$“: Wenn $A \in T \backslash M$, ist $A \vdash F_M$ klar nach der Definition von F_M und den Regeln der klassischen (bzw. vollständigen) Logik.

„$\Leftarrow$“: Gelte $A \vdash F_M$ und sei ohne Einschränkung der Allgemeinheit $M \in T \perp B$. Es gilt wegen der Maximalität von M für alle $C \in T \backslash M$ $M \cup \{C\} \vdash B$, deshalb gilt nach den Regeln der klassischen (bzw. vollständigen) Logik auch $M \cup F_M \vdash B$. Also kann A nicht in M sein, da sonst $M \vdash F_M$ und daher $M \vdash B$,

im Widerspruch zu $M \in T \perp B$.
(ii) Folgt unmittelbar aus (i).
(iii) Sei $M' \in M(T)$ mit $F_M \in T \backslash M'$; nach (i) gilt dann $F_M \vdash F_{M'}$.
Angenommen, es gibt ein $A \in M' \backslash M$. Das heißt nach (i) $A \vdash F_M$ und $A \not\vdash F_{M'}$, im Widerspruch zu $F_M \vdash F_{M'}$.
Also gilt $M' \subseteq M$.
Angenommen, es gibt ein $A \in M \backslash M'$. Sei ohne Einschränkung der Allgemeinheit $M \in T \perp B$ und $M' \in T \perp C$. Wegen $A \notin M'$ und der Maximalität von M' gilt $A \rightarrow C \in M' \subseteq M$. Also ist neben A auch $A \rightarrow C$ in M, also $C \in M$, also $C \rightarrow B \notin M$ wegen $M \in T \perp B$. Dann muß aber auch $C \rightarrow B \notin M'$ gelten, d.h. wegen der Maximalität von M' $(C \rightarrow B) \rightarrow C \in M'$, und dies heißt — da $M' \in M(T)$ eine Theorie ist! — nichts anderes als $C \in M'$, im Widerspruch zu $M' \in T \perp C$.
Also gilt $M' = M$. □

Wir kommen jetzt zum eigentlichen Beweis von Theorem 3.3.6(b).
Erfülle $\preceq$ (TB1) und (TB2), sei $\leq = \mathbf{TW}(\preceq)$ und $\preceq^* = \mathbf{TB}(\mathbf{TW}(\preceq))$ und seien M und M' in M(T). Wir zeigen: $M \preceq^* M' \Leftrightarrow M \preceq M'$.
„$\Rightarrow$“:Sei $M \preceq^* M'$, d.h. nach Definition 3.3.4(a)

$$\forall A \in T \backslash M' \; \exists B \in T \backslash M \; (A \leq B) \text{ , d.h. nach Definition 3.3.4(b)}$$
$$\forall A \in T \backslash M' \; \exists B \in T \backslash M \; \forall M'' \not\ni B \; \exists M''' \not\ni A \; (M'' \preceq M''') .$$

Nach Einsetzen von M für M'' ergibt sich

$$\forall A \in T \backslash M' \; \exists M''' \not\ni A \; (M \preceq M''') .$$

Man nehme speziell $F_{M'}$ für A ($F_{M'} \in T \backslash M'$ nach dem Hilfslemma, Teil (ii)) und erhält

$$\exists M''' \not\ni F_{M'} \; (M \preceq M''') .$$

Nach dem Hilfslemma, Teil (iii) ist aber jedes $M''' \not\ni F_{M'}$ gleich M', also

$$M \preceq M' .$$

„$\Leftarrow$“:Sei $M \not\preceq^* M'$, d.h. nach Definition 3.3.4

$$\exists A \in T \backslash M' \; \forall B \in T \backslash M \; \exists M'' \not\ni B \; \forall M''' \not\ni A \; (M'' \not\preceq M''') .$$

Man nehme speziell F_M für B ($F_M \in T \backslash M$ nach dem Hilfslemma, Teil (ii)) und beachte, daß nach dem Hilfslemma, Teil (iii) für $B = F_M$ schon $M'' = M$ ist, und so erhält man

$$\exists A \in T \backslash M' \; \forall M''' \not\ni A \; (M \not\preceq M''') .$$

Nach Einsetzen von M' für M''' ergibt sich

$$M \not\preceq M' . \; \square$$

3.3.7. Lemma Sei $\leq$ (auf T) gegeben und $\preceq = \mathbf{TB}(\leq)$ nach Definition 3.3.4(a) bestimmt; wenn $\leq$ die Bedingungen (TW1) und (TW2) erfüllt, dann erfüllt $\preceq$ auf Pot(T) (TB1), (TB2),

(TB3) $\forall M,M' (M \subseteq M' \Rightarrow M \preceq M')$ (Dominanz) und
(TB4) $\forall M,M' (M \preceq M \cap M' \vee M' \preceq M \cap M')$ (Durchschnitt).
Beweis: Sei $\preceq = \mathbf{TB}(\leq)$ nach Definition 3.3.4(a) und erfülle $\leq$ die Bedingungen (TW1) und (TW2).
(TB1) und (TB2) auf Pot(T) zeigt man wie im Beweis von Lemma 3.3.5(a).
(TB3): Seien M und M' beliebige Teilmengen von T mit $M \subseteq M'$. Angenommen $M \not\preceq M'$, d.h. nach Definition 3.3.4(a) $\exists A \in T \setminus M' \; \forall B \in T \setminus M \; (A \not\leq B)$. Wegen der Reflexivität von $\leq$ (die aus (TW2) folgt), folgt $\exists A \in T \setminus M' \; (A \notin T \setminus M)$, d.h. $\exists A \in M \setminus M'$, d.h. $M \not\subseteq M'$, im Widerspruch zur Voraussetzung.
(TB4): Seien M und M' beliebige Teilmengen von T. Angenommen $M \not\preceq M \cap M'$ und $M' \not\preceq M \cap M'$, d.h. nach Definition 3.3.4(a) $\exists A_1 \in T \setminus M \cap M' \; \forall B \in T \setminus M \; (A_1 \not\leq B)$ und $\exists A_2 \in T \setminus M \cap M' \; \forall C \in T \setminus M' \; (A_2 \not\leq C)$. Seien A_1 und A_2 solche Sätze, und sei ohne Einschränkung der Allgemeinheit $A_1 \leq A_2$ (der Fall $A_2 \leq A_1$ geht analog, und ein dritter Fall ist wegen (TW2) ausgeschlossen). Dann gilt wegen (TW1) $\forall B \in T \setminus M \; (A_2 \not\leq B)$ und $\forall C \in T \setminus M' \; (A_2 \not\leq C)$, d.h. $\forall D \in (T \setminus M) \cup (T \setminus M') \; (A_2 \not\leq D)$, d.h. $\forall D \in T \setminus (M \cap M') \; (A_2 \not\leq D)$. Da aber $A_2 \in T \setminus M \cap M'$, steht dies im Widerspruch zur Reflexivität von $\leq$. □

3.5.2. Lemma Sei $\langle T, \leq \rangle$ eine Theorie, die (TW1)–(TW6) erfüllt. Dann gilt:
(a) Für alle TW-Schnitte S und S' ist $S \subseteq S'$ oder $S' \subseteq S$.
(b) Satzmengen der Form $\{B: A < B\}$ und der Form $\{B: A \leq B\}$ sind TW-Schnitte.
(c) Der größte TW-Schnitt ist $Cn(\bot) = \{B: \bot \leq B\}$, der zweitgrößte TW-Schnitt ist $T = \{B: \bot < B\}$, dann folgen Teilmengen von T, der zweitkleinste TW-Schnitt ist $Cn(\emptyset) = \{B: \top \leq B\}$ und der kleinste TW-Schnitt ist $\emptyset = \{B: \top < B\}$.
(d) Die Klasse der TW-Schnitte ist unter Durchschnitts- und Vereinigungsbildung abgeschlossen.
(e) Ein TW-Schnitt $S \neq \emptyset$ ist eine Theorie, sofern Cn kompakt ist.

Beweis: (a) Seien S und S' Schnitte. Angenommen $S \not\subseteq S'$ und $S' \not\subseteq S$, d.h. $\exists A \in S \setminus S'$ und $\exists B \in S' \setminus S$. Da S ein Schnitt ist, muß $A \not\leq B$ gelten, und da S' ein Schnitt ist, muß $B \not\leq A$ gelten, im Widerspruch zu (TW2).
(b) $S = \{B: A < B\}$ ist ein Schnitt: Sei $C \in S$ und $C \leq D$. Wegen $C \in S$ gilt $A < C$, also wegen (TW1) $A < D$, also $D \in S$. $S = \{B: A \leq B\}$ geht analog.
(c) Folgt aus (TW3), (TW5) und (TW6).
(d) Sei $\mathbf{S}$ eine Menge von Schnitten. Betrachte $\bigcap \mathbf{S}$: Sei $A \in \bigcap \mathbf{S}$ und $A \leq B$. Da $A \in \bigcap \mathbf{S}$, ist $A \in S$ für alle $S \in \mathbf{S}$, also ist $B \in S$ für alle $S \in \mathbf{S}$, also ist $B \in \bigcap \mathbf{S}$,

also ist $\bigcap$S ein Schnitt. Betrachte $\bigcup$S: Sei A$\in\bigcup$S und A$\leq$B. Da A$\in\bigcup$S, ist A$\in$S für ein S$\in$S, also ist B$\in$S für ein S$\in$S, also ist B$\in\bigcup$S, also ist $\bigcup$S ein Schnitt.
(e) Sei S ein Schnitt und gelte S$\vdash$A. Wenn Cn kompakt ist, so gibt es $B_1,\ldots,B_n\in S$ mit $B_1\wedge\ldots\wedge B_n\vdash A$, also mit (TW3) $B_1\wedge\ldots\wedge B_n\leq A$. Durch mehrmalige Anwendung von (TW4) und (TW1) erhält man aber $B_i\leq B_1\wedge\ldots\wedge B_n$ für ein i$\in\{1,\ldots,n\}$, also mit (TW1) $B_i\leq A$ für ein i$\in\{1,\ldots,n\}$. Da $B_i\in S$ und S Schnitt, ist A$\in$S. □

3.5.5. Lemma Erfülle $\leq$ die Bedingungen (TW1) und (TW2). Dann gilt für alle Satzmengen M und M':

$$\forall A\notin M'\ \exists B\notin M\ (A\leq B) \Leftrightarrow \forall \text{ TW-Schnitte S } (S\subseteq M \Rightarrow S\subseteq M')\ .$$

Beweis: Erfülle $\leq$ (TW1) und (TW2) und seien M und M' beliebige Satzmengen. Wir zeigen

$$\neg\ (\forall A\notin M'\ \exists B\notin M\ (A\leq B)) \Leftrightarrow \neg\ (\forall \text{ TW-Schnitte S } (S\subseteq M \Rightarrow S\subseteq M'))\ .$$

„$\Rightarrow$“: Gelte $\exists A\notin M'\ \forall B\notin M\ (A\not\leq B)$, d.h. $\exists A\notin M'\ (\{B{:}A\leq B\}\subseteq M)$. Daraus folgt wegen der Reflexivität von $\leq$ (die aus (TW2) folgt)

$$\exists A\ (\{B{:}A\leq B\}\not\subseteq M' \wedge \{B{:}A\leq B\}\subseteq M)\ \text{, d.h. } \exists S_{\underline{A}}(S_{\underline{A}}\subseteq M \wedge S_{\underline{A}}\not\subseteq M')\ ,$$

also nach Lemma 3.5.2(b) insbesondere $\exists$ TW-Schnitt S (S$\subseteq$M $\wedge$ S$\not\subseteq$M').
„$\Leftarrow$“: Sei S ein TW-Schnitt mit S$\subseteq$M und S$\not\subseteq$M'. Dann gibt es ein A$\in$S mit A$\in$M$\setminus$M', und es gilt wegen (TW1) $S_{\underline{A}}=\{B{:}A\leq B\}\subseteq M$. Dies heißt aber $\exists A\notin M'\ (\{B{:}A\leq B\}\subseteq M)$, was mit $\exists A\notin M'\ \forall B\notin M\ (A\not\leq B)$ äquivalent ist. □

3.5.6. Lemma Erfülle $\leq$ die Bedingungen (TW1)–(TW6) und sei A ein beliebiger Satz. Dann gilt für jedes M in T$\perp$A:

$$(\forall M'\in T\perp A\ \forall \text{ TW-Schnitte S } (S\subseteq M' \Rightarrow S\subseteq M)) \Leftrightarrow S_A\subseteq M\ .$$

Beweis: Die Behauptung ist trivial für ein A mit $\vdash$A, da dann T$\perp$A=$\emptyset$. Gelte also $\not\vdash$A, d.h. mit (TW6) $S_A\neq\emptyset$.
„$\Rightarrow$“: Gelte $\forall M'\in T\perp A\ \forall$ TW-Schnitte S (S$\subseteq$M' $\Rightarrow$ S$\subseteq$M).
Da wegen der Reflexivität von $\leq$ A$\notin S_A$ ist und da wegen Lemma 3.5.2(e) S_A eine Theorie ist, gilt $S_A\not\vdash$A. Also gibt es (mit Auswahlaxiom) ein M'$\in$T$\perp$A mit $S_A\subseteq M'$. Da S_A ein T-Schnitt ist, gilt wegen der Voraussetzung auch $S_A\subseteq M$.
„$\Leftarrow$“: Gelte $S_A\subseteq M\in T\perp A$. Angenommen, es gibt ein M'$\in$T$\perp$A und einen TW-Schnitt S mit S$\subseteq$M' und S$\not\subseteq$M. Da $S_A\subseteq M$ und S$\not\subseteq$M, gibt es ein B$\in$S$\setminus S_A$. Da S ein TW-Schnitt ist, muß $S_{\underline{B}}:=\{C{:}\ B\leq C\}$ eine Teilmenge von S sein. Da aber B$\notin S_A$, gilt mit (TW2) B$\leq$A, also A$\in S_B\subseteq$S$\subseteq$M' , im

Widerspruch zu $M' \in T \perp A$. □

3.5.7. Lemma Erfülle $\leq$ die Bedingungen (TW1)-(TW6) und sei A ein Satz derart, daß $\not\vdash A$. Dann gilt:

$$\bigcap \{M \in T \perp A: S_A \subseteq M\} = T \cap \{B: A < A \vee B\} .$$

Beweis: Sei A so, daß $\not\vdash A$. Dies stellt sicher, daß $\{M \in T \perp A: S_A \subseteq M\} \neq \emptyset$ und wegen (TW6) auch, daß $S_A \neq \emptyset$.
„$\subseteq$": Sei $B \notin T \cap \{B: A < A \vee B\}$, d.h. $B \notin T$ oder $A \vee B \leq A$.
Im ersten Fall ist klar, daß $B \notin \bigcap \{M \in T \perp A: S_A \subseteq M\}$, da hier für alle $M \in T \perp A$ $B \notin M$ gilt.
Sei also $B \in T$ und gelte $A \vee B \leq A$. Zu zeigen: $\exists M \in T \perp A\ (S_A \subseteq M \wedge B \notin M)$.
Wegen der Maximalität der $M \in T \perp A$ ist $B \notin M$ äquivalent mit $B \rightarrow A \in M$, d.h. wir müssen zeigen, daß es ein $M \in T \perp A$ gibt mit $S_A \cup \{B \rightarrow A\} \subseteq M$. Dafür reicht es (wegen dem Auswahlaxiom) zu zeigen, daß $S_A \cup \{B \rightarrow A\} \not\vdash A$, äquivalent $S_A \not\vdash (B \rightarrow A) \rightarrow A$ oder einfach $S_A \not\vdash A \vee B$. Nach unserer Voraussetzung war aber $A \vee B \leq A$, d.h. $A \vee B \notin S_A$, und da nach Lemma 3.5.2(e) $S_A \neq \emptyset$ eine Theorie ist, heißt dies gerade $S_A \not\vdash A \vee B$.
„$\supseteq$": Sei $B \in T \cap \{B: A < A \vee B\}$ und sei $M \in T \perp A$ mit $S_A \subseteq M$. Zu zeigen: $B \in M$. Angenommen aber, daß $B \notin M$, dann ist wegen $B \in T$ und der Maximalität von M $B \rightarrow A \in M$, d.h. $\neg B \vee A \in M$. Wegen $A < A \vee B$ ist andererseits $A \vee B \in S_A \subseteq M$, also gilt $M \vdash A$, im Widerspruch zu $M \in T \perp A$. □

3.5.8. Theorem Sei $\langle T, \leq \rangle$ eine Theorie, welche (TW1)-(TW6) erfüllt, und sei $\mathbf{TB}(\leq)$ gemäß Definition 3.3.4(a) bestimmt. Dann gilt: Die $\mathbf{TB}(\leq)$-relationale PMCF (über T) ist identisch mit der $\leq$-ordnungsgemäßen Kontraktionsoperation.

Beweis: Für $\vdash A$ setzt (RPM) $T^-_A = T$ und Definition 3.4.1 ebenfalls $T^-_A = T$.
Für $\not\vdash A$ wird T^-_A nach (RPM) und Definition 3.3.4(a) genau dann durch die $\mathbf{TB}(\leq)$-relationale PMCF (über T) konstruiert, wenn mit $\preceq = \mathbf{TB}(\leq)$ gilt:

$$\begin{aligned}
T^-_A &= \bigcap \{M \in T \perp A: \forall M' \in T \perp A\ (M' \preceq M)\} && \text{(RPM)}\\
&= \bigcap \{M \in T \perp A: \forall M' \in T \perp A\ \forall A \in T \setminus M\ \exists B \in T \setminus M'\ (A \leq B)\} && \text{(Def. 3.3.4(a))}\\
&= \bigcap \{M \in T \perp A: \forall M' \in T \perp A\ \forall A \notin M\ \exists B \notin M'\ (A \leq B)\} && \text{(TW5)}\\
&= \bigcap \{M \in T \perp A: \forall M' \in T \perp A\ \forall \text{ TW-Schnitte } S\ (S \subseteq M' \Rightarrow S \subseteq M)\} && \text{(L. 3.5.5)}\\
&= \bigcap \{M \in T \perp A: S_A \subseteq M\} && \text{(L. 3.5.6)}\\
&= T \cap \{B: A < A \vee B\} && \text{(L. 3.5.7);}
\end{aligned}$$

d.h. genau dann, wenn $T^-{}_A$ durch die $\leq$-ordnungsgemäße Kontraktionsoperation nach Definition 3.4.1 konstruiert wird.
(Eränzung zu den Lemmata 3.3.5(a) und 3.3.7: Mit $\{M \in T\bot A: \forall M' \in T\bot A\ (M' \preceq M)\} = \{M \in T\bot A: S_A \subseteq M\}$ ist auch gezeigt, daß $\preceq = \mathbf{TB}(\leq)$ nach Definition 3.3.4(a) (TB0) erfüllt.) □

3.5.9. Lemma (a) Sei $\leq$ auf der Menge aller Sätze gegeben und $\preceq = \mathbf{TB}(\leq)$ auf der Klasse aller Satzmengen nach Definition 3.3.4(a) bestimmt. Erfülle $\leq$ die Bedingungen (TW1)–(TW3); wenn $\leq$ zusätzlich die Bedingung (TW5) erfüllt, so erfüllt $\preceq$ (TB6); wenn $\leq$ zusätzlich die Bedingung (TW6) erfüllt, so erfüllt $\preceq$ (TB5).
(b) Sei $\preceq$ auf der Klasse aller Satzmengen gegeben und $\leq = \mathbf{TW}(\preceq)$ auf der Menge aller Sätze nach Definition 3.3.4(b) bestimmt. Erfülle $\preceq$ die Bedingungen (TB1) und (TB2); wenn $\preceq$ zusätzlich die Bedingung (TB5) erfüllt, so erfüllt $\leq$ (TW6); wenn $\preceq$ zusätzlich die Bedingung (TB6) erfüllt, so erfüllt $\leq$ (TW5).

Beweis (a) Sei $\preceq = \mathbf{TB}(\leq)$ nach Definition 3.3.4(a), erfülle $\leq$ (TW1)–(TW3).
Erfülle $\leq$ zusätzlich (TW5). Sei $T \neq T_\bot$ und M eine beliebige Satzmenge. Betrachte nun die Bedingung $\forall M' \neq T_\bot (M' \preceq M)$; dies heißt nach Definition 3.3.4(a)

$$\forall M' \neq T_\bot\ \forall A \notin M\ \exists B \notin M'\ (A \leq B),$$

d.h. $\forall A \notin M\ \forall M' \neq T_\bot\ \exists B \notin M'\ (A \leq B)$, d.h. $\forall A \notin M\ \forall B\ (A \leq B)$, und dies ist nach (TW5) äquivalent mit $\forall A \notin M\ (A \notin T)$, d.h. $T \subseteq M$.
Erfülle $\leq$ zusätzlich (TW6). Sei M eine beliebige Satzmenge. Betrachte nun die Bedingung $\forall M' (M \preceq M')$; dies heißt nach Definition 3.3.4(a)

$$\forall M'\ \forall B \notin M'\ \exists A \notin M\ (B \leq A)\ ,$$

d.h. $\forall B\ \exists A \notin M\ (B \leq A)$, und dies ist wegen (TW3) und (TW1) äquivalent mit $\exists A \notin M\ (\top \leq A)$, was aus denselben Gründen mit $\exists A \notin M\ \forall B\ (B \leq A)$ äquivalent ist, und dies heißt nach (TW3) und (TW6) $\exists A \notin M\ (\vdash A)$.
(b) Sei $\leq = \mathbf{TW}(\preceq)$ nach Definition 3.3.4(b), erfülle $\preceq$ (TB1) und (TB2). Erfülle $\preceq$ zusätzlich (TB5). Sei A ein beliebiger Satz. Betrachte nun die Bedingung $\forall B\ (B \leq A)$; dies heißt nach Definition 3.3.4(b)

$$\forall B\ \forall M \not\ni A\ \exists M' \not\ni B\ (M \preceq M')\ ,$$

d.h. $\forall M \not\ni A\ \forall B\ \exists M' \not\ni B\ (M \preceq M')$, also insbesondere $\forall M \not\ni A\ \exists M' \not\ni \top\ (M \preceq M')$, d.h. wegen (TB5) und (TB1) $\forall M \not\ni A\ \forall M'\ (M \preceq M')$, d.h. wegen (TB5) $\forall M \not\ni A\ \exists B \notin M\ (\vdash B)$, d.h. $\vdash A$.
Erfülle $\preceq$ zusätzlich (TB6). Sei $T \neq T_\bot$ und A ein beliebiger Satz. Betrachte nun die Bedingung $\forall B\ (A \leq B)$; dies heißt nach Definition 3.3.4(b)

$\forall B\ \forall M \not\supseteq B\ \exists M' \not\supseteq A\ (M \preceq M')$,

d.h. $\forall M \neq T_\perp\ \exists M' \not\supseteq A\ (M \preceq M')$, und dies ist wegen (TB6) und (TB1) äquivalent mit $\exists M' \not\supseteq A\ (T \preceq M')$, was aus denselben Gründen mit $\exists M' \not\supseteq A\ \forall M \neq T_\perp\ (M \preceq M')$ äquivalent ist, und dies heißt nach (TB6) $\exists M' \not\supseteq A\ (T \subseteq M')$, d.h. $A \notin T$. □

3.5.10. Theorem Sei $T \neq T_\perp$ eine Theorie und $\preceq$ eine Ordnung auf der Klasse aller Satzmengen, die (TB0)–(TB2) und (TB6) erfüllt, und sei $\mathbf{TW}(\preceq)$ gemäß Definition 3.3.4(b) bestimmt. Dann gilt: Die $\mathbf{TW}(\preceq)$-ordnungsgemäße Kontraktionsoperation (über T) ist identisch mit der $\preceq$-relationalen PMCF.

Beweis: Für $\vdash A$ setzen sowohl relationale PMCFs als auch ordnungsgemäße Kontraktionsoperationen $T^-_A = T$. Gelte also von nun an $\not\vdash A$.
In diesem Fall wird T^-_A nach Definition 3.4.1 genau dann durch die $\mathbf{TW}(\preceq)$-ordnungsgemäße Kontraktionsoperation (über T) konstruiert, wenn mit $\leq = \mathbf{TW}(\preceq)$ und der Konvention, daß M und M' in diesem Beweis stets Elemente aus M(T) bezeichnen, gilt:

$B \in T^-_A \Leftrightarrow B \in T \cap \{B: A < A \vee B\}$ (Def. 3.4.1)

$\Leftrightarrow B \in T \cap \{B: \exists M' \not\supseteq A\ \forall M \not\supseteq A \vee B\ (M' \not\preceq M)\}$ (Def. 3.3.4(b))

Für $A \notin T$ kann wegen (TB6) mit $M' = T \neq T_\perp$ eine $\preceq$-maximale Satzmenge gefunden werden; dagegen gilt für jedes $B \in T$ auch $A \vee B \in T$, d.h. nach (TB6), daß jede Satzmenge, die $A \vee B$ nicht enthält, nicht $\preceq$-maximal ist. Also gilt $T \subseteq \{B: \exists M' \not\supseteq A\ \forall M \not\supseteq A \vee B\ (M' \not\preceq M)\}$, d.h. $T^-_A = T$. Da für $A \notin T$ $T \perp A = \{T\}$, gilt auch nach (RPM) und der Reflexivität von $\preceq$ $T^-_A = T$.
Für $A \in T$ hingegen kann man die Eigenschaft von maximalkonsistenten Teilmengen $M \in M(T)$ ausnutzen, daß für $A, B \in T$

$A \vee B \notin M \Leftrightarrow A \notin M \wedge B \notin M$

gilt. Denn sei ohne Einschränkung der Allgemeinheit $M \in T \perp C$. Es gilt $A \vee B \in T \setminus M$ genau dann, wenn $M \vdash (A \vee B) \rightarrow C$, d.h. wenn $M \vdash (A \rightarrow C) \wedge (B \rightarrow C)$, d.h. wenn $M \vdash A \rightarrow C$ und $M \vdash B \rightarrow C$, d.h. wenn $A \in T \setminus M$ und $B \in T \setminus M$.
Damit wissen wir, daß im Fall $A \in T$ gilt: $B \in T \cap \{B: \exists M' \not\supseteq A\ \forall M \not\supseteq A \vee B\ (M' \not\preceq M)\}$ genau dann, wenn

$B \in T \wedge \exists M' \not\supseteq A\ \forall M \in M(T)\ (M \not\supseteq A \wedge M \not\supseteq B \Rightarrow M' \not\preceq M)$. (*)

Daraus folgt quantorenlogisch

$\forall M \in M(T)\ (M \not\supseteq A \wedge M \not\supseteq B \Rightarrow \exists M' \not\supseteq A\ M' \not\preceq M)$. (**)

Um zu zeigen, daß umgekehrt auch (*) aus (**) folgt, nehmen wir an, daß zugleich (**) und die Negation von (*) gilt, d.h. daß

$\forall M \in M(T)\ (M \not\supseteq A \wedge M \not\supseteq B \Rightarrow \exists M' \not\supseteq A\ M' \not\preceq M) \wedge B \notin T$ (1. Fall)

oder

$$\forall M\in M(T)\ (M\not\supseteq A\wedge M\not\supseteq B \Rightarrow \exists M'\not\supseteq A\ M'\not\preceq M)\ \wedge$$
$$\forall M'\not\supseteq A\ \exists M\in M(T)\ (M\not\supseteq A\wedge M\not\supseteq B\ \wedge\ M'\preceq M)\quad (2.\ \text{Fall})$$

gilt. Im 1. Fall gilt wegen $B\notin T$ aber für jedes $M\not\supseteq A$ schon $M\not\supseteq B$, d.h. es folgt $\forall M\not\supseteq A\ \exists M'\not\supseteq A\ (M'\not\preceq M)$, und — da $A\in T$ — nach Lemma 2.4 von Alchourrón, Gärdenfors und Makinson (1985) $\forall M\in T\perp A\ \exists M'\in T\perp A\ (M'\not\preceq M)$, d.h. $\{M\in T\perp A\colon \forall M'\in T\perp A\ (M'\preceq M)\}=\emptyset$, obwohl $T\perp A\neq\emptyset$ (da $\not\vdash A$), im Widerspruch zu (TB0).

Im 2. Fall folgt wegen der Transitivität von $\preceq$ ebenfalls $\forall M\not\supseteq A\ \exists M'\not\supseteq A\ (M'\not\preceq M)$, was genauso widerlegt wird.

Damit ist bewiesen, daß unter den gegebenen Voraussetzungen (*) aus (**) folgt.

Die Bedingung (**) ist nun aber quantorenlogisch äquivalent mit $\forall M\not\supseteq A\ (\forall M'\not\supseteq A(M'\preceq M) \Rightarrow B\in M)$, oder, anders ausgedrückt, mit $B\in\bigcap\{M\not\supseteq A\colon \forall M'\not\supseteq A\ (M'\preceq M)\}$. Erneute Anwendung des Lemmas 2.4 von Alchourrón, Gärdenfors und Makinson (1985) ergibt, daß dies wiederum nichts anderes heißt als $B\in\bigcap\{M\in T\perp A\colon \forall M'\in T\perp A\ (M'\preceq M)\}$, d.h. $B\in T^{-}{}_{A}$, wenn $T^{-}{}_{A}$ durch die $\preceq$-relationale PMCF gemäß (RPM) konstruiert wird. □

3.5.11. Lemma Sei T eine Theorie. Eine schnittweise Kontraktionsoperation $\backslash$ (über T) erfüllt die Gärdenfors-Postulate (T-1)-(T-4) und (T-6)-(T-8), aber im allgemeinen nicht (T-5).

Beweis Für A und B mit $\vdash A$ (oder $\vdash B$) ist $T\backslash_A=T$ (bzw. $T\backslash_B=T$), womit man die (T-1)-(T-8) sofort verifiziert. Gelte also im folgenden $\not\vdash A$ und $\not\vdash B$.

(T-1): $T\backslash_A=S_A$ ist wegen (T-6) nicht leer, also nach Lemma 3.5.2(e) eine Theorie (Cn ist als kompakt vorausgesetzt).

(T-2): Sei $B\in T\backslash_A=S_A$, also $A<B$, d.h. $B\not\leq A$, also ist B nach (TW5) in T.

(T-3): Sei $A\notin T$, d.h. es gilt nach (TW5) $A\leq B$ für alle B. Wegen (T-2) muß für (T-3) nur $T\subseteq T\backslash_A$ gezeigt werden. Sei $C\in T$, also gibt es wegen (TW5) ein B mit $B<C$, also nach (TW1) $A<C$, also gilt $C\in S_A=T\backslash_A$.

(T-4): Wegen der Reflexivität von $\leq$ (folgt aus (TW2)) gilt $A\leq A$, also nicht $A<A$, d.h. $A\notin S_A=T\backslash_A$.

(T-6): Gelte $\vdash A\leftrightarrow B$, d.h. $A\vdash B$ und $B\vdash A$. Nach (TW3) gilt dann $A\leq B$ und $B\leq A$ und wegen (TW1) folgt daraus $T\backslash_A=S_A=S_B=T\backslash_B$.

(T-7): Sei $C\in T\backslash_A\cap T\backslash_B$, d.h. $A<C$ und $B<C$. Da nach (TW3) $A\wedge B\leq A$ (und $A\wedge B\leq B$), folgt nach (TW1) $A\wedge B<C$, d.h. $C\in S_{A\wedge B}=T\backslash_{A\wedge B}$.

(T-8): Sei $A\notin T\backslash_{A\wedge B}=S_{A\wedge B}$ und $C\in T\backslash_{A\wedge B}=S_{A\wedge B}$. Es gilt also nicht

$A \wedge B < A$, d.h. wegen (TW2) $A \leq A \wedge B$, und $A \wedge B < C$. Dann folgt mit (TW1) $A < C$, d.h. $C \in S_A = T^{\backslash}{}_A$.
Es gilt aber im allgemeinen *nicht* (TW5): Wir geben ein Gegenbeispiel. Seien A und B Sätze mit $A \not\vdash B$, $T := Cn(A \wedge B)$ und sei $\leq$ definiert durch

$$\forall C \in Cn(\emptyset)(C \doteq \top) \wedge \forall C, D \in T \backslash Cn(\emptyset)\ (\bot < C \doteq D < \top) \wedge \forall C \notin T\ (C \doteq \bot).$$

Man überlegt sich leicht, daß $\langle T, \leq \rangle$ (TW1)-(TW6) erfüllt. Es ist in diesem Fall $(T^{\backslash}{}_A)^+{}_A = (S_A)^+{}_A = (\{C: A < C\})^+{}_A = (\{C: \vdash C\})^+{}_A = Cn(A)$, jedoch gilt $T = Cn(A \wedge B) \not\subseteq Cn(A)$, da nach Voraussetzung $A \not\vdash B$. Das heißt aber, daß (T-5) hier verletzt wird. □

3.5.12. Lemma Sei eine Relation $\leq$ der theoretischen Wichtigkeit gegeben, die (TW1)-(TW6) erfüllt. Dann gilt für alle Sätze A

$$S_A \subseteq T \cap \{B: A < A \vee B\}.$$

Beweis Es ist zu zeigen, daß aus $A < B$ schon $B \in T \wedge A < A \vee B$ folgt.
Zu $B \in T$: Für $T = T_\bot$ ist $B \in T$ trivial. Für $T \neq T_\bot$ sagt (TW5), daß $B \in T$ genau dann, wenn ein A existiert mit $A < B$. Das ist aber nach Voraussetzung der Fall.
Zu $A < A \vee B$: Wegen (TW3) gilt $B \leq A \vee B$, also folgt aus $A < B$, d.h. $B \not\leq A$, mit (TW1) $A \vee B \not\leq A$, d.h. $A < A \vee B$. □

3.5.13. Theorem Sei T eine Theorie und $\leq$ eine Relation der theoretischen Wichtigkeit, die (TW1)-(TW6) erfüllt. Sei weiter $^-$ die $\leq$-ordnungsgemäße und sei $\backslash$ die $\leq$-schnittweise Kontraktionsoperation. Dann sind $^-$ und $\backslash$ revisionsäquivalent in dem Sinn, daß gilt: $\mathbf{L}(^-) = \mathbf{L}(\backslash)$.

Beweis Zu zeigen ist: $(S_A)^+{}_{\neg A} = (T \cap \{B: A < A \vee B\})^+{}_{\neg A}$. Für $\vdash A$ steht auf beiden Seiten die inkonsistente Theorie. Gelte von nun an $\not\vdash A$.
1. Fall: Sei $A \notin T$. Dann gilt sogar $S_A = T \cap \{B: A < A \vee B\}$. Denn einerseits gilt wegen (TW5) $S_A = T$, und andererseits gilt wegen (TW5) und (TW3) für alle $B \in T$ $A < B \leq A \vee B$, nach (TW1) also auch $A < A \vee B$.
2. Fall: Sei $A \in T$. Dann gilt — da S_A für $\not\vdash A$ eine Theorie ist —

$$\begin{aligned}
(S_A)^+{}_{\neg A} &= (\{B: A < B\})^+{}_{\neg A} = \\
&= \{C: \neg A \rightarrow C \in \{B: A < B\}\} = \\
&= \{C: A < \neg A \rightarrow C\} = \\
&= \{C: A < A \vee (\neg A \rightarrow C)\} = \\
&= \{C: \neg A \rightarrow C \in \{B: A < A \vee B\}\} = (\text{da } \{B: A < A \vee B\} \text{ Theorie!}) \\
&= (\{B: A < A \vee B\})^+{}_{\neg A} = \\
&= T_\bot \cap (\{B: A < A \vee B\})^+{}_{\neg A} = \\
&= T^+{}_{\neg A} \cap (\{B: A < A \vee B\})^+{}_{\neg A} =
\end{aligned}$$

$= (T \cap \{B: A<A\vee B\})^{+}{}_{\neg A}$.
(Man benötigt hier wesentlich die Tatsache, daß $\{B: A<A\vee B\}$ für $\not\vdash A$ eine Theorie ist. Dies zeigt man genauso, wie man beweist, daß TW-Schnitte Theorien sind.) □

3.6.2. Theorem Sei $\langle T,\leq\rangle$ eine Theorie und sei $\leq^*_A$ durch Definition 3.6.1 gegeben. Dann gilt:
(a) Erfüllt $\leq$ (TW1), so erfüllt auch $\leq^*_A$ (TW1); erfüllt $\leq$ (TW2), so erfüllt auch $\leq^*_A$ (TW2); erfüllt $\leq$ (TW3), so erfüllt auch $\leq^*_A$ (TW3); erfüllt $\leq$ (TW4), so erfüllt auch $\leq^*_A$ (TW4).
(b) Ist A ein Satz derart, daß $\not\vdash\neg A$, und erfüllt $\leq$ (TW1)-(TW4) und (TW6), so erfüllt $\langle T^*_A,\leq^*_A\rangle$ genau dann (TW5), wenn T^*_A die $\leq$-ordnungsgemäße Revision von T durch A ist.
(c) Ist A ein Satz derart, daß $\not\vdash A$, und erfüllt $\leq$ (TW3), so erfüllt $\leq^*_A$ nicht (TW6).

Beweis Sei $\langle T,\leq\rangle$ eine Theorie und $\leq^*_A$ durch Definition 3.6.1 gegeben.
(a) (TW1): Sei $B\leq^*_A C$ und $C\leq^*_A D$, d.h. $A\to B\leq A\to C$ und $A\to C\leq A\to D$, also mit (TW1) für $\leq$ $A\to B\leq A\to D$, d.h. $B\leq^*_A D$.
(TW2): Gelte $B\not\leq^*_A C$, d.h. $A\to B\not\leq A\to C$, also mit (TW2) für $\leq$ $A\to C\leq A\to B$, also $C\leq^*_A B$.
(TW3): Gelte $B\vdash C$, also auch $A\to B\vdash A\to C$, also nach (TW3) für $\leq$ $A\to B\leq A\to C$, also $B\leq^*_A C$.
(TW4): Nach (TW4) für $\leq$ gilt $A\to B\leq(A\to B)\wedge(A\to C)$ oder$A\to C\leq(A\to B)\wedge(A\to C)$, also gilt auch $A\to B\leq A\to(B\wedge C)$ oder $A\to C\leq A\to(B\wedge C)$, d.h. $B\leq^*_A B\wedge C$ oder $C\leq^*_A B\wedge C$.
(b) Erfülle $\leq$ (TW1)-(TW4) und (TW6), und sei A so, daß $\not\vdash\neg A$. Die Voraussetzung $\not\vdash\neg A$ sichert, daß $T^*_A\neq T_\perp$. Wenn dann $\langle T^*_A,\leq^*_A\rangle$ (TW5) erfüllt, so heißt dies

$T^*_A = \{B: \exists C\ (C<^*_A B)\}$
$= \{B: \exists C\ (A\to C<A\to B)\}$ Definition 3.6.1
$= \{B: \neg A<A\to B\}$ (TW3),(TW1)
$= \{B: A\to B\in\{D: \neg A<D\}\}$
$= (\{D: \neg A<D\})^{+}{}_A$ Definition 3.1.2, $\not\vdash\neg A$, $\{D: \neg A<D\}$ Theorie.

Die Voraussetzung $\not\vdash\neg A$ sichert mit (TW6), daß $\{D: \neg A<D\}$ nichtleer und damit nach Lemma 3.5.2(e) eine Theorie ist. Die Gleichung $T^*_A = (\{D: \neg A<D\})^{+}{}_A$ ist aber nach Definition 3.5.14 und Theorem 3.5.13 gerade die Bestimmung für eine ordnungsgemäße Revision T^*_A von T durch A.
(c) Sei A so, daß $\not\vdash A$. Wegen (TW3) für $\leq$ gilt $A\to B\leq A\to A$ für alle B, d.h.

$B \leq^*_A A$ für alle B, obwohl $\not\vdash A$, d.h. $\leq^*_A$ verletzt (TW6). □

Kapitel 4

Konditionalsätze und Theorienwandel: Revisionen, Expansionen und Additionen

4.1 Theorienrevisionsmodelle

Im letzten Kapitel haben wir von der Sprache, in der Theorien formuliert sind, weitgehend abstrahiert. Lediglich das Junktorenrepertoire der Aussagenlogik wurde vorausgesetzt. In diesem Kapitel wollen wir zusätzlich davon ausgehen, daß in der Objektsprache Konditionalsätze der Form Wenn A, dann B formulierbar sind. Konditionalsätze sollen gegenüber anderen Sätzen keinen von vornherein ausgezeichneten Status haben. Insbesondere seien Negationen von Konditionalsätzen erlaubt. Iterierte Konditionalsätze, Konditionalsätze also mit einem Konditionalsatz im Antzedens oder Konsequens, möchte ich ebenfalls nicht ausschließen, sie werden jedoch im folgenden nicht thematisiert werden.

Wie in Kapitel 3 deutlich geworden ist, sind Theorienrevisionsmodelle einfache formale Modelle für den Überzeugungs- oder Theorienwandel. Wir werden in diesem Kapitel vor allem Revisionen betrachten und Kontraktionen nur nebenbei erwähnen. Zwei Arten von Revisionen einer Theorie

T durch einen Satz A können unterschieden werden. Entweder ist A mit T konsistent (d.h. $T \not\vdash \neg A$); dann, so erwartet man, sollte die Angelegenheit eigentlich einfach sein. Oder aber A steht im Widerspruch mit T (d.h. $T \vdash \neg A$); dann wird die Aufgabe auf jeden Fall heikel — genau für diesen Fall wurden Theorienrevisionsmodelle entworfen. Das im letzten Kapitel vorgestellte Modell für Theorienwandel soll jedoch beide Fälle abdecken, wobei wichtig ist, daß Revisionen einer Theorie T immer ökonomisch oder „minimal" vonstatten gehen sollen, d.h. daß keine in T enthaltene Information unnötig preisgegeben wird.

Wir werden im folgenden mit einer vereinfachten Form der Gärdenforsschen Theorienrevisionsmodelle arbeiten. Unter einem *Theorienrevisionsmodell* oder *TRM* verstehen wir weiterhin ein geordnetes Paar $\langle \mathbf{T}, * \rangle$, wobei $\mathbf{T}$ eine Menge von Theorien ist und $*$: $\mathbf{T} \times \mathrm{Sent(L)} \rightarrow \mathbf{T}$ eine Funktion, die jeder Theorie $T \in \mathbf{T}$ und jedem Satz A die revidierte Theorie („Revision") T^*_A zuordnet. T^*_A soll wieder eine Theorie sein (d.h. $\langle \mathbf{T}, * \rangle$ soll (T*1) erfüllen) und interpretiert werden als die minimale Revision von T, die nötig ist, um A in T einzuverleiben. $\langle \mathbf{T}, * \rangle$ soll weiter intensional sein, d.h. (T*6) erfüllen. (T*1) und (T*6) werden i.f. normalerweise nicht mehr eigens erwähnt.

Bei den Überlegungen dieses Kapitels wollen wir von möglichst schwachen Voraussetzungen ausgehen. Anstelle der vollen Kollektion (T*1)–(T*8) werden wir — neben (T*1) und (T*6) — einen Minimalsatz von nur drei fundamentalen Kriterien für TRMe zugrunde legen. Er besteht aus dem *Erfolgskriterium* (T*2), dem *Konsistenzkriterium* (T*5) und einem *Stabilitätskriterium* (T*S):

(T*2) $A \in T^*_A$.
(T*5) $T^*_A = T_\perp \Rightarrow \vdash \neg A$.
(T*S) $A \in T \neq T_\perp \Rightarrow T^*_A = T$.

(T*S) folgt aus (T*3) und (T*4), ist jedoch sehr viel schwächer als die Konjunktion dieser beiden Bedingungen, da es nur Aussagen über den „trivialen" Fall $A \in T$ macht. Während (T*2) und (T*S) unproblematisch sein dürften, kann man über (T*5) vielleicht diskutieren.[1] Wenn der Satz A

[1] In vielen Arbeiten von Gärdenfors fehlt (T*5). So spielt etwa bei seiner Herausarbeitung des Lewisschen Systems **VC** in Gärdenfors (1979) (T*5) keine Rolle. Dies ist nicht verwunderlich, denn bei Lewis gibt es die sogenannte „vacuous truth" von Konditionalsätzen auch bei nichtkontradiktorischen Antezedenzien, was durch (T*5) zusammen mit dem Ramsey-Test (s.u.) ausgeschlossen wird. In Gärdenfors (1988) gelingt auch keine Anbindung von (T*5) an **VC**, denn das in Lemma 7.5 dort erwähnte Axiomenschema (A7) folgt, wie Gärdenfors (1979, S. 401) selbst zeigt, schon aus den übrigen

extrem weit hergeholt ist (wenn er z.B. *dem* fundamentalen Axiom von T widerspricht), dann könnte ein theoretischer Kollaps in T^*_A die Folge sein, und man wird die Landung in einem inkonsistenten T^*_A eventuell für möglich halten. Obgleich mir (T*5) nicht zu stark vorkommt, könnte man dieses Kriterium auch auf „intuitiv harmlose" oder „theoretisch denkbare" Sätze A beschränken. Die Argumentation unten würde an mehreren Stellen in entsprechender Weise eingeschränkt, aber nach meinem Dafürhalten nichts an Überzeugungskraft einbüßen. Deshalb werde ich gleich mit dem einfacheren und ebenfalls plausiblen (T*5) arbeiten.

Man beachte, daß (T*2), (T*5) und (T*S) Kriterien *für TRMe* sind. Die Quantifikation über alle Theorien T in **T** und über alle Sätze A habe ich nur der Übersichtlichkeit halber weggelassen, sie sollte jedoch stets im Kopf behalten werden.

Wir sind an TRMen interessiert, die den Wandel der Überzeugungen und Theorien von Menschen oder auch künstlichen Informationssystemen modellieren. Menschliche oder künstliche „Wissensbasen" sind aber nicht allwissend. Dies motiviert die folgenden Definitionen.

4.1.1. Definition Sei $\langle \mathbf{T}, * \rangle$ ein TRM und seien A und B (logisch voneinander unabhängige) beliebige Sätze der Objektsprache.
(a) Eine Theorie $T \in \mathbf{T}$ heiße *A-ignorant*, wenn weder A noch $\neg A$ in T ist;
(b) Eine Theorie $T \in \mathbf{T}$ heiße *A-B-ignorant*, wenn keine nichttautogigische wahrheitsfunktionale Kombination von A und B in T ist (d.h. wenn weder $A \vee B$ noch $A \vee \neg B$ noch $\neg A \vee B$ noch $\neg A \vee \neg B$ in T ist).
(c) $\langle \mathbf{T}, * \rangle$ heißt *schwach nichttrivial*, wenn gilt:
(SNT) Es gibt ein $T \in \mathbf{T}$ und einen Satz A derart, daß T A-ignorant ist.
(d) $\langle \mathbf{T}, * \rangle$ heißt *nichttrivial*, wenn gilt:
(NT) Es gibt eine Theorie $T \in \mathbf{T}$ und Sätze A und B derart, daß T A-B-ignorant ist.
Es ist klar, daß die A-B-Ignoranz einer Theorie ihre A-Ignoranz und daß (NT) (SNT) impliziert.

Ich gehe davon aus, daß jedes TRM intuitiv (T*2), (T*5), (T*S) und (NT) erfüllen sollte. Ich werde unten aber immer explizit angeben, ob und — wenn ja — wo ich von diesen Kriterien tatsächlich Gebrauch mache.

Axiomenschemata (vgl. Definition 4.2.1 unten).

4.2 Konditionalsätze und Gärdenfors' Trivialisierungstheorem

Wohl die wichtigste Anwendung, die TRMe bisher gefunden haben, ist die Analyse von Konditionalsätzen. Die Pionierarbeit bei dieser Anwendung wurde von Peter Gärdenfors (1979) geleistet. Der Schlüssel zur Interpretation von Konditionalsätzen besteht im sogenannten *Ramsey-Test*. Er ist hervorgegangen aus einer Fußnote von Frank P. Ramsey (1931) und läßt sich sehr natürlich formulieren als ein *Kriterium für TRMe* in Sprachen, die ein Konditionalkonnektiv wenn ... dann enthalten:

(R) Wenn A, dann B $\in$ T gdw. B $\in T^*_A$.

Gärdenfors (1979; 1988, Kapitel 7) hat die Analyse von Konditionalen mittels des Ramsey-Test und seines TRM-Konzepts sehr weit getrieben. Er zeigt, daß eine Abschwächung von (T*1)-(T*8)[2] zusammen mit (R) als Semantik für Konditionalsätze dienen kann, bezüglich der das inzwischen zu einer Art Standard gewordene System **VC** von Lewis (1973a) korrekt und vollständig ist. Ein Satz A heißt hierbei *gültig*, wenn er in allen TRMen $\langle \mathbf{T}, * \rangle$ gültig ist, was heißen soll, daß $\neg$A in keinem $T \neq T_\perp$ irgendeines BRMs $\langle \mathbf{T}, * \rangle$ ist.

Gärdenfors (1988) gibt folgende Axiomatisierung von **VC**:

4.2.1. Definition Das System **VC** besteht aus den Schlußregeln

(R1) Modus ponens

(R2) $\vdash$(B$\rightarrow$C) $\Rightarrow \vdash$(Wenn A, dann B)$\rightarrow$(Wenn A, dann C)

und den Axiomenschemata

(A1) Wahrheitsfunktionale Tautologien

(A2) ((Wenn A, dann B)$\wedge$(Wenn A, dann C)) $\rightarrow$ (Wenn A, dann B$\wedge$C)

(A3) (Wenn A, dann $\top$)

(A4) (Wenn A, dann A)

(A5) (Wenn A, dann B) $\rightarrow$ (A$\rightarrow$B)

(A6) (A$\wedge$B) $\rightarrow$ (Wenn A, dann B)

(A7) (Wenn A, dann $\neg$A) $\rightarrow$ (Wenn B, dann $\neg$A)

(A8) ((Wenn A, dann B)$\wedge$(Wenn B, dann A)) $\rightarrow$
((Wenn A, dann C)$\rightarrow$(Wenn B, dann C))

(A9) ((Wenn A, dann C)$\wedge$(Wenn B, dann C)) $\rightarrow$ (Wenn A$\vee$B, dann C)

[2] Die Modifikation der Gärdenfors-Postulate bezieht sich auf (K*4), (K*5) und (K*8). (K*5) kann ganz weggelassen werden (vgl. Fußnote 1). (T*4) wird abgeschwächt zu einer Hälfte von (T*S), nämlich zu $A \in T \neq T_\perp \Rightarrow T \subseteq T^*_A$. Zur (besonders interessanten) Abschwächung von (T*8) vgl. Fußnote 15.

(A10) ((Wenn A, dann C)∧¬(Wenn A, dann ¬B)) →
(Wenn A∧B, dann C).

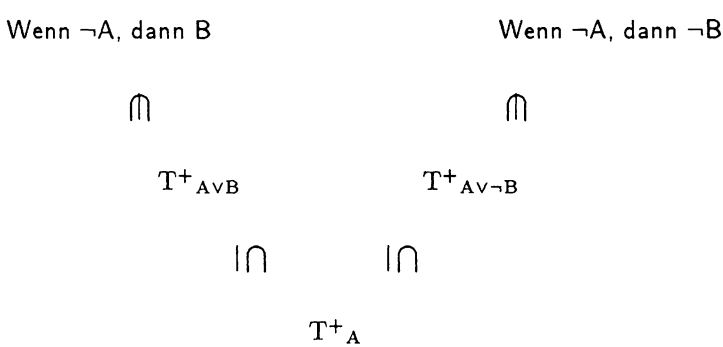
Neuerdings hat Gärdenfors aber seinem eigenen Ansatz einen schweren Schlag versetzt. In Gärdenfors (1986) formuliert er folgendes inzwischen berüchtigte *Trivialisierungstheorem*: Es gibt kein nichttriviales TRM, welches (T*2), (T*5), (R) und (T*P) erfüllt.[3]

(T*P) ist hierbei ein prima facie sehr plausibles *Erhaltungskriterium*: Wenn der Satz A, bzgl. dessen die Revision von T durchgeführt wird, mit T konsistent ist, dann wird in T^*_A nichts von T aufgegeben. Formal heißt dies:

(T*P) Wenn $\neg A \notin T$, dann $T \subseteq T^*_A$.

(T*P) folgt aus (T*4), ist aber viel schwächer als dieses. Ich will nun einen sehr einfachen Beweis des Theorems angeben, der überschaubarer ist als der in Gärdenfors (1986). Die Übersichtlichkeit soll helfen, die Gründe für die überraschende Trivialität zu erkennen. Alle Ergebnisse, die aufgrund des folgenden Beweises gewonnen werden, lassen sich auch auf Peter Gärdenfors' Beweis übertragen, wie leicht nachgeprüft werden kann.[4]

Sei nun $\langle \mathbf{T}, * \rangle$ nichttrivial und $T \in \mathbf{T}$ A-B-ignorant. Die Kernidee des Beweises veranschaulicht Skizze 4.1:



SKIZZE 4.1

[3] Gärdenfors hat einen anderen, nur unwesentlich verschiedenen Begriff von nichttrivialen TRMen.

[4] Mein Beweis erfordert keine iterierten Konditionalsätze. Dies kann man, wie David Makinson (Brief v. 16. 12. 1988 und Makinson 1989b), sogar als den größten Vorzug des Beweises ansehen.

T^+_A ($T^+_{A \vee B}$, $T^+_{A \vee \neg B}$) steht hier wieder für die Expansion von T bezüglich A (bzw. $A \vee B$ bzw. $A \vee \neg B$) gemäß Definition 3.1.2.

Wir benutzen jetzt die in der Behauptung von Gärdenfors' Trivialisierungstheorem erwähnten Voraussetzungen und schreiben den Beweis einmal Schritt für Schritt auf:

(1)	$A \vee B \in T^+_{A \vee B}$	Definition 3.1.2
(2)	$A \notin T^+_{A \vee B}$	A-B-Ignoranz von T, Definition 3.1.2
(3)	$A \vee B \in (T^+_{A \vee B})^*_{\neg A}$	(1), (2), (T*P)
(4)	$\neg A \in (T^+_{A \vee B})^*_{\neg A}$	(T*2)
(5)	$B \in (T^+_{A \vee B})^*_{\neg A}$	(3), (4), $(T^+_{A \vee B})^*_{\neg A}$ ist Theorie
(6)	Wenn $\neg A$, dann $B \in T^+_{A \vee B}$	(5), (R)
(7)	$T^+_{A \vee B} \subseteq T^+_A$	Definition 3.1.2
(8)	Wenn $\neg A$, dann $B \in T^+_A$	(6), (7)
(9)	Wenn $\neg A$, dann $\neg B \in T^+_A$	Ganz analog zu (1)–(8), mit $\neg B$ anstelle von B

Die Zwischenergebnisse (8) und (9), welche sagen, daß sowohl Wenn $\neg A$, dann B als auch Wenn $\neg A$, dann $\neg B$ in T^+_A ist, kollidieren mit (R) und (T*5):

(10)	$B, \neg B \in (T^+_A)^*_{\neg A}$	(8), (9), (R)
(11)	$(T^+_A)^*_{\neg A}$ inkonsistent	(10)
(12)	$\neg A$ konsistent	A-B-Ignoranz von T
(13)	$(T^+_A)^*_{\neg A}$ konsistent	(12), (T*5)
(14)	Widerspruch	(11), (13) .

Damit scheinen wir — unter Verwendung genau der von Gärdenfors angeführten Kriterien (T*2), (T*5), (T*P) und (R) — gezeigt zu haben, daß nur triviale TRMe diesen vier Kriterien genügen können.

Ist uns das aber tatsächlich gelungen? Haben wir keine Voraussetzungen unterschlagen? Doch. Wir haben vergessen, daß in den eben verwendeten Bedingungen über alle T *in* **T** quantifiziert wird. Wir haben diese Bedingungen ganz ungeniert *auf Expansionen* angewandt. Also haben wir stillschweigend vorausgesetzt, daß mit einem T im **T** eines TRMs auch die Expansionen dieses T in **T** sind. Das Gärdenforssche Theorem muß also genaugenommen so lauten:

4.2.2. Theorem (Gärdenfors) Es gibt kein nichttriviales TRM $\langle \mathbf{T}, * \rangle$, welches (T*2), (T*5), (R) und (T*P) erfüllt und in dem **T** gegenüber Expansionsbildung abgeschlossen ist.

Gärdenfors (1986, S. 85; 1988, S. 148) selbst erwähnt die Abgeschlossenheit von TRMen gegenüber Expansionen nur ganz beiläufig als technische

Voraussetzung. Dies birgt aber die Gefahr einer Petitio principii zugunsten von (T*P), denn die plausibelste Rechtfertigung der Abgeschlossenheit gegenüber Expansionen impliziert, wie wir gleich sehen werden, (T*P). Außerdem erscheinen die in Theorem 4.2.2 genannten ersten vier Bedingungen — mit der Ausnahme von (R) — durchaus nicht problematischer als die letztgenannte zu sein. Wir haben es also nicht mehr — wie Gärdenfors meint — mit einem Dilemma mit den zwei „Hörnern" (R) und (T*P) zu tun, sondern es ist uns eine zusätzliche Aufgabe aufgebürdet. Warum sollen denn Expansionen eigentlich in TRMen enthalten sein? Die naheliegende, und vielleicht auch die einzige Antwort auf diese Frage lautet: Weil Revisionen in bestimmten Fällen gerade Expansionen sind. So schreibt Peter Gärdenfors (1988, S. 54):

> The normal application area of a revision process is when the input A contradicts what is already in T, i.e. $\neg A \in T$. However, in order to have the revision function defined for all arguments, we can easily extend it to cover the case when $\neg A \notin T$. In this case, the revision is, of course, identified with an expansion.

Der Vorschlag wäre also:

(T*E) Wenn $\neg A \notin T$, dann $T^*_A = T^+_A$.

Dies ist äquivalent mit der Konjunktion der Konjunktion der Gärdenfors-Postulate (T*3) und (T*4). Durch (E) rechtfertigt sich die Abgeschlossenheit von TRMen gegenüber Expansionen natürlich schon über die Idee und Definition von TRMen. Außerdem erübrigt sich die eigene Nennung von (T*P), welches schon aus (T*E) folgt. Das Theorem lautet nun anders formuliert:

4.2.3. Korollar Es gibt kein nichttriviales TRM, welches (T*2), (T*5), (R) und (T*E) erfüllt.

Da wir (T*1) und (T*6) stets präsupponieren, heißt das einfach, daß es *kein nichttriviales BRM* gibt, *welches die grundlegenden Gärdenfors-Postulate und (R) erfüllt.* Gärdenfors würde angesichts dieses neuformulierten Dilemmas vermutlich wieder gegen (R) und für (T*E) votieren. Ich dagegen werde Argumente vorzubringen versuchen, daß man (R) beibehalten und (T*E) aufgeben sollte.

Bevor wir uns einer systematischeren Untersuchung des Verhältnisses von konsistenten Revisionen T^*_A, wo $\neg A \notin T$, und Expansionen T^+_A zuwenden, soll noch kurz erwähnt werden, daß *eine* Art Ausweg versperrt ist. Der Ramsey-Test (R) verlangt keine wirkliche Relevanz von A für B.

Insbesondere ist es eine Folge von (T*S), daß der Konditionalsatz Wenn A, dann B schon dann akzeptiert werden darf und muß, wenn man nur A und B für wahr hält. Als Kandidaten zur Vermeidung dieses intuitiven Mangels kommen mindestens folgende vier Abwandlungen von (R) infrage:

(R1) Wenn A, dann $B \in T$ gdw. $B \in T^*_A \wedge B \notin T$.
(R2) Wenn A, dann $B \in T$ gdw. $B \in T^*_A \wedge B \notin T^*_{\neg A}$.
(R3) Wenn A, dann $B \in T$ gdw. $B \in (T^-_B)^*_A$.
(R4) Wenn A, dann $B \in T$ gdw. $B \in T^*_A \wedge \neg A \in T^*_{\neg B}$.

(R1)-(R3) wurden in meiner Magisterarbeit (Rott 1984, S. 118f) angesprochen, wo (R2) der *Starke Ramsey-Test* genannt wurde. (R4) ist bei McCall (1983; vgl. Abschnitt 6.8) angelegt. In (R3) bezeichnet T^-_B die Kontraktion von T bzgl. B. Von den Gärdenforsschen Bedingungen (T-1)-(T-8) für Kontraktionen brauchen wir jetzt nur das *Stabilitätskriterium* (T-3):

(T-3) Wenn $B \notin T$, dann $T^-_B = T$.

Für (R1)-(R3) hat Gärdenfors (1987) gezeigt, daß diese Varianten des Ramsey-Tests nicht aus dem Dilemma herausführen. Wenn man (R1)-(R4) auf Skizze 4.1 anwendet, sieht man, daß erstens Wenn $\neg A$, dann $(\neg)B$ immer noch in $T^+_{A \vee (\neg)B}$ bleibt: Für (R1) ist zusätzlich zu zeigen, daß $(\neg)B \notin T^+_{A \vee (\neg)B}$, was nach Voraussetzung gilt; für (R2) ist zusätzlich zu zeigen, daß $(\neg)B \notin (T^+_{A \vee (\neg)B})^*_A$, wozu man wieder (T*E) benötigt; für (R3) ändert sich wegen (T-3) nichts; für (R4) ist zusätzlich zu zeigen, daß $A \in (K^+_{A \vee (\neg)B})^*_{(\neg)\neg B}$, was ganz analog wie $(\neg)B \in (K^+_{A \vee (\neg)B})^*_{\neg A}$ nachgewiesen wird. Der Widerspruch aus der Tatsache, daß sowohl Wenn $\neg A$, dann B als auch Wenn $\neg A$, dann $\neg B$ in T^+_A ist, folgt zweitens wie bei (R) (für (R3) braucht man noch einmal (T-3)). Damit haben wir das folgende

4.2.4. Korollar Es gibt kein nichttriviales TRM, welches
(a) (T*2), (T*5), (R1) und (T*E) *oder*
(b) (T*2), (T*5), (R2) und (T*E) *oder*
(c) (T*2), (T*5), (R3), (T*E) und (T-3) *oder*
(d) (T*2), (T*5), (R4) und (T*E) erfüllt.

Der Einbau einer Relevanzbedingung mittels der vier angeführten Varianten des Ramsey-Tests macht also in diesem Zusammenhang keinen wesentlichen Unterschied gegenüber (R) aus. Deshalb werden wir in den folgenden Abschnitten beim alten Ramsey-Test bleiben.

4.3 Expansionen und Additionen

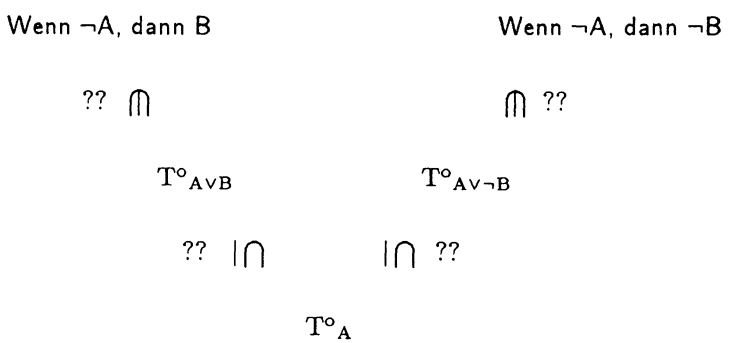
Es war unangenehm, daß wir im letzten Abschnitt die Abgeschlossenheit von TRMen gegenüber Expansionen hinnehmen mußten. In Anbetracht des Trivialisierungstheorems ist es nicht mehr klar, daß Expansionen spezielle Revisionen sind. Wir können sehen, ob und — wenn ja — wo zusätzliche Bedingungen, die in (T*E) enthalten sind, in den Beweis von Gärdenfors' Theorem einfließen, wenn wir statt mit Expansionen von vornherein mit *konsistenten Revisionen* oder kürzer *Additionen* T^*_A arbeiten, bei denen $\neg A \notin T$. Führen wir eine eigene, übersichtliche Notation ein: im folgenden stehe T°_A immer für die Revision von T bezüglich A in dem Fall, wo die Voraussetzung $\neg A \notin T$ gesichert und stillschweigend mitzudenken ist. (T*P) und (T*E) etwa lauten dann einfach so:

(T*P) $T \subseteq T^\circ_A$.
(T*E) $T^\circ_A = T^+_A$.

Für spätere Zwecke ist es hilfreich, (T*E) in zwei Hälften aufzuspalten:

(T°3) $T^\circ_A \subseteq T^+_A$.
(T*4) $T^+_A \subseteq T^\circ_A$.

(T°3), die Einschränkung von (T*3) auf konsistente Revisionen, ist keine echte Abschwächung von (T*3), da im Fall $\neg A \in T$ $T^+_A = T_\perp$ gilt. Deshalb nehmen wir im folgenden gleich (T*3) anstelle (T°3) her. Jetzt aber zum Beweis eines zu Theorem 4.2.2 analogen Satzes. Die neue Beweisidee ist in Skizze 4.2 dargestellt:



SKIZZE 4.2

Wenn wir in den obigen Beweisschritten (1)–(14) systematisch „$^+$“ durch „$^\circ$“ ersetzen wollen, sehen wir, daß an genau zwei Stellen, an denen wir uns auf Definition 3.1.2 bezogen, zusätzliche Voraussetzungen zum Tragen kommen müssen. In Schritt (1) kann (T*2) benutzt werden. Aber für Schritt (2) brauchen wir (T*3), und für Schritt (7) brauchen wir eine Bedingung, die ich *Monotonie von Additionen* nennen möchte:

(T°M) $\quad A \vdash B \Rightarrow T^\circ_B \subseteq T^\circ_A$.

Wir verwendeten im Beweis eine andere, äquivalente Formulierung dieser Bedingung, nämlich die Beziehung $T^\circ_{A\vee B}\subseteq T^\circ_A$. Wem im Gegensatz zu Gärdenfors die Bedingung (T*P) verdächtig ist, der will vielleicht die Schritte (3)–(5) durch eine neue Schlußkette ersetzen, die von einer Bedingung Gebrauch macht, die man *Konjunktivität von Additionen* nennen kann:

(T°∧) $\quad (T^\circ_A)^\circ_B = T^\circ_{A\wedge B}$.

Man beachte, daß für (T°∧) dreimal nachgeprüft werden muß, ob „$^\circ$“ statt „*“ stehen darf.[5] Anstelle von (3)–(5) oben (mit „$^\circ$“ statt „$^+$“) hätte man dann

(3′)	$(T^\circ_{A\vee B})^\circ_{\neg A} = T^\circ_{\neg A\wedge B}$	A-B-Ignoranz von T, (2), (T°∧)
(4′)	$B \in T^\circ_{\neg A\wedge B}$	(T*2)
(5′)	$B \in (T^\circ_{A\vee B})^\circ_{\neg A}$	(3′), (4′)

Eine andere Variationsmöglichkeit: (T°∧) würde es erlauben, auf (T°M) zu verzichten, wenn man nun doch wieder auf (T*P) zurückgreift. Anstelle von (7) oben würde dann die folgende Kette stehen:

(7′)	$\neg A \notin T^\circ_{A\vee B}$	A-B-Ignoranz von T, Def.3.1.2, (T*3)
(7″)	$(T^\circ_{A\vee B})^\circ_A = T^\circ_A$	(7′), (T°∧)
(7‴)	$T^\circ_{A\vee B} \subseteq T^\circ_A$	(7″), (T*P)

(T°M) und (T°∧) entsprechen Bedingungen (T^+M) und ($T^+\wedge$) für Expansionen, die aus Definition 3.1.2 folgen. Gärdenfors erwähnt (T^+M) und ($T^+\wedge$) explizit in seinen Aufsätzen über die Trivialitätsresultate (1986, 1987). Was ist aber nun gewonnen oder verloren durch die verschiedenen Beweisvarianten? Um Klarheit über die gegenseitigen Abhängigkeiten

[5] Eine Konjunktivitätsbedingung

(T*∧) $\quad (T^*_A)^*_B = T^*_{A\wedge B}$

für Revisionen im allgemeinen ist eindeutig fehl am Platze, wenn man (T*2), (T*5) und (T*S) nicht aufgeben will. Denn sei A konsistent und $\neg A\in T$. Mit (T*∧) und (T*S) erhalten wir dann die folgende Kette:

$$T^*_{A\wedge B} = (T^*_{\neg A\vee B})^*_A = T^*_A = (T^*_{\neg A\vee\neg B})^*_A = T^*_{A\wedge\neg B}\ .$$

Wegen (T*2) gilt $B\in T^*_{A\wedge B}$ und $\neg B\in T^*_{A\wedge\neg B}$, und da diese beiden Theorien als identisch mit T^*_A erwiesen wurden, muß T^*_A inkonsistent sein, im Widerspruch zu (T*5).

der verschiedenen Bedingungen zu erhalten, schreiben wir alles in einem kleinen Lemma zusammmen:

4.3.1. Lemma (a) $(T^*E) \Leftrightarrow (T^*3) \wedge (T^*4)$.
(b) $(T^*E) \Rightarrow (T^\circ M)$, $(T^*E) \Rightarrow (T^\circ\wedge)$.
(c) $(T^*4) \Leftrightarrow (T^*P) \wedge (T^\circ 2)$.
(d) $(T^\circ M) \wedge (T^*S) \Rightarrow (T^*P)$.
(e) $(T^\circ\wedge) \wedge (T^*2) \wedge (T^*S) \Rightarrow (T^*P)$.
(f) $(T^\circ\wedge) \wedge (T^*2) \wedge (T^*3) \wedge (T^*S) \Rightarrow (T^\circ M)$.

In (c) soll $(T^\circ 2)$ die Einschränkung von (T^*2) auf konsistente Revisionen bezeichnen:

$(T^\circ 2)$ $\quad A \in T^\circ{}_A$.

Beweis von Lemma 4.3.1: (a) Offensichtlich.
(b) Da Cn eine konservative Erweiterung der Aussagenlogik und jedes T bzgl. Cn abgeschlossen ist, gilt für alle T: mit $B\rightarrow C$ ist für ein A mit $A\vdash B$ auch $A\rightarrow C$ in T, und $A\rightarrow(B\rightarrow C)$ ist in T genau dann, wenn $(A\wedge B)\rightarrow C$ in T ist. Also folgt $(T^\circ M)$ und $(T^\circ\wedge)$ aus (T^*E) und Definition 3.1.2.
(c) Von links nach rechts: Für alle T ist mit B auch $A\rightarrow B$ in T, und $A\rightarrow A$ ist ebenfalls in T. Also gilt $T\cup\{A\} \subseteq T^+{}_A$, also mit (T^*4) auch $T\cup\{A\} \subseteq T^\circ{}_A$, also gilt (T^*P) und $(T^\circ 2)$. Von rechts nach links: Mit (T^*P) und $(T^\circ 2)$ gilt $T\cup\{A\} \subseteq T^\circ{}_A$. Da $T^+{}_A$ die kleinste Cn-abgeschlossene Menge ist, die $T\cup\{A\}$ enthält, und da $T^\circ{}_A$ als Theorie Cn-abgeschlossen sein muß, gilt $T^+{}_A \subseteq T^\circ{}_A$, d.h. (T^*4).
(d) Setze $\neg A$ oder T für B in $(T^\circ M)$.
(e) Sei $\neg A\notin T$ und $B\in T$. Dann ist mit (T^*S) und $(T^\circ\wedge)$ $T^\circ{}_A = (T^\circ{}_B)^\circ{}_A = T^\circ{}_{A\wedge B}$. (Auch im letzten Fall liegt tatsächlich eine Addition vor, da n.V. $\neg(A\wedge B)\notin T$.) Nach (T^*2) ist $B\in T^\circ{}_{A\wedge B}$, also auch $B\in T^\circ{}_A$, womit (T^*P) bewiesen ist.
(f) Nach (e) dürfen wir auch (T^*P) verwenden. Gelte $A\vdash B$ und sei $\neg B\notin T$, also auch $\neg A\notin T$. Wegen (T^*P) und $(T^\circ\wedge)$ ist $T^\circ{}_B \subseteq (T^\circ{}_B)^\circ{}_A = T^\circ{}_{A\wedge B} = T^\circ{}_A$. Damit sind wir fertig, wenn wir gezeigt haben, daß wir es beim mittleren Ausdruck mit einer zweifachen Addition zu tun haben. Hierzu muß man noch nachprüfen, daß $\neg A\notin T^\circ{}_B$: Wegen $\neg A\notin T$ und $\neg B\vdash\neg A$ ist $\neg B\vee\neg A\notin T$, d.h. es gilt $\neg A\notin T^+{}_B$, also nach (T^*3) auch $\neg A\notin T^\circ{}_B$. □

Lemma 4.3.1 führt uns vor Augen, daß wir kein einziges der grundlegenden Gärdenfors-Postulate einsparen können. (T^*1), (T^*2) und (T^*S) kann man als unproblematisch ansehen. Für den Beweisschritt (2) mit Additionen kommen wir nicht darum herum, (T^*3) zu fordern. Die Teile (c)–(e) des Lemmas zeigen schließlich, daß wir, unter der Voraussetzung von (T^*2)

und (T*S), auch (T*4) präsupponieren, egal ob wir mit (T°P), (T°M) oder (T°∧) arbeiten.

Wir müssen also — soweit ich sehe — die ganze Kraft von (T*E) in Anspruch nehmen. Es ist klar, daß man für den Trivialitätsbeweis als Argument gegen (R) mehr braucht als nur (T*P). (T*E), welches ohnehin die natürlichste und vielleicht sogar einzige Motivation der anderen Bedingungen lieferte, muß auch in seiner zweiten Hälfte, d.h. in (T*3), akzeptiert werden. Damit ist aber neben den Gärdenforsschen Kandidaten (R) und (T*P) ein weiterer Kandidat für Streichungen aufgetaucht, mämlich (T*3). Andererseits wissen wir aus den Teilen (a) und (b) des Lemmas, daß wir unter Voraussetzung von (T*3) für (T°M) bzw. (T°∧), die wir im Beweis brauchen, keine über (T*4) hinausgehende Rechtfertigung benötigen. Deshalb sind die Voraussetzungen für das Ergebnis, welches besagt, daß es keine nichttrivialen TRMe gibt, die die grundlegenden Gärdenfors-Postulate und (R) erfüllen, weder lückenhaft noch redundant. Wir hatten schon im letzten Abschnitt die aussagekräftigste Form des Trivialisierungstheorems gefunden.

Bei intuitiver Betrachtung der Skizze 4.2 drängt sich unmittelbar der Eindruck auf, daß für ein A-B-ignorantes T Wenn ¬A, dann (¬)B zwar in $T^{\circ}_{A \vee (\neg)B}$, nicht aber in T°_{A} sein sollte. Damit wäre der schwarze Peter bei (T°M) — bzw. via Lemma 4.3.1 bei (T*4) —, während (R) und (T*3) zunächst einmal entlastet sind. Sicher ist jedenfalls, daß (T*E) nicht mehr paßt, d.h. daß Expansionen nicht die richtige Methode sind, neue Sätze zu „addieren", wenn die Objektsprache Konditionalsätze enthält und diese nach dem Ramsey-Test interpretiert werden. Ich glaube jedoch, daß man beide Hälften, *sowohl* (T*4) *als auch* (T*3), opfern muß. Im nächsten Abschnitt werden Argumente gegen diese Bedingungen präsentiert.

4.4 Autoepistemische Allwissenheit

Der Ramsey-Test weist uns an, nachzusehen, was in T^*_A los ist: Wenn B in T^*_A ist, dann — und nur dann — ist Wenn A, dann B in T. Was aber, wenn B nicht in T^*_A ist? Über dieses negative Ergebnis sollten wir in T genauso Buch führen dürfen wie über das positive. Wenn B nicht in T^*_A ist, dann so scheint es, dürfen und müssen wir Es ist nicht der Fall, daß wenn A, dann B oder kurz ¬(Wenn A, dann B) in T akzeptieren. Wenn man einmal die Idealisierungen, die in deduktiv abgeschlossenen Theorien und im Ramsey-Test stecken, geschluckt hat, dann sollte man auch im gleichen Sinne die

Akzeptabilitätsbedingung für negierte Konditionalsätze annehmen. Theoretiker mit idealen deduktiven Fähigkeiten sind allwissend bezüglich ihres eigenen Theorienwandels. Wir halten folgende Akzeptabilitätsbedingung für negierte Konditionalsätze fest:

(R¬) ¬(Wenn A, dann B) ∈ T gdw. B $\notin T^*_A$.

Im folgenden bezeichne ich mit *autoepistemischer Allwissenheit* (*AEA*)[6] das Prinzip, daß in jedem T∈**T** für alle Satzpaare A und B entweder Wenn A, dann B oder ¬(Wenn A, dann B) ist.[7] Man beachte, daß wir es wieder mit einem Prinzip *für TRMe* zu tun haben. Ich halte (AEA) für genauso plausibel wie (R), möchte es also in der Plausibilitätsskala der Kriterien für TRMe zwar hinter (T*2), (T*5) und (T*S), aber noch vor (T*P), (T*E), (T°M), (T*3) etc. einordnen.

Zusammmen mit (T*S) impliziert die Allwissenheit von ideal deduzierenden Theoretikern bezüglich ihres eigenen Theorienwandels natürlich auch ihre Allwissenheit bezüglich ihrer Theorien. Die statische Spur der stärkeren dynamischen Forderung erhält man, wenn man die Konditionalsätze Wenn ⊤, dann A und ¬(Wenn ⊤, dann A) betrachtet. (T*S) sagt uns, daß sich diese Sätze, die wir ab jetzt mit □A und ¬□A abkürzen wollen, nicht auf echte Theorienrevisionen, sondern auf die je gegenwärtig vorliegenden Theorien beziehen. Mit (AEA) haben sie folgende Akzeptabilitätskriterien:

(R□) □A ∈ T gdw. A ∈ T .

(R¬□) ¬□A ∈ T gdw. A $\notin$ T .

(R□) und (R¬□) sind wieder Kriterien, die für alle T in einem TRM ⟨**T**,*⟩ gelten sollen. Da sie — wenn man (T*S) hat — schwächer sind als die entsprechenden Bedingungen für (negierte) Konditionalsätze, dürfte klar sein, daß (R□) und (R¬□) nicht nur über den Ramsey-Test und (AEA)

[6] Ich sage „autoepistemische" Allwissenheit, weil Theorien als die Menge der Überzeugungen eines (idealen) epistemischen Subjekts gedeutet werden können und weil ich den Bezug zur im nächsten Kapitel besprochenen „autoepistemischen Logik" herausstellen möchte.

[7] Man darf das Prinzip der autoepistemischen Allwissenheit nicht mit Stalnakers (1981) ‚Conditional excluded middle', d.h. mit dem konditionallogischen Axiomenschema (Wenn A, dann B) v (Wenn A, dann ¬B) verwechseln. Es scheint jedoch ein pragmatisches Faktum zu sein, daß man oft Wenn A, dann ¬B meint, wenn man ¬(Wenn A, dann B) äußert. Die m.E. richtige Lesart von ¬(Wenn A, dann B) erhält man am besten dialogisch. Anton behauptet: Wenn Gorbatschow kurz nach Amtsantritt gestorben wäre, dann stünden wir jetzt vor dem dritten Weltkrieg. Darauf protestiert Berta: Nein (das ist nicht ausgemacht/nicht unbedingt/das kann man nicht so sagen).

als interessant zu entdecken und rechtfertigen sind.[8] Insofern stellt das Folgende ein von (R) (zumindest zum Teil) unabhängiges Argument gegen die „Konkurrenten" von (R) dar.

Man beachte übrigens, daß etwa die inkonsistente Menge durch (R¬□) (wie auch schon durch (AEA)) aus **T** ausgeschlossen wird, da sie alle Sätze enthält. Um eine Kollision von (R¬□) mit (T*2) zu vermeiden, kann man zum Beispiel das Erfolgskriterium (T*2) ähnlich einschränken, wie wir es vom Erfolgskriterium (T–4) für Kontraktionen her kennen. Die neue Fassung wäre

(T*2′) $\not\vdash \neg A \Rightarrow A \in T^*_A$.

Mit (R□) und (R¬□) können wir die angekündigte Attacke *sowohl* gegen (T*P) *als auch* gegen (T*3) *einzeln* führen.

4.4.1. Theorem Es gibt kein schwach nichttriviales TRM, welches
(a)(T*2), (R¬□) und (T*4) *oder*
(b)(T*2), (R□), (R¬□) und (T*3) erfüllt.

Beweis: Sei $\langle \mathbf{T}, ^* \rangle$ ein schwach nichttriviales TRM, sei $T \in \mathbf{T}$ und A ein Satz derart, daß $A, \neg A \notin T$.

(a)	(1) $\neg\Box A \in T$	Voraussetzung $A \notin T$, (R¬□)
	(2) $\neg\Box A \in T^+_A$	(1), Definition 3.1.2
	(3) $\neg\Box A \in T^\circ_A$	(2), Vor. $\neg A \notin T$, (T*4)
	(4) $A \notin T^\circ_A$	(3), (R¬□)
	(5) $A \in T^\circ_A$	(T*2)
	(6)Widerspruch	(4), (5) .
(b)	(1) $\neg\Box A \in T$	Voraussetzung $A \notin T$, (R¬□)
	(2) $A \in T^\circ_A$	(T*2)
	(3) $\Box A \in T^\circ_A$	(2), (R□)
	(4) $\Box A \in T^+_A$	(3), (T*3)
	(5) $A \rightarrow \Box A \in T$	(4), Definition 3.1.2
	(6) $\neg\Box A \rightarrow \neg A \in T$	(5), Aussagenlogik
	(7) $\neg A \in T$	(1), (6), Aussagenlogik
	(8) Widerspruch	(7), Voraussetzung $\neg A \notin T$. □

Jetzt sind (T*4) und (T*3), d.h. *beide* Hälften der These „Konsistente Revisionen sind identisch mit Expansionen" diskreditiert, insoweit als Theoretiker die in ihren Theorien vorhandenen und fehlenden Sätze A mit „autoepistemischen" Sätzen der Form □A bzw. ¬□A festhalten.

[8] Dasselbe gilt für (R◇) unten. Vgl. dazu auch Levis (1979; 1988) ‚serious possibility' und Fuhrmanns (1989) ‚reflective modalities'.

Sehen wir uns den Beweis von Theorem 4.4.1, Teil (b), noch einmal genauer an. Der „Haken" an (T*3) war, daß es die Ableitung von (5) gestattete: wenn $A \rightarrow \Box A$ in T ist, kann — wie im Beweis demonstriert — T nicht A-ignorant sein. $A \rightarrow \Box A$, oder anders geschrieben:

$$(\top \wedge A) \rightarrow (\text{Wenn } \top, \text{ dann } A)$$

darf also kein Satzschema sein, welches in Cn ableitbar ist. Genau dies ist aber in der Konditionallogik von Lewis (1973a) und Gärdenfors (1979) der Fall. Diese Logik basiert auf dem Ramsey-Test (R) und den Eigenschaften der Gärdenforsschen TRMe. Sie hat

(A6) $\quad (A \wedge B) \rightarrow (\text{Wenn } A, \text{ dann } B)$

als Axiomenschema (s. Definition 4.2.1). Gärdenfors zeigt, daß (A6) genau dann in jeder Theorie eines TRMs ist, wenn dieses TRM eine Hälfte von (T*S), nämlich

$$\text{Wenn } A \in T \neq T_{\perp}, \text{ dann } T \subseteq T^*_A ,$$

erfüllt. Nun haben wir (T*S) nie in Zweifel gezogen. Andererseits haben wir gesehen, daß das Axiomenschema (A6) nur triviale TRMe zuläßt — unter der Voraussetzung von (R) und (AEA) (oder von (R$\Box$) und (R$\neg\Box$)). Wieso dürfen wir die Logik von Gärdenfors, die mit der prominenten Logik **VC** von Lewis identisch ist, nicht als die Logik Cn für TRMe verwenden?

Der entscheidende Punkt ist dieser: Gärdenfors benutzt in seinem Beweis, daß — unter (R) — (A6) in jedem T ist, genau die Prämisse, die er in seinem Trivialitätsbeweis so stiefmütterlich behandelt hat, nämlich die Abgeschlossenheit von TRMen gegenüber Expansionsbildung. Die einzig denkbare Rechtfertigung für diese Prämisse wäre, daß Expansionen spezielle Revisionen — eben Additionen — sind. Aber gerade Gärdenfors' eigenes Theorem ist im Ergebnis wohl so zu interpretieren, daß — unter (R) — Additionen eben *nicht* Expansionen sind. Die Prämisse für das obige Axiomenschema (A6), die auch Voraussetzung für die Gültigkeit anderer Gärdenforsscher Axiomenschemata ist, entbehrt damit ihrer Grundlage. (A6) sollte *kein* Axiom der Konditionallogik sein.

Wie bereits in Abschnitt 4.2 erwähnt, gilt wegen (R) und (T*S) folgendes: Wenn $A \wedge B \in T$, dann Wenn A, dann $B \in T$. Dies darf man nicht mit dem Enthaltensein des Axiomenschemas (A6) in Cn verwechseln. Man sieht im Gegenteil, daß aus der metasprachlichen Implikation

Für alle T gilt: Wenn $C \in T$, dann $D \in T$.

nicht auf die „entsprechende" objektsprachliche Implikation geschlossen werden darf, d.h. es folgt nicht

Für alle T gilt: $C \rightarrow D \in T$.

An unserem Beispiel mit (A6) machen wir folgende Beobachtung: (A6) ist aussagenlogisch natürlich äquivalent mit seiner Kontraposition ¬(Wenn A, dann B) → ¬(A∧B). Wenn jedoch (A6), also auch seine Kontraposition, in jedem T wäre, dann würde auch für jedes T gelten: Wenn ¬(Wenn A, dann B) ∈ T, dann ¬(A∧B) ∈ T. Oder umgekehrt formuliert: Wenn ¬(A∧B) ∉ T, dann ¬(Wenn A, dann B) ∉ T. Man betrachte nun ein T, in dem man überhaupt nichts über die Wahrheit A und B weiß, T also A-B-ignorant ist. Es gilt natürlich ¬(A∧B) ∉ T . Es ist aber intuitiv nicht unplausibel, daß ¬(Wenn A, dann B) ∈ T, denn der negierte Konditionalsatz schließt ja gerade aus, daß man etwas über den Zusammenhang von A und B weiß. Dies ist ein weiterer, offenbar von (AEA) unabhängiger Grund gegen die Annahme von (A6) als konditionallogisches Axiomenschema.

Diese Beobachtung könnte einen Ansatz für eine neue Art von Konditionallogik liefern, die auch negierte Konditionalsätze (oder möglicherweise-Konditionalsätze, s. Abschnitt 4.5) in Betracht zieht. Eine objektsprachliche materiale Implikation C→D wäre (dann und?) nur dann ein Theorem solch einer Logik, wenn aufgrund der gegebenen Postulate — (T*2), (T*5), (T*S), (R), (R¬□) etc. — zu zeigen ist, daß für alle Theorien T eines jeden TRMs zweierlei gilt: Wenn C∈T, dann D∈T *und* wenn ¬D∈T, dann ¬C∈T. Ich werde diese Idee hier aber nicht weiterverfolgen.

4.5 Möglicherweise-Konditionalsätze

In der eingangs des letzten Abschnitts beschriebenen Situation, in der B nicht in T^*_A ist, ist Wenn A, dann möglicherweise ¬B normalerweise eine natürlichere Formulierung als das umständliche Satzgefüge Es ist nicht der Fall, daß wenn A, dann B. Anstelle von (R¬□) können wir also folgende Akzeptabilitätsbedingung aufstellen:

(R◇→) Wenn A, dann möglicherweise $B \in T$ gdw. $\neg B \notin T^*_A$.

Entsprechend definieren wir ◇A durch Wenn ⊤, dann möglicherweise A und haben damit

(R◇) $\Diamond A \in T$ gdw. $\neg A \notin T$.

Man kann leicht nachprüfen, daß Teil (a) von Theorem 4.4.1 völlig analog mit (R◇) anstelle von (R¬□) zu beweisen ist. Für Teil (b) brauchen wir die (wenig problematische?) Zusatzbedingung, daß ¬□A und ◇¬A wie üblich austauschbar sind.

Wir wollen nun kurz nachsehen, welche der in der Literatur vorgebrachten Vorschläge zur Analyse von möglicherweise-Konditionalsätzen mit unseren Regeln in Einklang stehen. Die von Lewis (1973a) favorisierte not-would-not-Lesart, als Akzeptabilitätsbedingung geschrieben:

(NWN) Wenn A, dann möglicherweise $B \in T$ gdw.
Es ist nicht der Fall, daß wenn A, dann $\neg B \in T$.

wird natürlich über (R$\Diamond\!\!\rightarrow$) und (R$\neg$) sofort gestützt. Gärdenfors (1988, S. 154–156) akzeptiert zwar (R$\Diamond\!\!\rightarrow$), aber nicht (NWN), denn er hält $\neg$(Wenn A, dann B) für echt stärker als Wenn A, dann möglicherweise B. Allerdings bringt er hier ein recht blasses Argument und enthält sich außerdem einer jeden Stellungnahme zur Akzeptabilität von negierten Konditionalsätzen, so daß seine Zurückweisung von (NWN) insgesamt nicht zu überzeugen vermag. Lewis (1986) stellt (NWN) die oberflächensyntaktisch wohl natürlichste would-be-possible-Lesart gegenüber:

(WBP) Wenn A, dann möglicherweise $B \in T$ gdw.
Wenn A, dann $\Diamond B \in T$.

(WBP) erweist sich in unserem Modell als ebenso korrekt wie (NWN): die rechte Seite heißt mit (R) $\Diamond B \in T^*_A$, was mit (R$\Diamond$) wiederum äquivalent mit $\neg B \notin T^*_A$ ist, und dies ist nach (R$\Diamond\!\!\rightarrow$) genau die linke Seite. Dagegen muß der Vorschlag von Robert Stalnaker (1981, S. 98–101), den man als possibly-would-be-Lesart bezeichnen kann

(PWB) Wenn A, dann möglicherweise $B \in T$ gdw.
$\Diamond$(Wenn A, dann B) $\in T$.

zurückgewiesen werden. Denn hier ist die rechte Seite wegen (R$\Diamond$) gleichwertig mit $\neg$(Wenn A, dann B) $\notin T$, was mit (R$\neg$) einfach $B \in T^*_A$ heißt. Damit aber wäre der möglicherweise-Konditionalsatz vom reinen Konditionalsatz nach (R) nicht mehr zu unterscheiden. Die autoepistemische Allwissenheit ist dafür verantwortlich, daß der Möglichkeitsoperator $\Diamond$ vor einem Konditionalsatz keine Wirkung hat.

Als Analysen der tatsächlich in gesprochener und geschriebener Sprache vorkommenden Konditionalsätze mit möglicherweise im Konsequens überzeugt mich jedoch keiner dieser Vorschläge. Eine der Sprachwirklichkeit nähere Bedeutungsbeschreibung werde ich in Kapitel 6 versuchen.

4.6 Warum Additionen nicht parasitär von Expansionen abhängen

Die ganze Angelegenheit der Additionen wäre kein Problem, wenn das Enthaltensein von „modalisierten" Sätzen, d.h. von Sätzen mit Vorkommnissen von Konditionaloperatoren, $\Box$ oder $\Diamond$, in Theorien allein von den „nichtmodalen" Sätzen irgendwie ursprünglicherer Theorien abhinge. Solchermaßen ist die Ansicht von Isaac Levi (1988):

> I insist that revisions of knowledge or belief are in the first instance revisions of corpora expressible in non modal language ... (S. 69)
> What is the main feature of the position I am taking about belief revision? It is that all revisions are in the first instance revisions of corpora expressible in nonmodal language L. (S. 70)
> ... on the view being proposed here, the revisions of corpora expressible in L** [Sprache mit dem Konditionaloperator] are parasitic on the revisions of corpora expressible in L [Sprache der Aussagen- oder Prädikatenlogik]. (S. 66)

Levi möchte Konditionalsätze und ihren Verwandte aus Theorien („corpora") ausschließen, weil sie nicht Träger von Wahrheitswerten seien. Aus seinen Ausführungen in Levi (1988) geht nicht hervor, *warum* modalisierte Sätze nicht wahrheitsfähig sein sollen, und die entsprechende Darstellung in Levi (1979, S. 228–231) erscheint schwer verständlich. Vielleicht hat er folgende Rechtfertigung im Hinterkopf: Während die in der Sprache der Aussagen- und Prädikatenlogik formulierten Sätze „die objektive Welt" beschreiben, sind Konditionalsätze u.ä., wie (R) anscheinend zeigt, nur Sätze über Theorien und Theorienwandel, also über die Überzeugungen und den Wandel der Überzeugungen des Theoretikers, eines Subjekts. Auch wenn nicht ganz klar ist, weshalb das hinreichen sollte, Konditionalsätzen die Hineinnahme in Theorien zu verweigern, wäre dies immerhin ein qualitativer Schnitt zwischen modalisierten, subjektiven und nichtmodalisierten, objektiven Sätzen.[9]

[9] Levi (Brief vom 8. 1. 1989) hat inzwischen seine Argumentation präzisiert. Seine entscheidende Prämisse ist diese: Wenn ein Satz Träger eines Wahrheitswerts und nichtlogisch ist, dann „sollte es nicht ausgeschlossen sein, daß wir uns in einer Theorie ('belief set') des Urteils über seinen Wahrheitswert enthalten." Levi akzeptiert offenbar (AEA) und hält deshalb Konditionalsätze u.ä. für nicht wahrheitswertfähig. Ich möchte Levis

4.6.1 Der Realitätsgehalt modalisierter Sätze

So einfach gezogen, würde dieser Schnitt aber auf einem Fehlschluß beruhen. Aus (R), (R$\Box$), (R$\Diamond$) etc. kann man ebensowenig ablesen, daß Konditionalsätze etc. „nur" subjektive Sätze über Theorien(wandel) sind, wie man aus

$$A \wedge B \in T \text{ gdw. } A \in T \text{ und } B \in T$$

ablesen kann, daß $A \wedge B$ ein subjektiver Satz über Theorien ist. Wenn Konditionalsätze wirklich nur Sätze über die epistemischen Zustände irgendwelcher Subjekte, die eine Theorie haben, wären, dann könnte man sich bei Konditionalsätzen niemals irren — wir setzen ja autoepistemische Allwissenheit voraus. Man betrachte aber als Beispiel den folgenden Satz meiner Stammtischtheorie T_1:

> Wenn Gorbatschow kurz nach Amtsantritt gestorben wäre, dann wäre der Weltfrieden heute in viel größerer Gefahr, als er es tatsächlich ist.

Wir gehen jetzt einmal davon aus, daß dieser Konditionalsatz, nach allem was wir jetzt — in T_1 — wissen, tatsächlich plausibel ist. Es handelt sich um einen kontrafaktischen Konditionalsatz, da sowohl das Antezedens (A) als auch das Konsequens (C) als falsch unterstellt wird. Sollten wir aber später aus ganz sicherer Quelle erfahren, daß Gorbatschow wirklich kurz nach seinem Amtsantritt gestorben ist, würden wir völlig anders revidieren als bei der entsprechenden hypothetischen Annahme. Da C in der Wirklichkeit unmöglich der Fall sein kann, muß der Konditionalsatz realiter falsch gewesen sein. Es stellt sich — in T_2 — heraus (oder wird einfach nur erschlossen), daß die Sowjetunion heimlich einen Doppelgänger an Gorbatschows Stelle setzte und daß dieser seine Sache außerordentlich gut gemacht hat. Jedenfalls aber wird der Konditionalsatz Wenn A, dann C aufgegeben, da $\neg C$ (ein guter Kandidat für einen analytisch wahren Satz) gegen jede Veränderung resistent ist.[10]

Prämisse anzweifeln. Neben logischen Wahr- und Falschheiten scheint es noch viele andere wahrheitswertfähige Sätze zu geben, bezüglich derer „normale" Sprecher immer ein festes Urteil heben, etwa Eins und eins gibt zwei oder Ich habe Schmerzen. Folgendes Kriterium kommt mir aussichtsreich vor: Wenn man ein falsches Urteil über den Wahrheitswert eines Satzes haben kann, dann ist er tatsächlich Träger eines Wahrheitswerts. S. Abschnitt 4.6.1.

[10] Ähnlich zu verwertende Beispielsätze findet man bei Ramsey (1931, S. 249), Strawson (1952, S. 83), Mackie (1962, S. 71), Adams (1970, S. 90) und Stalnaker (1984, S.

Die Frage in der geschilderten Situation ist, in welchem Sinn man sich mit dem Konditionalsatz eigentlich geirrt hat. Einerseits hat man sich wohl über die Welt geirrt: Sie war einfach nicht so beschaffen, daß sie nach (oder wegen) Gorbatschows frühem Tod in Richtung atomaren Holocaust driftete. Andererseits hat man sich nicht geirrt, was die eigenen Überzeugungen anbetraf. In T_1 rechnete man unter der *hypothetischen Annahme* A mit einer Welt, in der C wahr ist. Deshalb war der Satz Wenn A, dann C zu akzeptieren. Jedoch war es schon in T_1 klar, daß aufgrund der *Neuinformation*, daß Gorbatschow tatsächlich so früh starb, eine Revision von T_1 völlig anders vorgenommen werden würde: Eine solche Neuinformation würde den Irrtum des Konditionalsatzes in T_1 sofort evident machen. An diesem Beispiel sieht man, daß Konditionalsätze zwar Akzeptabilitätsbedingungen haben, die auf hypothetische Revisionen Bezug nehmen, daß sie aber nichtsdestotrotz Aussagen über die wirkliche Welt sind. Offenbar geben hypothetische Annahmen Anstoß zu ganz anderen Revisionen als Neuinformationen.[11] Dieser grundlegende Unterschied, den man zwischen dem für den Ramsey-Test geeigneten $T_1{}^*{}_A$ und dem T_2 der obigen Schilderung machen muß, untergräbt auch das von Gärdenfors (1988, S. 166) angeführte Argument gegen den Ramsey-Test.

Die Dynamik unserer Theorien spiegelt genauso die Realität wider wie unsere Theorien selbst, und es gibt von daher keinen Grund, warum man diese Dynamik nicht mittels (R) (konditional-)satzförmig in den Theorien selbst fassen dürfen sollte. Konditionalsätze sind gegenüber anderen, unstrittig realitätsbeschreibenden Satzgefügen syntaktisch nicht ausgezeichnet, und semantisch offenbar aufs Engste verwandt mit Dispositionsprädikaten, Naturgesetzen, Kausalausdrücken, welche wohl allesamt objektive Ausschnitte der Wirklichkeit wiedergeben. Levi (1988, S. 75–78) unterscheidet zwar zwischen wahrheitswertfähigen Scheinkonditionalen, welche eigentlich Dispositionen ausdrücken, und „echten" Konditionalsätzen.[12]

105f), welch letzterer das Beispiel von Strawson aufgreift. — Ich vermute, es gibt einen engen Zusammenhang zwischen der unten erörterten Dichotomie „Neuinformation vs. hypothetische Annahme" und dem Problem „indikativische vs. konjunktivische Konditionalsätze", den ich aber an dieser Stelle nicht diskutieren möchte. Vgl. auch Kapitel 6, Fußnote 19.

[11] Es ist aber möglich, daß beide Arten von Revision denselben Rationalitätskriterien genügen. Der Unterschied liegt vermutlich in den völlig verschiedenartigen Relationen der theoretischen Wichtigkeit, die diesen beiden Revisionsvarianten zugrundeliegen. Vgl. Kapitel 3, Fußnote 35 und Kapitel 6, Fußnote 19.

[12] Es gibt noch weitere Arten von Konditionalsätzen bei Levi. Steht beispielsweise $\Diamond A$ im Antezedens eines Konditionalsatzes, so sei dieser durch Betrachtung der Kontraktion $T^-{}_{\neg A}$ zu analysieren (Levi 1988, S. 70–72). Levi hat keine einheitliche Analyse von

Aber er gibt uns kein Entscheidungskriterium an die Hand — und es ist auch keines in Sichtweite — wie man die Unterscheidung erkennen und ziehen kann.[13] Deshalb erscheint mir seine Hierarchie zwischen Sätzen erster Klasse (Sätzen in der Sprache der Aussagen- und Prädikatenlogik) und modalisierten Sätzen zweiter Klasse, die „parasitär" von den ersteren abhängen, nicht überzeugend.

4.6.2 Eine Abhängigkeit nichtmodalisierter Sätze von modalisierten Sätzen

Man mag trotz allem eine Trennung von „objektiven" und epistemisch-modalisierten Sätzen für möglich und für nötig halten. Dann läßt sich eine zweite Argumentationsstrategie gegen Levis vereinfachende Sichtweise in den sogenannten nichtmonotonen oder autoepistemischen Logiken finden, die in Kapitel 5 zur Sprache kommen werden. Im täglichen Leben müssen wir laufend auf der Basis unvollständigen Wissens Entscheidungen treffen und Handlungen ausführen. Obwohl wir eigentlich nicht explizit wissen, ob A oder ¬A der Fall ist, gehen wir in unseren Alltagstheorien bei zahllosen Instantiierungen von A davon aus, daß A und nicht ¬A der Fall ist: davon, daß mein Gegenüber nicht verrückt ist, daß das eben abgestellte Auto auch jetzt noch fahrtüchtig ist, daß das kleine Fritzchen (noch lebende) Eltern hat, daß der Vogel Tweety fliegt, daß der Angeklagte unschuldig ist usw. „Bis zum Beweis des Gegenteils" unterstellen wir für solche Sachverhalte einfach, daß sie vorliegen. Wir nehmen den Satz A (und nicht den Satz ¬A) in unsere „Theorien über die Welt" mit auf, weil wir uns auf der Basis dieser Theorien orientieren und handeln müssen.

Die Sorte von Sätzen A, die „normalerweise" oder „bei fehlender Gegenevidenz" für wahr gehalten werden, kann vermittels von „Axiomen" der Form $(\Diamond A) \rightarrow A$ (hier natürlich *nicht* als Schema zu lesen!) als „Default-Wissen" in Theorien hineingebracht werden. In den konsistenten „stabilen" Theorien T der autoepistemischen Logik von Moore (1985) (vgl. auch Stalnaker 1980) gehorchen $\Box$ und $\Diamond$ genau den durch $(R\Box)$ und $(R\Diamond)$ gegebenen Bedingungen. Eine A-Ignoranz kann so ausgeschlossen werden: Falls man ¬A nicht explizit weiß, hat man $\Diamond A$, und so mit dem Axiom $(\Diamond A) \rightarrow A$ den „Normalwert" A.

Wenn man will, kann man hier in Umkehrung von Levis Dictum davon sprechen, daß die nichtmodalisierten Sätze A „parasitär" von den moda-

Konditionalsätzen, die der Analyse vermittels (R) vergleichbar wäre.

[13] Siehe aber Dudman (1984).

len Sätzen ◇A abhängen. Es ergeben sich für diesen Ansatz eine Fülle interessanter Probleme eigener Art,[14] er berücksichtigt jedoch wichtige Intuitionen über unser alltägliches Räsonieren und hat sich, implementiert in Computern, bereits praktisch bewährt.

4.7 Konditionalsätze und Nichtmonotonie

Gärdenfors' Angriff auf die Kombination des Ramsey-Tests mit dem Erhaltungskriterium (T*P) ist nicht ganz stichhaltig gewesen. Tatsächlich ist der Ramsey-Test unverträglich mit dem Vorschlag, konsistente Revisionen mit Expansionen gleichzusetzen. Noch genauer genommen, kann man die Unvereinbarkeit von (R) mit der Kollektion sämtlicher grundlegender Gärdenfors-Postulate zeigen. Eine direkte Auswirkung von (R) auf (T*P) konnte ich ebensowenig wie Gärdenfors nachweisen.

Während Gärdenfors für (T*P) und gegen (R) plädiert, möchte ich hier für (R) und gegen (T*E) stimmen. Wenn man sich aufgrund der Diskussion in diesem Kapitel entscheiden müßte, *was* denn an der ganzen Angelegenheit inadäquat ist, dann würde intuitiv vermutlich am meisten gegen (T°M) — und damit gegen (T*4) — sprechen. Bei einer naiven Betrachtung der Skizze 4.2 wird man wohl zu dem Schluß kommen, daß Wenn ¬A, dann B zwar in $T°_{A \vee B}$, aber nicht in $T°_{A}$ hinein gehört. Die Varianten des Trivialitätsbeweises aus den Abschnitten 4.2 und 4.3 liefern also kein zwingendes Argument gegen den Ramsey-Test.

Die Qual der Wahl, ob (T*3) oder eher (T*4) aufzugeben sei, wird einem abgenommen, wenn man sich auf das Prinzip der autoepistemischen Allwissenheit einläßt. Ich möchte noch einmal betonen, daß (AEA) im Kontext des hier zugrundegelegten, ohnehin stark idealisierenden Theorienmodells *keine* besonders starke Forderung darstellt. Mit (AEA) zeigt sich, daß beide Hälften von (T*E), sowohl (T*3) als auch (T*4), inakzeptable Konsequenzen nach sich ziehen. Dies wurde in Abschnitt 4.4 mit negierten Konditionalsätzen gezeigt. Wenn man mit möglicherweise-Konditionalsätzen (Abschnitt 4.5) arbeitet, klappt die Trivialisierung von (T*4) problemlos, die von (T*3) benötigt jedoch als Zusatzvoraussetzung, daß negierte und möglicherweise-Konditionalsätze oder daß ¬□A und ◇¬A äquivalent sind. Gärdenfors bestreitet dies zwar, aber nur um den Preis ei-

[14] Beispielsweise existiert für eine beliebige „Axiomenmenge" T_o nicht immer eine eindeutig bestimmte Obermenge T, die deduktiv abgeschlossen ist und (R◇) erfüllt. Beispiel: $T_o = \{(\Diamond A) \rightarrow \neg B, (\Diamond B) \rightarrow \neg A\}$. Siehe Kapitel 5.

nes völligen und wahrscheinlich unumgänglichen Verzichts auf die Analyse von negierten Konditionalsätzen.

Derselbe Verzicht, dieselbe Lücke ist es auch, die es Gärdenfors ermöglicht, sein Programm einer Analyse von Konditionalsätzen durch den Ramsey-Test aufrecht zu erhalten. In Gärdenfors (1979) und in Gärdenfors (1988, Kapitel 7) ist die Identität von Additionen und Expansionen *allein* deshalb nicht mit eingebaut, weil nach Gärdenfors aus $B \notin T^*_A$ noch nicht $\neg$(Wenn A, dann B) $\in$ T folgen soll.[15] Trotz der so erkauften Immunität seines Programms zieht Gärdenfors die Konsequenz, aufgrund des Trivialitätstheorems den Ramsey-Test aufzugeben. Das scheint mir, wie gesagt, nicht hinreichend motiviert. Aber es muß auch zugegeben werden, daß die Preisgabe von (T*E) und der Abgeschlossenheit von TRMen unter Expansionsbildung, welche ich in diesem Kapitel befürwortet habe, einen ebenfalls ganz entscheidenden Schlag gegen das Gärdenforssche Programm darstellt. So oder so gerät Gärdenfors' Ramsey-Test-Rekonstruktion der Lewisschen Standardlogik **VC** für kontrafaktische Konditionalsätze ins Wanken. Am Ende von Abschnitt 4.4 haben wir indes gesehen, daß diese Logik mit ihrem Axiomenschema (A$\wedge$B)$\rightarrow$(Wenn A, dann B) als Grundlage für TRMe ohnehin nicht adäquat sein kann — wenn man autoepistemische Allwissenheit voraussetzt.[16]

Es ist klar, daß (T*E) nichts taugt, wenn wir (R$\neg$), (R$\neg\Box$), (R$\Diamond\rightarrow$) oder (R$\Diamond$) zur Anwendung bringen. Wie unmittelbar einsichtig, führen in (schwach) nichttrivialen TRMen Expansionen z.B. aus (R$\Diamond$) heraus. Sei T eine A-ignorante Theorie aus einem TRM, welches (R$\Diamond$) erfüllt. Mit A,$\neg$A$\notin$T ist dann $\Diamond\neg$A in T und damit auch in T^+_A. Aber auch A ist in T^+_A, weswegen T^+_A nicht mehr (R$\Diamond$) genügen kann, also nicht im TRM ist. *Keine* „A-erfolgreiche" Obermenge eines A-ignoranten T bleibt innnerhalb von (R$\Diamond$). Ja, es gibt in (R$\Diamond$) erfüllenden TRMen sogar keine zwei Theorien T_1 und T_2, die durch echte Inklusion geordnet sind: für jedes A $\in T_1 \backslash T_2$ ist $\Diamond\neg A \in T_2 \backslash T_1$.

Diese gravierenden Feststellungen bzgl. negierter und möglicherweise-

[15] Der entscheidende Punkt ist Gärdenfors' Modifikation von (T*8) zu

(T*8′) Wenn $\neg$(Wenn A, dann $\neg$B) $\in$ T, dann $(T^*_A)^+_B \subseteq T^*_{A\wedge B}$.

Mit (R$\neg$), (T*S) und $\top$ für A würde aus dieser Bedingung und (T*7) sofort (T*E) folgen. Ohne (R$\neg$) tritt dieser Mangel nicht auf, jedoch ist dann (T*8′) eigentlich gar nicht zu verstehen. Die kritische Frage bleibt unbeantwortet: Unter welchen Umständen soll $\neg$(Wenn A, dann $\neg$B) in T sein? — Auf das analoge Problem bzgl. (T*P) statt (T*E) hat Isaac Levi (1988, S. 68, 80) hingewiesen.

[16] Eine unabhängige, intuitiv motivierte Kritik an diesem Axiomenschema werden wir in Abschnitt 6.3 vorbringen.

Konditionalsätze einerseits und die intuitive Fragwürdigkeit von (T°M) in Skizze 4.2 andererseits lassen es als äußerst zweifelhaft erscheinen, daß es in auf Sprachen, die nur „einfache" Konditionalsätze enthalten, basierenden TRMen so etwas wie monotones Lernen gibt. Wenn wir das $T^\circ_{A \vee B}$ aus Skizze 4.2 jetzt einmal mit T′ bezeichnen, dann ist es nicht unplausibel anzunehmen, daß T'°_A mit T°_A identisch ist. Mit $T^\circ_{A \vee B} \subseteq T^\circ_A$ gerät dann auch $T' \subseteq T'^\circ_A$, d.h. das Erhaltungskriterium (T*P), ins Zwielicht. Außer in sehr speziellen Fällen wird man vermutlich *immer* Konditionalsätze der Form Wenn ¬A, dann ... finden, die in T′, aber nicht in T'°_A sind. Denn die „größere Unwissenheit" von T′ läßt Raum für Möglichkeiten, die im „wissenderen" T'°_A verloren gehen.[17] Man kann hier darüber spekulieren, ob es vielleicht Argumente dafür gibt, daß auch in TRMen, die nur (R) erfüllen, keine zwei durch echte Inklusion geordnete Theorien existieren sollten. Mit formalen Ergebnissen in dieser Richtung kann ich aber bis jetzt nicht aufwarten.

Wenn man aber davon ausgeht, daß in Sprachen mit (ggf. negierten oder möglicherweise-) Konditionalsätzen keine zwei Theorien T_1 und T_2 mit $T_1 \subseteq T_2$ existieren, dann ist es klar, warum die von Gärdenfors ins Zentrum seiner Diskussion des Trivialitätsbeweises gerückte *Monotoniebedingung für TRMe*

(T*M) Wenn $T_1 \subseteq T_2$, dann $T_1{}^*_A \subseteq T_2{}^*_A$

ganz und gar unschädlich ist. (T*M) folgt, wie Gärdenfors (1986) vermerkt, trivial aus (R): Sei $B \in T^*_A$, dann ist nach (R) Wenn A, dann B in T, also nach dem Antezedens von (T*M) auch in T′, also ist B nach (R) auch in T'^*_A , womit wir das Konsequens von (T*M) bewiesen haben. Es mag durchaus sein, daß Gärdenfors (1986, S. 86f; 1988, S. 159f) mit seinen Argumenten gegen (T*M) im Prinzip recht hat, daß (T*M) aber dennoch wahr ist — eben trivialerweise, weil es keine echte Inklusionen zwischen Theorien gibt. Dem Ramsey-Test zum Vorwurf zu machen, daß er (T*M) impliziert, ist also keine vielversprechende Strategie.[18]

Eine weitere Folge der Ergebnisse dieses Kapitels ist es, daß es keinen rechten Sinn mehr macht, von „Erweiterungen" („expansions") und

[17] Es sei denn, der folgende sehr spezielle Fall liegt vor: Mittels der Relation der theoretischen Wichtigkeit ist ein „Gedächtnis" in T'°_A eingebaut, welches sagt, daß unter der Annahme ¬A von T'°_A auf T′ zurückgegangen werden soll. Ein solches Gedächtnis, das die zuletzt erworbene Information am niedrigsten bewertet, ist aber im allgemeinen nicht wünschenswert.

[18] Gärdenfors hat auch noch andere Argumente gegen den Ramsey-Test.

„Verkleinerungen" („contractions") zu sprechen.[19] Wenn wir $\neg\Box$ oder $\Diamond$ in der Objektsprache haben und diese Operatoren den Akzeptabilitätsbedingungen (R$\neg\Box$) bzw. (R$\Diamond$) gehorchen, dann gibt es keine echten Erweiterungen und Verkleinerungen mehr. Die einzige Art von Theorienwandel ist die Revision. Entsprechend wäre es suggestiver, scheinbare Expansionen und Kontraktionen gleich als Revisionen anzuschreiben: Statt $T^{\circ}{}_{A}$ — bei Gärdenfors identisch mit $T^{+}{}_{A}$[20] — schreiben wir besser $T^{*}{}_{\Box A}$ und statt $T^{-}{}_{A}$ besser $T^{*}{}_{\neg\Box A}$ oder $T^{*}{}_{\Diamond\neg A}$.[21] (Es ist leicht nachzuprüfen, daß die Erfolgs- und Stabilitätskriterien (T*2) und (T*S) zusammen mit (R$\Box$), (R$\neg\Box$) bzw. (R$\Diamond$) auch den Erfolg und die Stabilitität von so eingeführten Additionen und „Kontraktionen" garantieren.) Diese Schreibweise macht deutlich, daß jeder Theorienwandel — auch der prima facie „konsistente" — in gewissem Sinn mit den ursprünglichen Theorien im Widerstreit steht: Eine Addition widerspricht einer vorher einkalkulierten Möglichkeit, eine „Kontraktion" einer vorher angenommenen Notwendigkeit (Unmöglichkeit). Ich hoffe, in Abschnitt 4.6 hinreichend plausibel gemacht zu haben, daß dies keine künstlichen, aufgesetzten Widersprüche sind, sondern solche, die wesentlich verstrickt sind in der komplexen Vernetztheit unserer Theorien über die Welt.

Ein logisches Phänomen bleibt sonderbar und störend. Wenn $\neg A \notin T$, dann möchte man $T^{\circ}{}_{A}$ eigentlich gerne mit $Cn(T \cup \{A\})$ ansetzen. Wie wir sahen, ist dies aber wegen beispielsweise (R$\Diamond$) nicht möglich. Doch es gibt eine Lösung, die unserem Wunsch entgegenkommt. Man darf hierzu eine Theorie T nicht als einen erratischen Block auffassen, sondern als eine Satzmenge, die sich aus dem „expliziten Wissen" T_e und dem „impliziten Wissen" $T_i = T \backslash T_e$ zusammensetzt. Das implizite Wissen wird aus dem expliziten mit Hilfe von (R$\Box$), (R$\neg\Box$) und (R$\Diamond$) und einer zugrundegelegten Logik Cn abgeleitet. In Abschnitt 4.6 habe ich dafür plädiert, daß T_e *in aller Regel* keineswegs nur ein gewisser nichtmodalisierter Teil von T ist.

Nachdem wir also $T = Cn(T_e)$ haben, können wir jetzt $T^{\circ}{}_{A}$ mit $Cn(T_e \cup \{A\})$ ansetzen. Was ist damit gewonnen? Mit $Cn(T \cup \{A\})$ — die urspüngliche Idee — erhielten wir bei jeder Art von Logik Cn eine Obermenge von T, und dies war verantwortlich für den Verstoß zum Beispiel

[19] Die Wahl meiner Bezeichnung „Addition" kann allerdings auch nur durch den Mangel an besseren Namen gerechtfertigt werden.

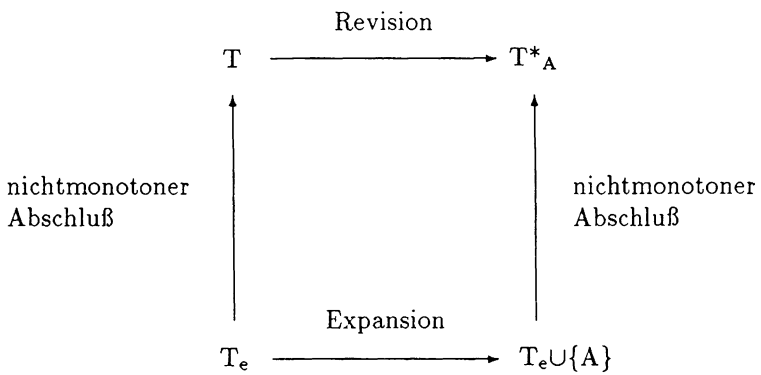
[20] Siehe aber die Diskussion bei Fußnote 15.

[21] Fuhrmann (1989, S. 132) macht in seiner Argumentation gegen Levi den umgekehrten Vorschlag und definiert Revisionen als spezielle Kontraktionen mittels $T^{*}{}_{A} = T^{-}{}_{\neg\Box A}$.

gegen (R◇). Mit $Cn(T_e \cup \{A\})$ hingegen erhalten wir nur dann sicher eine Obermenge von T, wenn die Logik Cn *monoton* ist, d.h. wenn für Cn gilt:

Wenn $T_e \subseteq T'_e$, dann $Cn(T_e) \subseteq Cn(T'_e)$.

Daß die in unserem alltäglichen Denken verwendetete Logik aber allem Anschein nach nicht monoton ist, dies ist ein Beitrag der Forschung auf dem Gebiet der Künstlichen Intelligenz, aus der die sogenannten nichtmonotonen oder autoepistemischen Logiken erwachsen sind. Durch die Aufspaltung einer Theorie in einen expliziten und einen erschlossenen impliziten Anteil und durch die Verwendung einer nichtmonotonen Logik erhält die Intuition, daß konsistente Revisionen oder „Additionen" einfach identisch mit dem logischen Abschluß der alten Wissensbasis plus des neuen Satzes sind, zu guter Letzt doch wieder neue Hoffnung. Die Grundzüge der Idee werden in Skizze 4.3 dargestellt.



SKIZZE 4.3

Zumindest einige der in diesem Kapitel georteten Verwicklungen sollten sich dadurch auflösen lassen.[22] Da das sogenannte nichtmonotone Schließen einerseits von beträchtlichem philosophischen, logischen und wissenschaftstheoretischen Interesse ist, da es andererseits aber in philosophischen Fachkreisen bisher noch wenig Verbreitung gefunden hat,[23] will ich nun in einem eigenen Kapitel gewisse für den vorliegenden Zusammenhang besonders relevante Entwicklungen der nichtmonotonen Logik aufzeigen.

[22] Eine präzise Ausarbeitung der in diesem Kapitel angelegten Ideen findet man jetzt in Rott (1990)

[23] Dieser Satz wurde 1987 geschrieben!

Kapitel 5

Nichtmonotone und autoepistemische Logik

5.1 Schlüsse aus der Unwissenheit

Bei der Verwaltung unserer Alltagstheorien ziehen wir laufend *Schlüsse aus unserem Nichtwissen*. Erinnern wir uns an die in Abschnitt 4.6.2 genannten Beispiele. Wenn ich nicht weiß, daß mein Gesprächspartner verrückt ist, nehme ich an, daß er es nicht ist. Wenn niemand feststellt, daß dieses Auto kaputt ist, gehe ich davon aus, daß es fährt. Ohne explizite Information, daß das Gegenteil der Fall ist, unterstellen wir — meistens ohne lang nachzudenken — daß der Vogel namens Tweety fliegen kann, daß Fritzchen Eltern hat und daß der Angeklagte nicht schuldig ist. Die zugrundeliegende Überlegung hier ist: „Wenn es anders wäre, würde (sollte, müßte) ich es schon wissen."

Die Bedeutung solcher Schlüsse ergibt sich aus der Notwendigkeit, ständig auf der Basis unvollständigen Wissens Entscheidungen zu treffen und zu handeln. Wenn wir keine Evidenz für das Gegenteil haben, setzen wir „Default-Werte" an, die „normalerweise", „unter normalen Umständen" zutreffend sind. Erst wenn eine Handlung fehlschlägt, wird neu überlegt: welche Informationen haben mir gefehlt, um mein Wissen adäquat zu machen?

„*Wissen*" meint hier und im folgenden eigentlich nicht Wissen im strengen Sinn: Ich setze nicht voraus, daß das, was wir „wissen", auch wahr ist.

Bessere Bezeichnungen wären also „zu wissen glauben", „überzeugt sein von", „für wahr halten", „akzeptieren" oder „als (Bestandteil der) momentane(n) Theorie über die Welt haben". Aus Bequemlichkeitsgründen und weil sie sich eingebürgert hat, bleibe ich aber bei der schlampigeren Ausdrucksweise. *Explizites Wissen* (die Datenbasis, Grundinformationen, nicht problematisierte Evidenz, vorgegebene Prämissen etc.) wird im folgenden durch *„(nichtlogische) Axiome"* bezeichnet. Dagegen heißt *„Wissen"* alles, was aus diesen Axiomen (kalkülisiert oder auch nicht) ableitbar ist. Entsprechend ist *„Nichtwissen"* das, was aus den Axiomen nicht ableitbar und mit ihnen konsistent ist.

Sei S nun eine Menge von gegebenen Axiomen. Dann haben mit den eben eingeführten Bezeichnungen die obigen Beispiele die folgende informelle Struktur:

(5.1.1) Wenn A konsistent mit S ist, dann darf man (aus S) auf A schließen.

Bei der Betrachtung von (5.1.1) wird deutlich, daß die sog. „Monotonie" der klassischen und vieler anderer Logiken (z.B. intuitionistische Logik, Modallogik) verloren geht. Eine Logik Cn kann als (Hüllen-)Operator auf Satzmengen S aufgefaßt werden, der alle logischen Konsequenzen von S liefert. Die *Monotoniebedingung* für Cn kann man dann so schreiben:

(M) $\quad S_1 \subseteq S_2 \Rightarrow Cn(S_1) \subseteq Cn(S_2)$.

Die üblichen kalkülisierten Logiken erfüllen (M),[1] was daran liegt, daß ihre Schlußregeln die Axiome „permissiv" (Minsky 1974) machen: Sie haben stets die Form „Wenn man auf A_1, A_2,..., A_n schließen kann, dann kann man auch auf B schließen."

Daß die Bedingung (5.1.1), wenn man sie irgendwie in die Logik Cn einbauen kann, (M) unmöglich macht, sieht man schon am Beispiel eines sowohl mit A als auch mit $\neg$A konsistenten S. Sei S' definiert durch $S'=S\cup\{\neg A\}$. Offenbar gilt in Cn nach Anwendung der „Schlußregel" (5.1.1) $A\in Cn(S)$, während man auf keinen Fall $A\in Cn(S')$ haben will — sonst wäre $Cn(S')$ ja inkonsistent, obwohl S' konsistent ist. (Hier ging als Voraussetzung die auch i.f. stets präsupponierte *Reflexivität* von Cn ein, d.h. die Bedingung $S'\subseteq Cn(S')$.)

Eine Schwierigkeit von (5.1.1) haben wir bisher ohnehin unterschlagen: Wenn S mit A und mit $\neg$A konsistent ist (d.h. wenn man auf der Basis der

[1] Ganz neu ist die Infragestellung von (M) jedoch nicht; wie Segerberg (1982, S. 37) bemerkt, wird (M) in verschiedenen Varianten der Relevanz-, Quanten- und der induktiven Logik zurückgewiesen.

Informationen von S nichts über A weiß), dann könnte man (5.1.1) ja im Prinzip sowohl auf A als auch auf ¬A anwenden und damit auf A und ¬A schließen, d.h. auf eine Inkonsistenz. Das ist natürlich unerwünscht. Auf welchen Satz soll man (5.1.1) aber anwenden?[2]

Dieses Argument zeigt schon, daß (5.1.1) als allgemeine Schlußregel untauglich ist — wie man natürlich auch intuitiv erwartet. Man kann nun die Möglichkeit wählen, eine Liste von inhaltlich motivierten Schlußregeln der Art (5.1.1) für eine gewisse Menge von Sätzen A_1, A_2, A_3 ... explizit anzugeben; die A_i's sind dann nicht mehr als Metavariable für beliebige Sätze zu verstehen, sondern als konkrete objektsprachliche Sätze. Eine Verallgemeinerung dieser Idee, nämlich Schlußregeln der Form

(5.1.2) Wenn man A aus S erschließen kann und wenn B konsistent mit S ist, dann darf man (aus S) auf C schließen.

legt Reiter (1980) seiner „Logic for Default Reasoning" zugrunde (wobei C eine strukturelle Beziehung zu A und B hat). Eine aktuelle Bestandsaufnahme und neue Entwicklungen dieses Ansatzes findet man in Etherington (1987).

Eine allgemeinere, elegantere und ehrgeizigere Strategie ist es, eine objektsprachliche Ausdrucksmöglichkeit für „A ist konsistent (mit S)" einzuführen.[3] Sei im folgenden der objektsprachliche Ausdruck „MA" (M — für „möglich" — ist dabei ein Satzoperator) die entsprechende Symbolisierung. Dann kann man (5.1.1) zu dem untadelig aussehenden

(N) Wenn A konsistent mit S ist, dann darf man (aus S) auf MA schließen.

abschwächen und die inhaltlichen Überlegungen der im Einleitungsabschnitt geschilderten Art als Axiome anschreiben, z.B. M(nicht-verrückt(a)) → nicht-verrückt(a). Diese Idee wurde zuerst von McDermott und Doyle (1980) untersucht. Später wurde der Ansatz von McDermott (1982), Moore (1984, 1985) und Konolige (1988) verbessert. In diesem Kapitel sollen

[2] In der Praxis ist dieses Problem nicht ganz so groß. Man benutzt ja i.a. keine Satzbuchstaben, sondern hat als „kleinste Einheiten" atomare Formeln mit irgendwie „natürlichen" Prädikaten und Objektbezeichnungen. Wenn man z.B. S = {Student(Franz), Student(Olli), Dozent(Maria)} als Axiomenmenge hat, schließt man typischerweise auf nicht-Dozent(Franz), nicht-Dozent(Olli) und nicht-Student(Maria), obwohl ja das jeweilige Gegenteil genauso konsistent mit S wäre. Allgemein pflegt man bei „natürlichen" Prädikaten P (5.1.1) normalerweise eher auf ¬P(x) als auf P(x) anzuwenden. Das grundlegende philosophische Problem, was denn ein „natürliches" Prädikat ist, wäre dabei natürlich erst noch zu lösen.

[3] Konolige (1988) hat jedoch gezeigt, daß der Unterschied zwischen dieser und Reiters Methode doch nicht ganz so groß ist, wie es scheint. Vgl. Abschnitt 5.8.

die dabei entwickelten „nichtmonotonen" und „autoepistemischen" Logiken vorgestellt werden.[4]

Bevor wir dies tun, wollen wir aber darauf hinweisen, in welch engem Zusammenhang die folgenden Überlegungen mit den in Kapitel 4 diskutierten Problemen von TRMen stehen. Der Anschluß kann durch das folgende kleine Theorem hergestellt werden:

5.1.3. Theorem Sei $\langle \mathbf{T},^* \rangle$ ein TRM und S eine Axiomenmenge derart, daß $Cn(S) \in \mathbf{T}$.

(a) Erfüllt $\langle \mathbf{T},^* \rangle$ die Bedingungen (T*4) und (R), dann gilt:
Wenn $S \not\vdash \neg A$, dann $S \vdash$ Wenn A, dann B für alle $B \in S$.

(b) Erfüllt $\langle \mathbf{T},^* \rangle$ die Bedingung (R$\Diamond$), dann gilt:
Wenn $S \not\vdash \neg A$, dann $S \vdash \Diamond A$.

Der Beweis ist einfach. Seien die Voraussetzungen des Theorems erfüllt, erfülle $\langle \mathbf{T},^* \rangle$ für Teil (a) (T*4) und (R), und gelte $S \not\vdash \neg A$; d.h. $\neg A \notin Cn(S) \in \mathbf{T}$, also nach (T*4) $(Cn(S))^*_A = (Cn(S))^\circ_A \supseteq (Cn(S))^+_A \supseteq Cn(S) \supseteq S$. Mithin ist jedes $B \in S$ in $(Cn(S))^*_A$, woraus wegen (R) folgt, daß für jedes $B \in S$ Wenn A, dann B in Cn(S) ist, d.h. $S \vdash$ Wenn A, dann B. Erfülle für Teil (b) $\langle \mathbf{T},^* \rangle$ nun (R$\Diamond$) und gelte $S \not\vdash \neg A$; d.h. $\neg A \notin Cn(S) \in \mathbf{T}$, also nach (R$\Diamond$) $\Diamond A \in Cn(S)$, d.h. $S \vdash \Diamond A$. □

In Teil (b) von Theorem 5.1.3 haben wir einen mit (N) praktisch identischen Ausdruck stehen. Wenn wir *annehmen*, daß $\mathbf{T}$ die Klasse *aller* Theorien ist, insbesondere also für jede Axiomenmenge S Cn(S) enthält,[5] und daß $\langle \mathbf{T},^* \rangle$ (R$\Diamond$) erfüllt, dann erhalten wir in (b) die uneingeschränkte „Schlußregel" (N). Wir werden aber bald sehen, daß (N) und also auch die eben gemachte Annahme nicht unproblematisch sind.

Durch Teil (a) von Theorem 5.1.3 stellt sich folgendes heraus: Unter der *Annahme* von nur (T*4) und (R) für ein TRM $\langle \mathbf{T},^* \rangle$, bei dem $\mathbf{T}$ die Klasse aller Theorien ist, muß die Logik Cn eine Art Konditionallogik enthalten. Cn zeigt hinsichtlich von Konditionalsätzen ein nichtmonotones Verhalten, das dem Verhalten gegenüber Sätzen der Form $\Diamond A$ völlig analog ist. Dieser Aspekt wurde meines Wissens noch überhaupt nicht gesehen und ist in seiner Tragweite sehr schwer einzuschätzen. Für die Annahme hier gilt, wie in Kapitel 4 argumenitert, jedoch derselbe Vorbehalt wir für die Annahme bei Teil (b). Anstatt in dem unbekannten Gebiet einer nichtmonotonen

[4] Eine ausführliche Einführung in Default Logic, Nonmonotonic Logic und „Circumscription", den ebenfalls vieldiskutierten alternativen Ansatz McCarthys (1980), findet man in Łukaszewicz (1986). Vergleiche auch Kapitel 5 von Turner (1984).

[5] Beachte, daß Cn(S) wegen der vorausgesetzten Idempotenz von Cn eine Theorie im Sinne von Definition 3.1 ist: $Cn(Cn(S)) = Cn(S)$.

Konditionallogik ohne Leitfaden herumzuirren,[6] will ich nun daran gehen, die Ergebnisse der Erforschung des inzwischen etwas besser bekannten Terrains der nichtmonotonen Logik auf der Grundlage einer Regel wie (N) darzustellen.

5.2 Nichtmonotone Logik I

An einem Beispiel machen wir uns klar, daß (N) als uneingeschränkte Schlußregel nicht so harmlos ist, wie man zunächst meinen könnte. Betrachten wir die Sätze

A: Ich gewinne den 1.Preis im Preisausschreiben.
B: Ich muß mein Auto verkaufen.

Angenommen, aus meiner aktuell akzeptierten Axiomenmenge S ist weder $\neg$A noch $\neg$B ableitbar, d.h. sowohl A als auch B sind mit S konsistent. Also scheint man mit (N) auf MA und MB schließen zu dürfen: Möglicherweise gewinne ich im Preisausschreiben, aber es ist auch möglich, daß ich mein Auto verkaufen muß. Nun kann mein S aber durchaus den Satz MA$\rightarrow\neg$B enthalten: Solange es möglich ist, daß ich im Preisausschreiben gewinne, verkaufe ich mein Auto auf keinen Fall. Dann ist jedenfalls die simultane Anwendung der Regel (N) unmöglich; denn wenn man MA (als Resultat der Anwendung von (N) auf A) zu S dazunimmt, ist $\neg$B ja ableitbar, was die Anwendung von (N) auf B blockiert. Man kann nur *entweder* MA (und damit auch $\neg$B) *oder* MB (und damit, zumindest intuitiv, nicht mehr $\neg$B) haben. Man muß aufpassen, daß man mit (N) keine Denkfehler macht.

Es ist die „Nicht-Permissivität" der Regel (N), die solche Schwierigkeiten hervorruft. Sehen wir nun, wie McDermott und Doyle (1980) das Problem zu lösen versuchen.

Sei R eine beliebige Menge von Sätzen der Sprache der Prädikatenlogik erster Stufe. Dann bezeichne

(5.2.1) $M(R) := \{MA: \neg A \notin R\}$

die Menge der *R-Möglichkeiten*. Für eine gegebene Axiomenmenge S wird durch

$$NM_S(R) := Th(S \cup M(R))$$

ein Operator auf Formelmengen definiert, der zu einer beliebigen Formelmenge R die aus S und den R-Möglichkeiten klassisch (d.h. insbesondere monoton) ableitbaren Theoreme liefert. Th(R) bezeichne hier die Menge

[6] Vgl. jetzt aber Rott (1990).

der aus R, vermittels der Prädikatenlogik erster Stufe zu erhaltenden Theoreme.)

5.2.3. Definition Eine Formelmenge T mit $T = NM_S(T)$ heißt *Fixpunkt von NM_S* oder einfach *Fixpunkt von S*.

Ein Fixpunkt T ist eine Obermenge der klassisch erhaltenen Theorie Th(S) von S, die ihre eigenen (d.h. die T-) Möglichkeiten enthält und in der bereits alle möglichen Schlüsse daraus gezogen sind. Ein solches T kann als vernünftige nichtmonotone Erweiterung von S betrachtet werden.

Nicht jedes S hat einen Fixpunkt (z.B. hat $S = \{MA \rightarrow \neg A\}$ keinen), und viele S haben mehrere Fixpunkte (z.B. hat $S = \{MA \rightarrow \neg B, MB \rightarrow \neg A\}$ zwei). Im ersten Fall gibt es keine vernünftige nichtmonotone Erweiterung von S; im letzteren Fall gibt es mehrere, und McDermott und Doyle sehen nur das als Folgerung von S an, was in allen Fixpunkten enthalten ist:

(5.2.4) $TH(S) := \bigcap \{T: T \text{ ist Fixpunkt von } S\}$.

(Beachte: $\bigcap\emptyset = \{A: A \text{ ist Formel von } L\} = T_\perp$, d.h. $\bigcap\emptyset$ ist die „inkonsistente Theorie".) Statt $A \in TH(S)$ schreibt man auch $S \mid\!\sim A$.[7] Im Gegensatz zu McDermott und Doyle, die die Axiomenmenge S selbst als „Theorie" bezeichnen, würde ich eine Sprechweise bevorzugen, nach der TH(S) die Theorie auf der Grundlage der Axiome S ist.

Die Modelltheorie von McDermott und Doyle ist so aufgebaut:

5.2.5. Definition Eine *Interpretation* V von Formeln über der Objektsprache L ist ein Paar $\langle X,U\rangle$, wobei X eine nichtleere Menge und U eine übliche Interpretationsfunktion über X ist, mit den üblichen Bestimmungen für V bzgl. atomarer Sätze, bzgl. $\neg$, $\rightarrow$ und $\forall$, aber ohne Bestimmung bzgl. M. Ein *monotones Modell einer Satzmenge S* ist eine Interpretation V mit $V(A)=1$ für alle $A \in S$. Ein *nichtmonotones Modell einer Axiomenmenge S* ist ein Paar $\langle T,V\rangle$, wobei T ein Fixpunkt von S und V ein monotones Modell von T ist.

---

[7] Der „Beweisbarkeits"- und der „Theorem"-Begriff der nichtmonotonen Logik ist ohne Rückgriff auf den Beweisbegriff definiert und leider völlig inkonstruktiv. Eine etwas anschaulichere äquivalente Definition von $S \mid\!\sim A$ findet man bei Davis (1980). Dazu geht man von einer Aufzählung $A_1, A_2, A_3, \ldots$ aller Formeln aus (die Menge der Formeln sei abzählbar). Für eine gegebene Formelmenge S definiert man induktiv:

$S_0 := S$,

$$S_{i+1} := \begin{cases} S_i \cup \{MA_i\} & \text{falls } S_i \text{ und } A_i \text{ konsistent,} \\ T_\perp & \text{falls es ein B mit } MB \in S_i \text{ und } S_i \vdash \neg B \text{gibt,} \\ S_i & \text{sonst.} \end{cases}$$

Dann sei $S_\infty := \bigcup S_i$. Wie S_∞ aussieht, ist i.a. von der Aufzählung der Formeln abhängig; dies entspricht den möglicherweise verschiedenen Fixpunkten bei McDermott und Doyle. Es gilt $S \mid\!\sim A$ genau dann, wenn $S_\infty \vdash A$ für *jede* Aufzählung $A_1, A_2, A_3, \ldots$.

Mit diesen Bestimmungen kann ein Adäquatheitssatz (Korrektheit und Vollständigkeit) für die nichtmonotone Logik bewiesen werden:

5.2.6. Theorem $S \mid\!\sim A$ gilt genau dann, wenn $V(A)=1$ für alle Modelle $\langle T,V\rangle$ von S.

Zur Illustration seien einige weitere Eigenschaften von McDermott und Doyles nichtmonotoner Logik aufgezählt:

- Das Deduktionstheorem in der Form „Wenn $A \mid\!\sim B$, dann $\mid\!\sim A \rightarrow B$." gilt nicht.
  Alle Axiomensysteme ohne Vorkommnis von M haben genau einen Fixpunkt.

- $S = \{MA \rightarrow A\}$ hat genau einen Fixpunkt.

- Wenn T_1 und T_2 Fixpunkte sind mit $T_1 \subseteq T_2$, dann ist $T_1 = T_2$.

- Der Kompaktheitssatz in der Form „Wenn $S \mid\!\sim A$, dann schon $S_0 \mid\!\sim A$ für eine endliche Teilmenge $S_0 \subseteq S$." gilt nicht.

- Es gibt konsistente Axiomenmengen mit inkonsistenten Teilmengen, z.B. ist $S = \{MA \rightarrow \neg A, \neg A\}$ konsistent, nicht aber $S' = \{MA \rightarrow \neg A\}$.

- Sei S konsistent und $S_s := \{A: A \in TH(S')$ für alle konsistenten Obermengen S' von $S\}$ die Menge der „sicheren" Formeln von S. Dann ist S_s die kleinste Menge R derart, daß die folgenden Bedingungen gelten: (i) $S \subseteq R$, (ii) $Th(R)=R$, (iii) Wenn $A \in R$, dann auch $MA \in R$. (Nicht jede beweisbare Formel ist sicher.)

- Für endliche Axiomenmengen S, die in der Sprache der nichtmonotonen Aussagenlogik (mit M) formuliert sind, gibt es ein effektives Beweisverfahren, d.h. hier ist der Beweisbarkeitsbegriff entscheidbar.

- $S = \{MA \rightarrow B, \neg B\}$ ist inkonsistent (was von McDermott und Doyle als Problem gesehen wird).

Trotz des Adäquatheitssatzes ist dieser erste Ansatz zur nichtmonotonen Logik aber letztlich völlig unbefriedigend. Denn die Semantik ist so schwach, daß in einem Modell V $V(MA)=1$ gelten kann, obwohl $\neg A$ „ableitbar ist" (McDermott und Doyle, S. 53). Entscheidende Probleme bestehen darin, daß Konsistenz keine distributive Eigenschaft ist, d.h. daß aus $M(A \wedge B)$ nicht $M(A)$ folgt, und daß die Theorie $S = \{MA, \neg A\}$ nicht inkonsistent ist.

5.3 Nichtmonotone Logik II: Nichtmonotone Modallogik

Bei McDermott und Doyle (1980) ist die Semantik so schwach, daß es keinerlei Verknüpfung der Wahrheitswerte von A und $\neg M\neg A$ gibt. Zum Zwecke größerer Übersichtlichkeit verwenden wir i.f. die Schreibweise LA („Notwendigerweise A") als Abkürzung für $\neg M\neg A$.

McDermott (1982) versucht diesen Mangel dadurch zu beheben, daß er die monotonen Elemente „Th" und „$\vdash$" umdeutet. Anstelle der Prädikatenlogik erster Stufe sollen nun die bekanntesten modallogischen Systeme als „monotone Grundlage" dienen und mehr Zusammenhang zwischen „modalisierte" und „nichtmodalisierte" Formeln bringen. Als Schlußregel tritt neben die Regeln des Modus ponens und der universellen Generalisierung die Notwendigkeitsregel

(L) $\quad \vdash A \Rightarrow \vdash LA$.

(L) erlaubt den Schluß von A auf „Notwendigerweise A". Die Regel ist in diesem Zusammenhang so zu verstehen: Wenn A aus irgendwelchen Axiomen S logisch ableitbar ist, dann ist A *relativ zu S* notwendig. (L) überträgt dies in die Objektsprache.

Die folgenden in der Modallogik besonders verbreiteten Axiomenschemata werden in McDermotts Untersuchung einbezogen:

(1) $\quad LA \rightarrow A$
(2) $\quad L(A \rightarrow B) \rightarrow (LA \rightarrow LB)$
(3) $\quad \forall x(LA) \rightarrow L(\forall xA)$
(4) $\quad LA \rightarrow LLA$
(5) $\quad MA \rightarrow LMA$

Durch (1)–(3) wird das System **T**, durch (1)–(4) das System **S4** und durch (1)–(5) das System **S5** bestimmt.[8]

Auf die Motivation von (1)–(5) soll hier nicht eingegangen werden. McDermott betont noch ausdrücklicher als McDermott und Doyle die Interpretation von M als „ist konsistent" (im beweistheoretischen Sinne) und von L als „ist ableitbar". [9]

[8] Der Einschluß der „Barcan-Formel" (3) in **T** und **S4** ist unüblich; (3) folgt quantorenlogisch aus der Konjunktion von (1), (2), (4) und (5) (s. Hughes und Cresswell 1968, S. 145). Ebenso wie die besprochenen Autoren selbst gehe ich i.f. auf keine der äußerst heiklen Probleme der modalen Quantorenlogik ein, d.h. ich klammere insbesondere (3) aus.

[9] Es ist allerdings sehr fraglich, ob (L) und (1)–(5) wirklich zu dieser Interpretation passen. Wenn L z.B. für die (monotone) Ableitbarkeit in der Peano-Arithmetik **PA**

Ein solcher Ansatz ist aber, so McDermott, unvollständig, wenn man über keine Regel der Art

(N′) Wenn $\not\vdash \neg A$, dann $\vdash MA$.

verfügt. Doch diese Regel sei so „nicht wohlgeformt" und außerdem „zirkulär".[10] Also wendet McDermott (1982) genau dieselbe Konstruktion wie McDermott und Doyle an, nur mit **T**-Th, **S4**-Th und **S5**-Th anstelle von (**PL1**-)Th.

Symbolisiere $\mid\sim$ jetzt noch einmal denjenigen nichtmonotonen Ableitbarkeitsbegriff, der auf der (monotonen) Ableitbarkeitsrelation der Prädikatenlogik erster Stufe aufbaut. Dann gilt das

5.3.1. Theorem Für alle geschlossenen Instanzen α von (1)–(5) und für beliebige Axiomenmengen S gilt $S \mid\sim \alpha$.

Dieses Ergebnis McDermotts darf aber nicht zur Annahme verleiten, (1)–(5) seien damit von vornherein überflüssig. Denn es ist eine Eigenschaft von nichtmonotonen Theorien, daß aus $S\mid\sim A$ nicht automatisch $TH(S\cup\{A\}) = TH(S)$ folgt. (Beispiel ohne (L) und (1)–(5): $S = \{MA\rightarrow\neg B, MB\rightarrow\neg A, \neg A\rightarrow LC\}$ hat $\neg B$ als Theorem; erweitert man aber S um das S-Theorem $LC\rightarrow C$, so fällt $\neg B$ aus $TH(S\cup\{LC\rightarrow C\})$ wieder heraus.) Immerhin gilt, wie man sich leicht überlegt, die Cut-Regel: aus $S\mid\sim A$ folgt $TH(S\cup\{A\}) \subseteq TH(S)$ („Antiklimax" bei McDermott 1982, S. 38).

Die Modelltheorie von modalen Theorien ist komplizierter als die von nichtmodalen Theorien. Zuerst der monotone Fall.

5.3.2. Definitionen Eine *modale Interpretation V* von Formeln über einer modalen Sprache L ist ein Quadrupel $\langle W,R,X,U\rangle$, wobei $W\neq\emptyset$ eine Menge möglicher Welten, $X\neq\emptyset$ eine Individuenmenge, $R \subseteq W\times W$ die sog. „Zugänglichkeitsrelation" zwischen möglichen Welten und $U: L\times W \rightarrow$

---

stehen soll, dann ist entweder die Schlußregel (L) oder das Axiomenschema (1) fehl am Platze. Einerlei, ob man die Logik der **PA**-beweisbaren oder die Logik der **PA**-beweisbarerweise **PA**-beweisbaren Formeln haben will, in beiden Fällen gilt der Satz von Löb, welcher durch

(SL) $L(LA\rightarrow A)\rightarrow LA$

zu formalisieren ist. Und es ist leicht zu zeigen, daß (L) und (A4) und (SL) zusammen zum Widerspruch führen. Genaueres dazu findet man in Boolos (1979) und Smorynski (1984).

[10] Der eigentliche Haken an „Schlußregeln" wie (N) und (N′) scheint mir zu sein, daß die Nichttheoreme der Prädikatenlogik nicht rekursiv aufzählbar sind. Während man Regeln der Form $S\vdash A$ in der Beweispraxis als „Wenn S *bewiesen* ist, dann kann man auf B schließen" lesen kann, ist die Lesart von (N′) als „Wenn $\neg A$ unbewiesen ist, dann kann man auf MA schließen" offenkundig inadäquat — hier ist die Un*beweisbar*keit von $\neg A$ gefragt. Es sind vielleicht eher praktische als theoretische Gründe, die (N′) in dieser Form nutzlos machen.

{0,1} eine Bewertungsfunktion mit den üblichen Bestimmungen für $\neg$, $\rightarrow$ und $\forall$ ist (bei festem w$\in$W) und der folgenden Bestimmung für M: U(MA,w) = 1 genau dann, wenn U(A,w') = 1 für ein w' mit wRw'.[11] Bei gegebenem V heißt A *wahr (falsch) in w*, wenn V(A,w)=1 (=0), und A heißt *wahr (falsch) in V*, wenn V(A,w)=1 für alle w$\in$W (bzw. wenn V(A,w)=0 für ein w$\in$W) ist. V heißt **T**-*Interpretation*, wenn R reflexiv ist, **S4**-*Interpretation*, wenn R reflexiv und transitiv ist, und **S5**-*Interpretation*, wenn R reflexiv, transitiv und symmetrisch ist. Eine Interpretation V mit V(A)=1 für alle A$\in$S heißt *(modales) Modell von S*.

Für (monotone) modallogische Theorien gilt bekanntlich der Adäquatheitssatz:

5.3.3. Theorem Für jede Formelmenge S und jede Formel A gilt: S $\vdash_{\mathbf{T(S4,S5)}}$ A genau dann, wenn A in allen **T**-(**S4**-,**S5**-) Modellen von S wahr ist.

Der nichtmonotone Fall wird deshalb so kompliziert, weil er die *semantische Lokalitätsbedingung* („Wenn $\mathcal{V}_1$ die Menge der S_1-Modelle und $\mathcal{V}_2$ die Menge der S_2-Modelle ist, dann ist $\mathcal{V}_1 \cap \mathcal{V}_2$ die Menge der Modelle von $S_1 \cup S_2$.") verletzt. Beispiel: Sei $S_1 = \{MA \rightarrow A\}$ und $S_2 = \{\neg A\}$. Dann ist in allen S_1-Modellen A, in allen S_2-Modellen aber $\neg$A wahr; trotzdem ist $S_1 \cup S_2$ konsistent in dem Sinn, daß $S_1 \cup S_2$ Modelle hat (in denen $\neg$A wahr ist).

Für die folgenden Definitionen, die nicht leicht zu durchschauen sind, benutzen wir die Abkürzung V(A)=1 anstelle von V(A,w)=1 für alle w$\in$W. Die Präfixe **T**-, **S4**-, **S5**- denke man sich immer mitgeschleppt.

5.3.4. Definitionen Die *Zufälligkeiten von V bzgl. einer Axiomenmenge S* sind definiert durch Z(V,S) := {MA: V(MA)=1 und es gibt ein Modell V' von S mit V'(MA)$\neq$1}. Ein *unverbindliches Modell V einer Axiomenmenge S* ist ein modales Modell von S mit der Eigenschaft V(MA)=1 für alle A derart, daß es ein Modell V' von S$\cup$Z(V,S) und eine Welt w' in W' mit V'(A,w')=1 gibt.[12]

In unverbindlichen Modellen sind ungerechtfertigte Notwendigkeiten (Sätze der Form $\neg$MA) ausgeschlossen. Wenn die intuitive Grundlage vielleicht auch nicht ganz klar ist, so gibt es doch einen *Satz*, der zeigt, worauf

[11] Wenn W, R und X fest sind, setzen wir i.f. V mit U gleich.

[12] S. McDermott (1982, S. 41). Die komplizierte Definition von Z(V,S) braucht man eigentlich gar nicht. Denn V' ist genau dann ein Modell von S $\cup$ Z(V,S) = S $\cup$ {MA: V(MA)=1 und es gibt ein Modell V'' von S mit V''(MA)$\neq$1}, wenn V' ein Modell von S $\cup$ {MA: V(MA)=1} ist. Mit der Abkürzung M(V):={MA: V(MA)=1} ist ein Modell V von S also genau dann unverbindlich, wenn für alle Sätze A gilt: Wenn ein Modell V' von S$\cup$M(V) existiert mit V'($\neg$A)$\neq$1, dann ist V(MA)=1.

die Definitionen 5.3.4 zugeschnitten sind:

5.3.5. Theorem V ist genau dann ein **T**-(**S4**-,**S5**-)Modell eines Fixpunktes von S, wenn V ein unverbindliches **T**-(**S4**-,**S5**-)Modell von S ist.

Im Beweis der Richtung v.r.n.l. zeigt McDermott, daß Th(S∪Z(V,S)) bei unverbindlichen Modellen V von S ein Fixpunkt von S ist.

5.3.6. Korollar A ist genau dann in allen unverbindlichen **T**-(**S4**-,**S5**-)Modellen von S wahr, wenn $S \mid\!\sim_{\mathbf{T(S4,S5)}} A$.

McDermott weist auf eine Schwierigkeit der intendierten Interpretation von MA als „A ist wahr in einer möglichen Welt, die mit dem konsistent ist, was ‚der Roboter' glaubt" hin, die entsteht, wenn der Roboter nur Theoreme glauben darf. Die Theorie $S = \{MA \rightarrow \neg B, MB \rightarrow \neg A\}$ etwa hat weder $\neg A$ noch $\neg B$ als Theorem, d.h. sowohl A als auch B ist mit S konsistent; aber nur eine der beiden Möglichkeiten MA und MB, so McDermott, ist „wahr in der wirklichen Welt". Dies ist für McDermott ein Argument dafür, daß Roboter vielleicht besser zunächst nur *einen* Fixpunkt suchen und erst bei Bedarf (d.h. wenn Widersprüche auftreten) Änderungen vornehmen sollten (vgl. auch Doyles (1979) TMS und Reiters Default Logic). Jedenfalls scheint es keine ganz befriedigende intuitive Interpretation von MA nach McDermotts Vorstellung zu geben.

Der nächste Satz zeigt, daß im Falle von **S5** unverbindliche Modelle ausreichen:

5.3.7. Theorem Wenn es ein **S5**-Modell der Axiomenmenge S gibt, so gibt es auch ein unverbindliches **S5**-Modell.

Einerseits ist das schön, denn Theorem 5.3.7 zeigt, daß das nichtmonotone Schließen hier also nicht zu Widersprüchen führt; andererseits hat Theorem 5.3.7 eine unerwünschte Konsequenz:

5.3.8. Korollar Wenn $S \mid\!\sim_{\mathbf{S5}} A$, dann schon $S \vdash_{\mathbf{S5}} A$.

Das bedeutet, daß der ganze nichtmonotone Aufsatz auf **S5** nichts zu dem altbekannten monotonen System **S5** dazutut. Er erlaubt keine neuen Schlüsse aus der Unwissenheit. Wie schwach $\mid\!\sim_{S5}$ ist, sieht man am Beispiel $S = \{MA \rightarrow A\}$; es gilt *nicht* $S \mid\!\sim_{\mathbf{S5}} A$, wie man wünschen würde (denn es gibt einen Fixpunkt von S mit $M\neg A$, und $MA \rightarrow A, M\neg A \vdash_{\mathbf{S5}} \neg A$; außerdem gilt ja $S \not\mid\!\sim_{\mathbf{S5}} A$). Also kann man den Inferenzapparat von **S5** nicht sinnvollerweise in seiner ganzen Stärke für die nichtmonotone Logik verwenden. In den nichtmonotonen Versionen von **T** und **S4** tritt das Problem der Reduktion auf den monotonen Fall zwar nicht auf, dafür kann die Konsistenz dieser Systeme nur für den aussagenlogischen Teil bewiesen werden.

Der aussagenlogische Teil der nichtmonotonen **T** und **S4** (und natürlich **S5**) ist entscheidbar. McDermott gibt ein Beweisverfahren für endliche aus-

sagenlogische **S4**-basierte Theorien explizit an. Als Nebenergebnis davon ergibt sich, daß das nichtmonotone **S4** (ohne nichtlogische Axiome) konsistent ist.

5.4 Autoepistemische Logik

Moore (1985) knüpft an die Selbstkritik von McDermott (und Doyle) an und präsentiert einen konstruktiven Alternativvorschlag. Nach seiner Auffassung ist die nichtmonotone Logik eigentlich nicht so sehr zur Modellierung von „Default Reasoning" mit Prämissen wie Die meisten Vögel fliegen oder Ein typischer Vogel fliegt geeignet, sondern besser zu verstehen, wenn man L ganz konsequent als Es wird geglaubt, daß und M als Es ist konsistent zu glauben, daß liest. Unabhängig davon, ob und wie sich technische Unterschiede ergeben mögen, gibt es unterschiedliche Gründe für die Nichtmonotonie von Default und „Autoepistemic" Reasoning. Moore (S. 79f) schreibt:

> Default reasoning is nonmonotonic because, to use a term from philosophy, it is *defeasible*: its conclusions are tentative, so, given better information, they may be withdrawn. Purely autoepistemic reasoning, however, is not defeasible. ... As Stalnaker (1980) has observed, autoepistemic reasoning is nonmonotonic because the meaning of an autoepistemic statement is context-sensitive; it depends on the theory in which the statement is embedded. ... default reasoning is nonmonotonic because it is defeasible, autoepistemic reasoning is nonmonotonic because it is indexical.

Moore nennt die Menge der Formeln, die sich aus einer Axiomenmenge herleiten lassen eine *autoepistemische Theorie* (*AE-Theorie* oder kurz *Theorie*). Sie ist aufzufassen als die Menge *aller* Überzeugungen einer Person oder Maschine, die über ihre eigenen Überzeugungen reflektieren kann. Moore beschränkt sich auf Theorien auf aussagenlogischem Niveau.

Die einfache Modelltheorie von AE-Theorien sei explizit entwickelt.

5.4.1. Definitionen Eine *Interpretation einer Theorie T* ist eine Zuordnung von Wahrheitwerten zu den Formeln der Sprache von T mit den üblichen wahrheitsfunktionalen Bestimmungen der klassischen Aussagenlogik (ohne Bestimmungen für L). Ein *Modell von T* ist eine Interpretation von T, die alle Formeln in T wahr macht. Eine *AE-Interpretation einer*

Theorie T ist eine Interpretation von T, in der für jede Formel A der Sprache von T LA genau dann wahr ist, wenn A∈T. Ein *AE-Modell von T* ist eine AE-Interpretation von T, die alle Formeln in T wahr macht.

5.4.2. Definitionen Eine Theorie T ist genau dann *korrekt bzgl. einer Axiomenmenge S*, wenn jede AE-Interpretation von T, in der alle Formeln in S wahr sind, ein AE-Modell von T ist. Eine Theorie T ist genau dann *semantisch vollständig*, wenn T jede Formel enthält, die in jedem AE-Modell von T wahr ist.

Die syntaktische Charakterisierung von Theorien, die bzgl. einer Axiomenmenge S korrekt und semantisch vollständig sein sollen, geschieht nach einem Vorschlag von Stalnaker (1980) durch die Angabe von Abgeschlossenheitseigenschaften:

5.4.3. Definition Eine Theorie T heißt genau dann *stabil*, wenn gilt:
(a) Wenn $A_1,\ldots,A_n \in T$ und $A_1,\ldots,A_n \vdash B$, dann $B \in T$.[13]
(b) Wenn $A \in T$, dann $LA \in T$.
(c) Wenn $A \notin T$, dann $\neg LA \in T$.

Die folgenden Sätze zeigen, daß Wahrheit und Vollständigkeit von stabilen Theorien allein von ihren *objektiven* Formeln, d.h. den Formeln ohne Vorkommnis von L und M, abhängen:

5.4.4. Theorem Sei T eine stabile Theorie. Dann ist jede AE-Interpretation von T, die ein Modell der objektiven Formeln von T ist, ein AE-Modell von T.

5.4.5. Theorem Enthalten zwei stabile Theorien dieselben objektiven Formeln, so sind sie identisch.

5.4.6. Theorem Eine Theorie ist genau dann semantisch vollständig, wenn sie stabil ist.

Für stabile Theorien auf einer Axiomenbasis S gilt offenbar $Th(S \cup \{LA{:}A \in T\} \cup \{\neg LA{:}A \notin T\}) \subseteq T$.[14] Was jetzt noch fehlt, ist eine Garantie dafür, daß aus S nicht mehr geschlossen wird als intuitiv erlaubt. Moores Vorschlag drängt sich geradezu auf:

5.4.7. Definition Eine Theorie T heißt *fundiert in einer Axiomenmenge S* genau dann, wenn $T \subseteq Th(S \cup \{LA{:}A \in T\} \cup \{\neg LA{:}A \notin T\})$.

Das nächste Theorem zeigt, daß man mit Definition 5.4.7 das bekommt, was Moores Semantik will:

5.4.8. Theorem Eine Theorie T ist genau dann korrekt bzgl. einer Prämissenmenge S, wenn T fundiert in S ist.

[13] $\vdash$ ist hier die Ableitbarkeitsrelation der gewöhnlichen Aussagenlogik.
[14] „Th" soll hier die aussagenlogischen Konsequenzen liefern.

Als vernünftige Theorien bei einer gegebenen Axiomenmenge S werden also die stabilen, in S fundierten Erweiterungen von S ausgezeichnet, die Moore *stabile Erweiterungen von S* nennt. Als Menge der *Theoreme von S* bezeichnet er die Menge der Formeln in *einer* stabiler Erweiterung von S (bzw. $\emptyset$, wenn keine solche existiert). Diese „von-innen-Perspektive" ist nach Moore der autoepistemischen Auffassung angemessener als die „von-außen-Perspektive" des Durchschnitts aller stabilen Erweiterungen. Mir erscheint aber nicht jede stabile Erweiterung von S vernünftig. So kann man nachprüfen, daß die inkonsistente Theorie $T_\perp$ eine stabile Erweiterung von $S = \{LA \rightarrow A, \neg A\}$ ist, doch es ist sicher wünschenswert, nur konsistente stabile Erweiterungen in Betracht zu ziehen.[15] Durch diese Beobachtung möchte ich erste Zweifel daran aufwerfen, ob die Fundiertheitsbedingung in Definition 5.4.7 wirklich das leistet, wozu sie gedacht ist. Ich komme in Abschnitt 5.7 auf dieses Thema zurück.

Moores Konstruktion hat offensichlich viele Gemeinsamkeiten mit der von McDermott und Doyle. Wie bei dieser haben in der autoepistemischen Logik die Axiomenmengen $S = \{MA \rightarrow \neg A\}$ und $S = \{MA \rightarrow B, \neg B\}$ keine stabile Erweiterung, $S = \{MA \rightarrow \neg B, MB \rightarrow \neg A\}$ deren zwei (das letzte Beispiel wird in Abschnitt 5.7 ausführlich besprochen). Während bei McDermott und Doyle aber ein Fixpunkt T nur „allwissend" bzgl. seiner eigenen (der T-)Möglichkeiten, d.h. bzgl. seines *Nichtwissens*, sein muß, müssen Moores stabile Erweiterungen auch über ihr eigenes *Wissen* Buch führen.[16] Mit der Abkürzung

$$L(R) := \{LA: A \in R\}$$

wäre der entscheidende Operator also

(5.4.9) $\quad NM_S(R) := Th(S \cup M(R) \cup L(R))$.

Fixpunkte des so definierten Operators NM_S sind stabile Erweiterungen von S.[17] Damit verschwinden die zentralen Probleme von McDermott und

[15] Wie dies auch Halpern und Moses (1984) tun, deren Arbeit hiermit als Ergänzung zu Moore (1985) empfohlen sei. $T_\perp$ ist übrigens die einzige Theorie, die eine stabile Obermenge einer anderen stabilen Theorie sein kann. Denn seien T_1 und T_2 zwei verschiedene stabile Theorien mit $T_1 \subsetneq T_2$; dann ist für $A \in T_2 \setminus T_1$ wegen Definition 5.4.3(c) $\neg LA \in T_1 \subsetneq T_2$ und wegen Definition 5.4.3(b) $LA \in T_2$, also gilt wegen Definition 5.4.3(a) $T_2 = T_\perp$.

[16] Die Entwicklung in der nichtmonotonen Logik vollzieht sich also gewissermaßen gegenläufig zu der in der Konditionallogik. Während hier der Ramsey-Test das Wissen (in T^*_A) prüft und erst negierte oder möglicherweise-Konditionalsätze auf das Nichtwissen (in T^*_A) zugreifen, wurde dort zuerst das Nichtwissen und dann das Wissen (in T) in Betracht gezogen.

[17] Hier geht ein, daß $\neg\neg A$ klassisch äquivalent mit A ist.

Doyles Logik: Nun folgt M(A) aus M(A∧B), und S = {MA, ¬A} ist jetzt inkonsistent.

Durch die Hinzunahme der modalen Schlußregel (L) hat McDermotts nichtmonotone modale Logik noch mehr Ähnlichkeit mit dem NM$_S$ von Moore, in welches L(R) eingeht. Die Menge der modalen Fixpunkte von S im Sinne von McDermott ist aber eine *echte* Teilmenge der stabilen Erweiterungen von im Sinne von Moore; so gibt es z.B. eine stabile Erweiterung von S = {LA→A}, welche A enthält, aber keinen modalen Fixpunkt von S mit dieser Eigenschaft. Moore (1985, S. 87) meint, McDermott expliziere „gerechtfertigt-glauben", während er selbst einfach „glauben" expliziere.

Nachdem (L) Moores Billigung gefunden hat, erhebt sich die Frage, wie es mit den modallogischen Axiomenschemata (1)–(5) aus autoepistemischer Sicht steht (das quantorenlogische Schema (3) bleibt unberücksichtigt). Moore hält (2), (4) und (5), die die Klauseln (a), (b) bzw. (c) aus Definition 5.4.3 „beschreiben", für unproblematisch, das Schema (1) — gelesen als Wenn die Person/Maschine A glaubt, dann ist A wahr — jedoch für unangebracht. Zwar ist LA→A in jeder stabilen Theorie enthalten und damit ein nichtmonotones Theorem jeder Prämissenmenge S, aber *als Prämisse* selbst erlaube sie, zusammen mit (5), ungerechtfertigte Selbstbegründungen von beliebigen Sätzen A, woraus sich auch der Kollaps des nichtmonotonen **S5** in das monotone **S5** ergebe.[18]

Moore empfiehlt nicht — wie McDermott — den Rückzug von **S5** auf **S4** oder gar **T**, sondern die Streichung von (1) aus **S5**, wonach der aussagenlogische Teil des *schwachen* **S5** (Stalnaker 1980) oder **K45** (Chellas 1980) übrig bleibt. Daß er (2), (4) und (5) als Axiome aber gar nicht braucht, folgt aus der Tatsache, daß alle Instanziierungen dieser Axiomenschemata in jeder AE-Interpretation einer jeden stabilen Theorie wahr sind, und dem folgenden Resultat:

[18] Ich habe den Verdacht, daß in Moores Argumentation hier drei Fehler stecken. Erstens suggeriert die o.g. Lesweise von (1) eine „von-außen-Perspektive", wo eigentlich eine „von-innen-Perspektive" angebracht wäre — warum kann sonst LA→A ein universelles Theorem sein? (Vgl. hierzu auch Levesque 1990). Zweitens ergibt, bei Zugrundelegung des Theorembegriffs von McDermott (Durchschnitt aller Fixpunkte), die Hinzunahme eines Theorems als Axiom höchstens weniger, nie aber mehr Theoreme — wie können die ungerechtfertigten Selbstbegründungen dann zustande kommen? Drittens soll das aus (1) und (5) folgende MLA→A die Selbstbegründungen liefern — aber zum einen kann es nach Moores eigener Definition keine stabile Theorie T geben mit MLA∈T, ohne daß A∈T (denn sei A∉T, dann ¬LA∈T, dann L¬LA∈T, also, sofern T konsistent, ¬L¬LA∉T, d.h. MLA∉T; wenn T inkonsistent, so ist A trivialerweise in T), und zum anderen muß MLA ja auch irgendwie gerechtfertigt sein.

5.4.10. Theorem Wenn A in jeder AE-Interpretation von T wahr ist, dann ist T genau dann eine stabile Erweiterung von SU{A}, wenn T eine stabile Erweiterung von S ist.

Moore resümiert, daß McDermott in die falsche Richtung ging, indem er die nichtmonotone Logik als eine Logik der Beweisbarkeit statt als Logik der Überzeugungen betrachtete.[19] Während McDermott (1982) das Fehlen eines jeglichen Zusammenhangs zwischen den Wahrheitswerten von A und LA in McDermott und Doyle (1980) monierte, betont Moore, daß in der Logik der Überzeugungen gar kein Grund für einen Zusammenhang dieser Wahrheitswerte zu sehen ist. Das wirkliche Problem der ursprünglichen nichtmonotonen Logik sei es gewesen, daß „die ‚wenn'-Hälfte der semantischen Definition von L — daß LA genau dann wahr ist, wenn A geglaubt wird — nicht in der Logik ausgedrückt wurde" (Moore 1985, S. 90).

5.5 Mögliche-Welten-Semantik für autoepistemische Logik

Die autoepistemische Logik und ihre Semantik sind inkonstruktiv, so daß die Existenz von stabilen Erweiterungen sehr schwer zu beweisen ist. In Moore (1984) wird eine alternative Semantik entwickelt, die es erlaubt, endliche Modelle für AE-Theorien zu konstruieren und die Existenz von korrekten und vollständigen Theorien auf der Basis einer Axiomenmenge S zu beweisen.

In der in Moore (1985) präsentierten Semantik gibt es keinen systematischen Zusammenhang zwischen der Wahrheit etwa von LA und L(A∧B). Wenn man aber die Überzeugungen ideal rationaler Personen oder Maschinen modellieren will, dann sollte LA aus L(A∧B) folgen. Solchermaßen idealisierte stabile AE-Theorien können durch spezielle Mögliche-Welten-Modelle für modale Logiken à la Kripke dargestellt werden.

5.5.1. Definitionen Eine *vollständige* **S5**-*Struktur* ist ein geordnetes Paar $\mathbf{K}=\langle W,U\rangle$, wobei $W\neq\emptyset$ eine Menge möglicher Welten und $U: L\times W \rightarrow \{0,1\}$ eine Bewertungsfunktion mit den üblichen Bestimmungen für ¬ und → und der folgenden Bestimmung für L ist:
$U(LA,w) = 1$ genau dann, wenn $U(A,w') = 1$ für alle $w'\in W$.[20]

[19] Ironischerweise beanstandet Moore genau das Axiomenschema, welches in der Logik der **PA**-beweisbaren **PA**-Beweisbarkeit beanstandet wird (vgl. Fußnote 9). Die Begründungen für den Wegfall von (1) sind aber von Grund auf verschieden.

[20] Vgl. damit den Begriff der modalen Interpretation nach Definition 5.3.2. Die Indi-

Eine Formel A heißt genau dann *wahr in einer (vollständigen* **S5**-*) Struktur* $\mathbf{K}=\langle W,U\rangle$, wenn $U(A,w) = 1$ für alle $w\in W$.

Die folgende Beobachtung wurde unabhängig von Moore, von Halpern und Moses, von Fitting und von van Benthem gemacht:

5.5.2. Theorem Eine Formelmenge T ist genau dann eine stabile AE-Theorie, wenn es eine vollständige **S5**-Struktur **K** gibt, so daß T die Menge der in **K** wahren Sätze ist.

Mit diesem Satz im Rücken kann man jede AE-Interpretation einer beliebigen stabilen AE-Theorie T durch eine sog. Mögliche-Welten-Interpretation charakterisieren:

5.5.3. Definitionen Eine *Mögliche-Welten-* oder kurz *MW-Interpretation* von T ist ein geordnetes Paar $\langle \mathbf{K},V\rangle$, wobei $\mathbf{K}=\langle W,U\rangle$ eine vollständige **S5**-Struktur, T die Menge der in **K** wahren Sätze und V eine Bewertung der nichtmodalen („objektiven") Formeln der Sprache von T ist. Eine Formel A ist *wahr in* $\langle \mathbf{K},V\rangle$ genau dann, wenn

(a) A eine Satzkonstante ist und $V(A)=1$ oder
(b) A eine Formel LB ist und $U(B,w)=1$ für alle $w\in W$ oder
(c) A ansonsten nach wahrheitsfunktionaler Rekursion wahr ist.

$\langle \mathbf{K},V\rangle$ heißt genau dann *Mögliche-Welten-* oder kurz *MW-Modell* von T, wenn jede Formel von T wahr in $\langle \mathbf{K},V\rangle$ ist.

In MW-Interpretationen ist W als die Menge der mit den Überzeugungen der Person oder Maschine verträglichen Welten aufzufassen, und V repräsentiert die Wahrheiten der wirklichen Welt. Theorem 5.5.2 zeigt, daß es für jede AE-Interpretation (jedes AE-Modell) einer stabilen Theorie T eine entsprechende MW-Interpretation (ein entsprechendes MW-Modell) von T gibt und umgekehrt.

Schließlich zeigt Moore noch:

5.5.4. Theorem Eine MW-Interpretation $\langle \mathbf{K},V\rangle$ von T, $\mathbf{K}=\langle W,U\rangle$, ist ein MW-Modell von T genau dann, wenn es ein $w\in W$ gibt mit $V(A)=U(A,w)$, für alle Satzkonstanten A.

Als Anwendungsbeispiel der MW-Semantik für die autoepistemische Logik *beweist* Moore, daß $S = \{\neg LA\rightarrow B, \neg LB\rightarrow A\}$ zwei stabile Erweiterungen hat, eine mit A und ohne B und eine mit B und ohne A. In Moore (1985) hatte er für dieses Ergebnis nur Plausibilitätsüberlegungen

viduenmenge X wird nicht gebraucht, da sich Moore nur auf aussagenlogischem Niveau bewegt, und die Zugänglichkeitsrelation ist hier die triviale $R=W\times W$. (Hinsichtlich der modallogischen Gültigkeit ist es übrigens egal, ob R eine Äquivalenzrelation oder die triviale Relation ist; vgl. Hughes und Cresswell 1968, S. 74).

anführen können.[21] Zwecks Vereinfachung der Notation können einfache MW-Interpretationen $\langle \mathbf{K},V\rangle$ als Paare $\langle W,v\rangle$ angegeben, wobei W an der ersten Stelle des Paars $\mathbf{K}=\langle W,U\rangle$ steht und v die durch V „erzeugte" mögliche Welt ist. Mögliche Welten wiederum sollen durch die in ihnen wahren Literale (Satzkonstanten bzw. Negationen von Satzkonstanten) charakterisiert werden (wodurch man U einsparen kann).

Für das Anwendungsbeispiel betrachten wir zunächst die durch $W = \{\{A,B\},\{A,\neg B\}\}$ charakterisierte stabile Theorie T. Die einzigen MW-Interpretationen von T, die Modelle von $S = \{\neg LA\rightarrow B,\ \neg LB\rightarrow A\}$ sind, sind offenbar

$$\langle W,v_1\rangle = \langle\{\{A,B\},\{A,\neg B\}\},\ \{A,B\}\rangle \quad \text{und}$$
$$\langle W,v_2\rangle = \langle\{\{A,B\},\{A,\neg B\}\},\ \{A,\neg B\}\rangle\ .$$

Es ist $v_1\in W$ und $v_2\in W$, also hat man — nach Theorem 5.5.4 — beide Male MW-Modelle von T. Also ist T korrekt bzgl. S; da T nach Theorem 5.5.2 auch stabil ist und beide Sätze von S wahr in W sind, d.h. $S\subseteq T$, ist T eine stabile Erweiterung von S. T enthält A, aber nicht B.

Ganz analog erhält man eine stabile Erweiterung T′ von S, die B, aber nicht A enthält. Hat man aber eine Theorie T″, die A *und* B enthält, so ist das entsprechende $W'' = \{\{A,B\}\}$. Dann ist aber $\langle W'',v\rangle = \{\{A,B\},\{\neg A,\neg B\}\}$ eine MW-Interpretation von T″, unter der beide Sätze aus S wahr, aber einige Sätze aus T″ falsch (z.B. A und B) sind. Deshalb ist T″ nicht korrekt bzgl. S, also ist T″ keine stabile Erweiterung von S, also gibt es keine stabile Erweiterung von S, die A und B enthält.

5.6 Kritische Beispiele

Moores autoepistemische Logik ist eine natürliche und ausgereifte Form der nichtmonotonen Logik. Wenn ihre Überlegenheit gegenüber McDermotts System auch nicht völlig einsichtig ist (s. Fußnote 14), so hat sie auf jeden Fall den Vorzug der besseren Handhabbarkeit. Ich werde nun die im vorigen Abschnitt zur Verfügung gestellte Methode verwenden, um zwei Beispiele zu analysieren, die darauf hinweisen, daß die autoepistemische Logik *zumindest* in der Anwendung noch erhebliche Probleme macht. In diesen Anwendungen habe ich bestimmten Ausdrücken, die „M" enthalten, die Verbalisierung normalerweise zugeordnet, also eine Default-Reasoning- und keine strenge autoepistemische Interpretation zugrunde gelegt, da in

[21] Moore (1985) ist vor Moore (1984) entstanden und wurde schon 1983 auf der 8. IJCAI in Karlsruhe präsentiert.

der Praxis fast ausschließlich die erstere Interpretation auftaucht.

5.6.1 Die autoepistemische Logik liefert zu wenig Schlußfolgerungen

Das erste Beispiel stammt von Touretzky (1986, S. 16–18), dem es als Kritik an McDermott und Doyle (1980) diente. Wir betrachten die ornithologische Minitheorie mit den Axiomen

> Vögel können normalerweise fliegen.
> Ein jeder Strauß ist ein Vogel.
> Strauße können normalerweise nicht fliegen.

Der Grund für die Formulierung des letzten Axioms mit „normalerweise" ist, daß wir nicht ausschließen wollen, daß irgendwann besondere, flugfähige Strauße entdeckt oder gezüchtet werden oder schon worden sind.

Um unsere Theorie auf ein rein aussagenlogisches Niveau zu bringen, wollen wir sie als eine Theorie über Henry formulieren. Die Satzbuchstaben sollen folgende Bedeutung haben:

O: Henry ist ein Strauß.
V: Henry ist ein Vogel.
F: Henry kann fliegen.

Unter Verwendung naheliegender Formalisierungsregeln hat unsere Theorie die Axiomenbasis

$$S = \{V \wedge MF \rightarrow F\ ,\ O \rightarrow V\ ,\ O \wedge M\neg F \rightarrow \neg F\}\ .$$

In unserem alltäglichen Denken würden wir, sollten wir erfahren, daß Henry tatsächlich ein Strauß ist, nichtmonoton folgern, daß Henry *nicht* fliegen kann. Was sagt Moores System dazu?

Wir nehmen an, daß Henry ein Strauß ist, betrachten also die neue Axiomenmenge $S \cup \{O\}$. Zuerst machen wir die Angelegenheit durch Benennen und Umformulieren der Axiome etwas übersichtlicher:

Axiom 1	$\neg V \vee F \vee L\neg F$
Axiom 2	$\neg O \vee V$
Axiom 3	$\neg O \vee \neg F \vee LF$
Axiom 4	O

Axiom 2 und Axiom 4 legen fest, daß auf jeden Fall $O \wedge V$ gilt. Deshalb kommen als Kandidaten für stabile Erweiterungen nur die folgenden drei Theorien in Frage; wir repräsentieren sie, wie am Ende des vorangegangenen Abschnitts, durch die Mengen der mit der jeweiligen Theorie verträglichen Welten.

T_1 $W_1 = \{\{O,V,\neg F\}\}$
T_2 $W_2 = \{\{O,V.F\}\}$
T_3 $W_3 = \{\{O,V,\neg F\},\{O,V,F\}\}$

In T_1 ist LF falsch und L$\neg$F wahr. Deshalb ist Axiom 1 immer erfüllt und Axiom 3 „erzwingt" $\neg$F. Also ist die einzige „wirkliche Welt" v_1, die $\langle W_1,v_1\rangle$ zu einem Modell von S macht, $v_1 = \{O,V,\neg F\}$. Es gilt $v_1 \in W_1$, d.h. W_1 ist korrekt bzgl. S, also ist T_1 eine stabile Erweiterung von S. In T_2 ist LF wahr und L$\neg$F falsch. Deshalb ist Axiom 3 immer erfüllt und Axiom 1 erzwingt F. Also ist die einzige wirkliche Welt v_2, die $\langle W_2,v_2\rangle$ zu einem Modell von S macht, $v_2 = \{O,V,F\}$. Es gilt $v_2 \in W_2$, d.h. W_2 ist korrekt bzgl. S, also ist T_2 eine stabile Erweiterung von S.

Damit hat unsere Analyse bestätigt, daß sich Touretzkys Feststellungen über McDermott und Doyles Logik auf Moore übertragen lassen: Es gibt zwei stabile Erweiterungen von S, die uns beide zu einer Behauptung über die Flugfähigkeit von Henry veranlassen, jedoch zu einander widersprechenden Behauptungen. Intuitiv würde man vielleicht gerne auf die weniger verpflichtende Theorie T_3, die sozusagen T_1-oder-T_2 darstellt, ausweichen wollen. Aber T_3 ist überhaupt keine Erweiterung von S: Weder Axiom 1 noch Axiom 3 sind in T_3 enthalten.

Moores Theorie läßt uns also im Ungewissen darüber, ob Henry jemals von Boden wegkommen wird, und zwar unabhängig davon, ob man McDermott und Doyles oder Moores Theorembegriff (s.o.) zugrundelegt. Es gibt keinen Mechanismus, der uns hilft zu erkennen, daß die spezifischere Strauß-Regel die allgemeinere Vogel-Regel im Falle des Straußen Henry außer Kraft setzen sollte. Die gewünschte Konklusion ist nicht herausgekommen.

Es gibt einen Ausweg für die nichtmonotone Logik, den Touretzky beschreibt: das explizites Auflisten aller Ausnahmen der Vogel-Regel. In unserem Fall ersetzt man im ersten Axiom M(F) durch M(F$\wedge\neg$O), d.h. aus Axiom 1 wird

Axiom 1′ $\quad \neg V \vee F \vee L(Ov\neg F)$.

Um zu sehen, wie sich dies auswirkt, betrachten wir wieder T_1-T_3. An der Argumentation bzgl. T_1 ändert sich nichts, denn mit L($\neg$F) ist auch L(Ov$\neg$F) wahr, und T_1 bleibt eine stabile Erweiterung der neuen Axiome. T_3 ist immer noch keine Erweiterung der schwächer gewordenen Axiomenbasis: Axiom 3 ist nicht in T enthalten. Eine Veränderung ergibt sich jedoch in Bezug auf T_2: In T_2 sind sowohl L(F) als auch L(Ov$\neg$F) wahr, d.h. die Axiome 1′ und 3 sind automatisch erfüllt. Damit ist $v = \{O,V,\neg F\}$

eine wirkliche Welt derart, daß $\langle W_2,v\rangle$ ein Modell der neuen Axiome ist; da aber $v \notin W_2$, ist T_2 nicht mehr korrekt bzgl. S, d.h. T_2 fällt als stabile Erweiterung von S weg. Mit T_1 als einziger verbliebener Theorie haben wir das gewünschte nichtmonotone Theorem $\neg F$.

Diese Idee löst zwar das technische Problem, sie opfert aber den Witz der nichtmonotonen Logik, nämlich das Verarbeiten von impliziten normalerweise-Wendungen. Man muß hier eine ganze Liste von Ausnahmen in den Skopus von M mit einbeziehen: Strauße, Emus, Nandus, Kasuare, Kiwis, Pinguine, Riesenalke, chinesische Seidenhühner und vielleicht noch andere mehr. Im allgemeinen Fall wird das eine recht große Anzahl von Sonderklauseln ausmachen. Die natürlichsprachlichen Prämissen sind nach ihrer Formalisierung nicht mehr wiederzuerkennen. Darum kann dieser Weg nicht als Fortschritt der nichtmonotonen oder autoepistemischen Logik bezeichnet werden.

5.6.2 Die autoepistemische Logik liefert zu viele Schlußfolgerungen

Das zweite Beispiel stammt von Łukaszewicz (1986, S. 14f) und behandelt eine Miniaturtheorie meiner Freizeitgestaltung mit den Axiomen

> Normalerweise gehe ich am Freitag zum Angeln.
> Wenn ich krank bin, gehe ich normalerweise nicht zum Angeln.

Im zweiten Axiom kann die Einschränkung „normalerweise" z.B. der traurigen Tatsache Rechnung tragen, daß ich auf Wunsch meines Chefs eine weniger ernste Krankheit hintanstellen und mit ihm zum Angeln gehen muß.

Durch unser letztes Beispiel vorgewarnt, wollen wir diesmal gleich eine Ausnahmeregelung mit einbauen, die die Anwendbarkeit der Regeln im Konkurrenzfalle organisiert. Dabei sollen die Satzbuchstaben jetzt mit folgender Bedeutung versehen sein:

F: Es ist Freitag.
A: Ich gehe zum Angeln.
K: Ich bin krank.

Die Axiomenbasis gestalten wir — Lukaszewicz folgend — dann so, daß mir meine Gesundheit mehr wert ist als mein Freitagsvergnügen, daß also die Ausnahme des Krankheitsfalls in die Freitagsregel aufgenommen ist:

$$S = \{F \wedge M(\neg K \wedge A) \rightarrow A \ , \ K \wedge M\neg A \rightarrow \neg A\} \ .$$

Wir wollen diese kombinierte Theorie jetzt für sich alleine betrachten, also kein „Faktenwissen" über den Wochentag oder meinen Gesundheitszustand als zusätzliches Axiom aufnehmen. Intuitiv sollte sich dann wohl kein Zusammenhang, auf keinen Fall aber ein *zwingender* (d.h. nicht nur „als Default" vorliegender) Zusammenhang zwischen Freitagen und Krankheitstagen ergeben. In Moores System aber kommt ein solcher Zusammenhang heraus.

Genauer gesagt, behauptet Lukaszewicz, daß in der autoepistemischen Logik die Axiomenmenge S genau eine stabile Erweiterung besitze, in welcher $F \rightarrow A$ und $K \rightarrow \neg A$, also auch $F \rightarrow \neg K$ enthalten ist. Ich möchte hier nicht auf die Existenz- und Eindeutigkeitsbehauptung eingehen, sondern mich mit einem Beweis begnügen, daß es keine stabile Erweiterung von S gibt, die nicht $F \rightarrow \neg K$ (Am Freitag bin ich nicht krank.) enthält. Wir schreiben uns die Axiome wieder übersichtlicher auf:

Axiom 1 $\quad \neg F \vee A \vee L(K \vee \neg A)$
Axiom 2 $\quad \neg K \vee \neg A \vee LA$

Angenommen nun, es gibt eine stabile Erweiterung T von S, in der $F \rightarrow \neg K$ nicht enthalten ist. Das heißt, daß es im entsprechenden W eine Welt w gibt, in der $F \wedge K$ wahr ist. Wir unterscheiden nun zwei Fälle:

1. Fall: $w = \{F,K,A\}$. Wegen Axiom 2, das auch in w wahr sein muß, muß LA wahr in w sein, d.h. A muß in jedem $w' \in W$ wahr sein. Aber $v = \{\neg F,(\neg)K,\neg A\}$ ist eine wirkliche Welt derart, daß $\langle W,v \rangle$ in jedem Falle ein Modell von S ist; doch $v \notin W$, also ist T nicht korrekt bzgl. S, also kann T keine stabile Erweiterung von S sein.

2. Fall: $w = \{F,K,\neg A\}$. Wegen Axiom 1, das auch in w wahr sein muß, muß $L(K \vee \neg A)$ wahr in w sein, d.h. $K \vee \neg A$ muß in jedem $w' \in W$ wahr sein. Aber $v = \{(\neg)F,\neg K,A\}$ ist eine wirkliche Welt derart, daß $\langle W,v \rangle$ in jedem Falle ein Modell von S ist; doch $v \notin W$, also ist T nicht korrekt bzgl. S, also kann T keine stabile Erweiterung von S sein.

Damit haben wir die Annahme, es gebe eine stabile Erweiterung T von S, in der $F \rightarrow \neg K$ nicht enthalten ist, zum Widerspruch geführt. In Moores System folgt aus unserer harmlosen Freizeittheorie das kategorische Theorem, daß man am Freitag nicht krank sein kann, was eine offensichtlich inadäquate Folgerung ist.

Aber vielleicht war unsere (d.h. Lukaszewicz') Formalisierung nicht gut. Versuchen wir es einmal mit der Idee, die eventuell doch nicht ganz durchschaute Ausnahmeregelung wegzulassen, d.h. mit der Axiomenbasis

$$S = \{F \wedge MA \rightarrow A \ , \ K \wedge M\neg A \rightarrow \neg A\} \ .$$

Dann hätte man

Axiom 1′ $\neg F \vee A \vee L\neg A$,

der 1. Fall oben bliebe der gleiche, und im 2. Fall hätte $L(\neg A)$ die gleiche Wirkung wie $L(K \vee \neg A)$ im ursprünglichen Beispiel. Also hilft dieser Schachzug nicht.

Ein anderer Rettungsversuch geht in die entgegengesetzte Richtung, indem er die einseitige Ausnahmeregelung ausbaut zu einer zweiseitigen, wonach jedes Axiom auf das andere Rücksicht nimmt:

$$S = \{F \wedge M(\neg K \wedge A) \rightarrow A \ , \ K \wedge M(\neg F \wedge \neg A) \rightarrow \neg A\} \ .$$

Damit ergibt sich

Axiom 2′ $\neg K \vee \neg A \vee L(F \vee A)$;

der 2. Fall oben bliebe der gleiche, und im 1. Fall hätte $L(F \vee A)$ die gleiche Wirkung wie $L(A)$ im ursprünglichen Beispiel. Also ist auch dieser Ausweg versperrt.

Abgesehen davon, daß die eben versuchten Variationen eine unerwünschte Symmetrie in die Freitags- und die Krankheitsregel hineinbringen, können auch sie das Problem, eine völlig unmotivierte Konklusion zu beseitigen, nicht lösen. Wie Lukaszewicz anmerkt, löst sich das Problem dann, wenn man als Zusatzaxiom entweder F oder K oder $F \wedge K$ hat.[22]

Es ist aber unbefriedigend, wenn man zu einer Theorie stets irgendwelche Rand- oder Anfangsbedingungen dazunehmen *muß*, um kontraintuitive Schlußfolgerungen zu vermeiden.

Wir ziehen also das Fazit, daß das Beispiel von Lukaszewicz ebenso wie Touretzkys Beispiel ein echtes Problem für Moores autoepistemische Logik darstellt. Man kann beweisen, daß gewisse Formalisierungen zu wenig bzw. zu viele Schlußfolgerungen liefern. Es ist nicht ausgeschlossen, daß es Formalisierungen gibt, die offensichtliche Fehlschlüsse umgehen. Doch dann müßte man erst noch eine eigene Theorie der Formalisierung von natürlichsprachlichen Axiomen entwickeln, um mit der autoepistemischen Logik befriedigende Resultate zu erzielen.

[22] Nachrechnen zeigt: Die einzige stabile Erweiterung von $S \cup \{F\}$ ist $\{\{F,K,A\}, \{F,\neg K,A\}\}$, die einzige stabile Erweiterung von $S \cup \{K\}$ ist $\{\{F,K,\neg A\},\{\neg F,K,\neg A\}\}$, und die einzige stabile Erweiterung von $S \cup \{F,K\}$ ist $\{\{F,K,\neg A\}\}$. Also ist $F \rightarrow \neg K$ in keiner der stabilen Erweiterungen enthalten. Zudem liefert Moores System immer die intuitiv richtigen Theoreme über meine Anglerei. Man kann sich allerdings fragen, ob es sinnvoll ist, daß die Welt $\{F,K,A\}$ in der stabilen Erweiterung von $S \cup \{F\}$ möglich ist, denn bei freitäglicher Krankheit bleibe ich ja normalerweise zu Hause.

5.7 Autoepistemische Logik ohne Fundiertheit

In der nichtmonotonen und autoepistemischen Logik gibt es das notorische Problem, was man denn mit Axiomenmengen machen soll, die keinen oder die mehrere Fixpunkte bzw. stabile Erweiterungen haben. Man kann sich, wie Gelfond (1987), auf den Standpunkt stellen, solche Axiomenbasen seien an sich schon unvernünftig, und versuchen, die eindeutig erweiterbaren Axiomenmengen syntaktisch zu charakterisieren. Ich glaube aber, das letzte Wort, wie eine richtige autoepistemische Theorie auszusehen hat, ist noch nicht gesprochen. In Abschnitt 5.4 habe ich Zweifel geäußert, ob die Fundiertheitsforderung der autoepistemischen Logik wirklich das leistet, was man von ihr verlangen würde. In diesem Abschnitt möchte ich noch kurz die Frage aufwerfen, ob man die Fundiertheit tatsächlich braucht.

Mein Alternativvorschlag besteht einfach darin, die inkonsistente Theorie $T_\perp$ auszuschließen. Wenn wir das tun, ist es klar, daß man in den Klauseln (b) und (c) der Definition 5.4.3 stabiler Theorien „genau dann, wenn" anstelle von „wenn ..., dann" schreiben können. Setzen wir keine weiteren Restriktionen (also insbesondere keine Fundiertheitsbedingung) an, so sehen wir, daß die in der nichtmonotonen und autoepistemischen Logik fixpunkt- bzw. erweiterungslosen Axiomenmengen $\{MA{\rightarrow}\neg A\}$ und $\{MA{\rightarrow}B, \neg B\}$ plötzlich Sinn machen: Wie man sich leicht überlegt, ist $MA{\rightarrow}\neg A$ in einer stabilen Theorie T genau dann, wenn $\neg A$ in T ist; und $MA{\rightarrow}B$ und $\neg B$ sind gleichzeitig in einer stabilen Theorie T genau dann, wenn $\neg A \wedge \neg B$ in T ist. Dies scheint im Einklang mit unseren Intuitionen zu stehen. Die Schwierigkeiten der nichtmonotonen Logik mit diesen Beispielen verschwinden.

Es bleiben aber echte Probleme. Betrachten wir noch einmal das bereits am Ende von Abschnitt 5.5 untersuchte Axiomenpaar $S = \{MA{\rightarrow}\neg B, MB{\rightarrow}\neg A\}$, das ich jetzt mit einem *Hochzeitsrätsel* illustrieren will. Anna und Berta sind zwei Schwestern. Sie sind beide in denselben Mann verliebt, nennen wir ihn Olli, und wären glücklich, ihn zu heiraten. Aber die Schwesternliebe ist noch größer, und so sagt Anna zu Berta: „Solange ich nicht weiß, daß du Olli nicht heiratest, werde ich ihn auch nicht heiraten." Und Berta sagt dasselbe zu Anna. Es scheint korrekt, das gegenseitige Versprechen durch S zu formalisieren (ohne wesentliche Vorkommnisse von Zeitparametern oder ähnlichem). Wie Moore (1984) nachweist (vgl. Abschnitt 5.5), hat S in seiner autoepistemischen Logik zwei stabile

Erweiterungen. Eine davon enthält $\neg A$, aber nicht $\neg B$ (noch B), die andere enthält $\neg B$, aber nicht $\neg A$ (noch A). Läßt man die Fundiertheit fallen und fordert stattdessen $T \neq T_\perp$ für stabile Theorien T, so ändert sich die Sachlage nicht wesentlich. Man stellt fest, daß S genau dann Teilmenge einer stabilen Theorie T ist, wenn $\neg A$ oder $\neg B$ in T enthalten ist: Sei o.E.d.A. $\neg A$ in T ($\neg B$ in T geht genauso), dann ist auch $\neg MA$ in T, woraus die Elemente von S aussagenlogisch folgen, also ebenfalls in T sind. Für die umgekehrte Richtung nehmen wir an, die beiden Axiome, aber weder $\neg A$ noch $\neg B$ seien in T; aus dem letzten Teil der Annahme folgt, daß MA und MB in T sind, woraus sich aus dem ersten Teil der Annahme mit Modus Ponens ergibt, daß $\neg B$ und $\neg A$ in T sind, im Widerspruch zum letzten Teil der Annahme.

Man beachte, daß die verzwickte Lage nicht durch die simple Monogamiebehauptung $\neg A \vee \neg B$ erfaßt werden kann — das wäre intuitiv zu schwach. Man beachte gleichfalls, daß die andere symmetrische Lösung, nämlich $\neg A \wedge \neg B$, intuitiv viel zu stark wäre — Anna muß nur „Diesen Kerl kann ich unmöglich heiraten" sagen, und Berta kann das Aufgebot bestellen (wenn Olli einverstanden ist). Durch den Verzicht auf die Fundiertheitsbedingung geht hier der Schutz vor der intuitiv zu starken Lösung verloren.

Intuitiv hat man keine Probleme, die liebenden Schwestern zu verstehen, eine formale Aufdröselung aber ist bisher nicht gelungen (jedenfalls bei der vorgeschlagenen Formalisierung). Das ist das Rätsel: *Wie* sehen die Überzeugungen von Anna und Berta aus, nachdem sie sich das völlig aufrichtige und völlig symmetrische Versprechen gegeben haben?

5.8 Fortschritte der autoepistemischen Logik

Neuerdings ist es Kurt Konolige (1988) gelungen, weitere Einsichten über die Eigenschaften der autoepistemischen Logik zu gewinnen. Wir können hier nicht auf seinen bemerkenswerten Nachweis eingehen, daß sich autoepistemische und Default-Logik ineinander übersetzen lassen und in einem gewissen Sinn sogar als äquivalent bezeichnet werden können.[23] Für uns sind

[23] Die autoepistemische Logik gestattet aber die Formulierung mehrerer verschiedener Begriffe von stabilen Erweiterungen und ist insofern flexibler als die Default-Logik. Außerdem gilt die Äquivalenz vermutlich nur auf aussagenlogischem Niveau, auf welches wir uns im folgenden wieder beschränken wollen.

vor allem die vorbereitenden Resultate Konoliges interessant.

Als erstes gibt Konolige drei verschiedene Charakterisierungen stabiler Erweiterungen:

5.8.1. Theorem T ist genau dann eine stabile Erweiterung einer Axiomenmenge S, wenn eine der drei folgenden Bedingungen erfüllt ist:
(a) T ist die Menge derjenigen Sätze, die alle AE-Interpretationen von T erfüllen, welche Modelle von S sind.
(b) T ist die Menge derjenigen Sätze, die alle AE-Interpretationen erfüllen, welche Modelle von $S \cup L(T) \cup M(T)$ sind.
(c) T ist die Menge derjenigen Sätze, die alle AE-Interpretationen stabiler Theorien erfüllen, welche Modelle von $S \cup L(T_0) \cup M(T_0)$ sind, wobei $L(T_0) = \{LA: A \text{ objektiv} \wedge A \in T\}$ und $M(T_0) = \{MA: A \text{ objektiv} \wedge \neg A \notin T\}$ ist.

Im Teil (b) von Theorem 5.8.1 ist das semantische Analogon von Moores Definition (vgl. (5.4.9)) zu finden. Wegen der Adäquatheit der Aussagenlogik kann man deren syntaktische Ableitbarkeitsrelation anstelle der semantischen Folgerungsrelation setzen. Wie ist aber Teil (c) des Theorems syntaktisch wiederzugeben, was bewirkt die Quantifikation über alle AE-Interpretationen *stabiler Theorien*? Während bei Moore die Rolle des modallogischen Systems **K45** (des „schwachen **S5**") nicht ganz durchsichtig war, zeigt Konolige, daß es hier am richtigen Ort ist:[24]

5.8.2. Theorem T ist genau dann eine stabile Erweiterung einer Axiomenmenge S, wenn T ein Fixpunkt des durch

$$NM_S(R) := Th_{\mathbf{K45}}(S \cup L(R_0) \cup M(R_0))$$

definierten Operators NM_S ist.

Konolige übt weiter intuitive Kritik an der Mooreschen Fundiertheitsbedingung. Ihm kommen ganz ähnliche Zweifel, wie wir sie in Abschnitt 5.4 geäußert haben, wenn er kritisiert, daß $S = \{LA \rightarrow A\}$ eine stabile Erweiterung hat, die A enthält. (Zur Erinnerung: Wir hatten bemängelt, daß die inkonsistente Theorie $T_\perp$ eine stabile Erweiterung von $S = \{LA \rightarrow A, \neg A\}$ ist.) Um diese Inadäquatheit auszuschalten, verstärkt Konolige die Mooresche Bedingung:

5.8.3. Definition T ist genau dann *mäßig fundiert in einer Axiomenmenge S*, wenn T ein Fixpunkt des durch

$$NM_S(R) := Th_{\mathbf{K45}}(S \cup L(S) \cup M(R_0))$$

definierten Operators NM_S ist.

Mäßig fundierte Theorien sind wenig verpflichtende Theorien über die

[24] Lemma 3.4 und Proposition 3.5 in Konolige (1988) sind ziemlich nachlässig formuliert und bewiesen, führen aber zu einem korrekten Ergebnis.

Welt:

5.8.4. Theorem T ist genau dann mäßig fundiert in S, wenn T eine minimale stabile Erweiterung von S ist in dem Sinn, daß es keine stabile Theorie T′ mit $S \subseteq T'$ gibt, deren objektiver Teil eine echte Teilmenge des objektiven Teils von T ist.

Aber auch diese Verschärfung des offenbar zu schwachen Fundiertheitsbegriffs von Moore reicht nach Konolige nicht aus. Er bringt ein — allerdings schwer einzuschätzendes — Beispiel, welches zeigen soll, daß Definition 5.8.3 immer noch zu viele Fixpunkte zuläßt. Wichtiger ist noch, daß für die gegenseitige Anpassung von autoepistemischer und Default-Logik der folgende Begriff eine zentrale Rolle spielt:

5.8.5. Definition Sei S eine Axiomenmenge in Normalform und T eine Erweiterung von S. Sei $S' \subseteq S$ die Menge der Sätze, deren objektiver Teil in T enthalten ist. Dann heißt T genau dann *streng fundiert in S*, wenn T ein Fixpunkt des durch

$$NM_S(R) := Th_{\mathbf{K45}}(S' \cup L(S') \cup M(R_0))$$

definierten Operators NM_S ist.

Wir wollen uns hier nicht darum kümmern, wie die Normalform eines Axioms nach Konolige aussieht. Es genüge die Bemerkung, daß jede beliebige Formel der modalen Aussagenlogik **K45**-äquivalent mit einer Formel ist, die keine Verschachtelungen von L und M enthält. In S streng fundierte Theorien sind auch mäßig fundiert in S. Eine sonderbare Eigenschaft des in Definition 5.8.5 eingeführten Begriffs ist es, daß **K45**-äquivalente Axiomenmengen S und S′ (in Normalform) verschiedene stark fundierte Erweiterungen haben können. Die starke Fundiertheit ist also sensitiv gegenüber der oberflächensyntaktischen Formulierung der Axiome. Konolige findet daran nichts Anstößiges, zumal diese Eigenschaft ja eine Entsprechung in der Default-Logik findet.

Wir sehen also, daß es unterschiedliche Rationalitätskriterien für „richtige" nichtmonotone Erweiterungen von (modalisierten) Axiomenmengen gibt. Das letzte Wort darüber, was intuitiv gerechtfertigt werden kann und soll, ist sicher noch nicht gesprochen. Auf diesem philosophisch und formal äußerst interessanten Gebiet werden so schnell Fortschritte erzielt wie wohl auf keinem anderen, auf welches ich zu sprechen komme. Deshalb ist es sicher, daß zum Zeitpunkt des Erscheinens dieses Buchs bereits weitreichende neue Erkenntnisse vorliegen.[25]. Wir wollen die Diskussion an dieser Stelle abbrechen und den Faden aus Kapitel 4 wieder aufnehmen.

[25] Auf Marek und Truszczyński (1989) und Levesque (1990) sei besonders hingewiesen

Kapitel 6

Konditionalsätze und Erklärungen

6.1 Wenn-dann und weil

In den letzten beiden Kapiteln haben wir eine Möglichkeit ausgekundschaftet, wie man die Interpretation von Konditionalsätzen gesehen durch den Ramsey-Test aufrechterhalten kann. Ein entscheidender Punkt war, daß für Sprachen mit Konditionalsätzen oder autoepistemischen Modaloperatoren eine „konsistente" Revision nicht mehr mit einer Expansion gleichgesetzt werden durfte. Im vorliegenden Kapitel wir diese Einsicht noch offensichtlicher, indem wir verschiedene Arten von Konditionalsätzen unterscheiden werden. Sei T eine Theorie, in der weder A noch $\neg$A enthalten ist. In T ist offen, ob A wahr oder falsch ist. Deshalb ist, wie ich unten argumentiere, Wenn A, dann B als indikativischer Konditionalsatz auszudrücken: Wenn A der Fall *ist*, dann *ist* B der Fall. Im Gegensatz dazu ist $\neg$A in $T^+_{\neg A}$ enthalten, d.h. A wird in $T^+_{\neg A}$ als kontrafaktisch angesehen. Deshalb muß ein kontrafaktischer Konditionalsatz formuliert werden: Wenn A der Fall *wäre*, dann *wäre* B der Fall. Weil $\neg$A in $T^+_{\neg A}$ ist, deshalb *darf* man in $T^+_{\neg A}$ keinen indikativischen Konditionalsatz mit dem Antezedens A akzeptieren. Da aber andererseits $T^+_{\neg A}$=Cn(T$\cup\{\neg A\}$) für monotone Logiken Cn eine Obermenge von T ist, ist in diesem Fall der indikativische Konditionalsatz doch in $T^+_{\neg A}$. Daraus folgt, daß entweder eine nichtmonotone Logik Cn zu verwenden ist oder daß T^+_A nicht gleich der konsistenten Revision T^*_A,

der Addition T°_{A}, sein kann.

Wie wir aber in Kapitel 5 gesehen haben, hat das nichtmonotonen Schließens im allgemeinen ziemlich vertrackte Eigenschaften, und über den Inhalt von Additionen, die keine Expansionen sind, wissen wir bis jetzt leider so gut wie nichts. Da wir das Revisionsmodell auf wissenschaftlichen Theorienwandel anwenden und dabei Explikate für Reduktionen und intertheoretische Erklärungen liefern wollen, sind wir auf Mittel angewiesen, mit denen man weiter kommt. Deshalb wollen wir für unsere weiteren „praktischen" Aufgaben genau das tun, was wir in Kapitel 4 (besonders Abschnitt 4.6) aus philosophischen Gründen für fragwürdig erklärt haben. Wir werden darauf achten, daß wir bei der Addition eines „objektiven" Satzes A zu einer Theorie T keine Konditionalsätze oder (autoepistemisch interpretierten) Modalsätze von T nach T°_{A} mitschleppen. Der „objektive Kern" $\mathbf{Obj}(T^{\circ}_{A})$ von T°_{A} wird einfach durch die Erweiterung $(\mathbf{Obj}(T))^{+}_{A}=Cn(\mathbf{Obj}(T)\cup\{A\})$ des „objektiven Kerns" $\mathbf{Obj}(T)$ von T festgelegt. Des weiteren setzen wir wieder eine *monotone Logik* Cn voraus. Dies sind, wie gesagt, rein pragmatische Entscheidungen, die — jedenfalls beim gegenwärtigen Stand der Forschung — nötig sind, um zu Ergebnissen zu kommen. Man möge sich anhand der Adäquatheit der im folgenden präsentierten Ergebnisse ein Bild davon machen, wie gravierend die Folgen der philosophischen Nachlässigkeit sind.

Der Ausgangspunkt der folgenden Überlegungen ist der Gedanke, daß wenn(-dann) und weil(-deshalb)[1] verwandte Konjunktionen sind. Dies wurde schon häufig bemerkt. Goodman (1954, S. 14) hält Weil A, B für äquivalent mit seiner „Kontraposition", dem konjunktivischen Konditionalsatz Wenn ¬B, dann ¬A und nennt entsprechend weil (since) das „faktische Konditional". Ryle (1963, S. 317) sieht zumindest eine wahrheitsfunktionale Gleichheit zwischen Weil A, B und dem konjunktivischen Konditionalsatz Wenn ¬A, dann ¬B. Die meistverbreitete Ansicht ist aber wohl bereits bei Ramsey (1931, S. 248) zu finden, der Wenn p, dann q und Weil p, q in einem Atemzug nennt und feststellt: „weil ist nur eine Variante von wenn-dann, wenn p als wahr bekannt ist". Philosophiestudenten in den ersten Semestern können ziemlich genau dasselbe in Ulrich Blaus hochgeschätzter Propädeutik (1982/83, S. 43) lernen: „Eng verwandt mit wenn-dann sind die partiell wahrheitsfunktionalen Operatoren weil, da, also, deshalb, etc. Mir scheint, sie sind im wesentlichen die Konjunktion von A und wenn A, dann B." Obgleich die Meinungen über den genauen Zusam-

[1] Ich gehe davon aus, daß das deutsche weil und die englischen since und because synonym sind.

menhang zwischen wenn-dann und weil weit auseinandergehen, möchte ich im folgenden die Ansicht von Ramsey und Blau die *Normalanalyse* der Beziehung von wenn-dann und weil nennen.

Bemerkenswert ist das unterschiedliche Maß an Aufmerksamkeit, das den beiden Konnektiven in der modernen Philosophie zuteil geworden ist. Während wenn-dann weitreichendste Beachtung in den mittlerweile wuchernden Konditionallogiken erfahren hat, hat weil bislang ein Mauerblümchendasein fristen müssen. Ich meine, man sollte dem Wörtchen weil mehr Gerechtigkeit widerfahren lassen, und ich werde eine parallele Analyse der beiden Konjunktionen versuchen. Dabei will ich mich nicht von vornherein auf eine der einleitend erwähnten Ansichten verpflichten, sondern mit der Idee beginnen, daß weil immer auf eine Erklärung abzielt. Zunächst werde ich die grundlegenden Ideen zweier vielversprechender Ansätze der Analyse von Konditionalsätzen und Erklärungen wiedergeben, wobei der Begriff der Erklärung mit dem des Grundes in Zusammenhang gebracht wird. Das gemeinsame Merkmal dieser Analysen ist, daß sie auf den Überzeugungen und Überzeugungsänderungen eines epistemischen Subjekts aufbauen oder, anders gewendet, daß sie auf ein Theorienrevisionsmodell relativiert werden. Im Zusammenhang der gemeinsamen Betrachtung von wenn-dann und weil wird sich die Analyse von Konditionalsätzen durch den Ramsey-Test dann als inadäquat herausstellen. Die Normalanalyse von weil deutet darauf hin, daß man den aus Kapitel 4 bekannten *Starken Ramsey-Test* (R2) verwenden sollte, ich werde aber gemäß der zuvor angeführten Explikationen von Gründen und Erklärungen eine alternative Analyse von weil vorschlagen. Diese Analyse kann zugleich für wenn-dann hergenommen werden, indem sie zur Definition eines (natürlichsprachlich nicht realisierten) *universellen (Pro-)Konditionals* führt. Auf natürliche Weise kann auch ein *universelles Kontrakonditional* und ein *universelles Nichtkonditional* eingeführt werden. Danach stelle ich eine Handvoll von allgemeinen Thesen über die Einordnung von indikativischem und konjunktivischem wenn-dann, wenn-dann möglicherweise, auch wenn, weil und obwohl im Koordinatensystem der universellen Konditionale auf. Die relativ komplizierten Akzeptabilitätsbedingungen dieser natürlichsprachlichen Konjunktionen reduziere ich auf handliche Formulierungen, indem ich die *Randbedingungen* bezüglich der Akzeptanz von „Antezedenzien" und „Konsequenzen" mit verarbeite. Dadurch können die allgemeinen Thesen aus der Perspektive der am Ende resultierenden Bedingungen noch einmal beurteilt werden. Schließlich vergleiche ich meinen Ansatz mit dem meines Wissens einzigen neueren Versuch einer systematischen gemeinsamen Be-

handlung von wenn-dann und weil, der Theorie von Storrs McCall.

Es gibt wohlbegründete Zweifel daran, ob es angemessen ist, Konditionalsätzen und Erklärungen objektive Wahrheitswerte zuzuschreiben. Diesen Zweifeln werde ich gerecht, indem ich *epistemische* oder *theorienrelative* Interpretationen als Ausgangspunkt wähle und ein Modell verwende, welches einen Rahmen für Akzeptabilitätsbedingungen anstelle von Wahrheitsbedingungen bietet: das in Kapitel 3 bearbeitete Revisionsmodell von Peter Gärdenfors. Neu gegenüber ähnlichen Interpretationen verschiedener Sorten von Konditionalsätzen (vgl. Gärdenfors 1981; 1988, Abschnitt 7.3) ist die konjunktive Aufnahme einer Relevanzbeziehung durch den Starken Ramsey-Test und die Aufnahme von Randbedingungen bezüglich der Akzeptanz von Antezedenzien und Konsequenzen in die Bedeutung von Konditionalsätzen. Sprachphilosophisch mag es fragwürdig erscheinen, Relevanz und Randbedingungen mit zur Semantik von Konditionalsätzen zu rechnen. Vielleicht wären beide Punkte besser als Bedingungen einer angemessenen Verwendung von Konditionalsätzen im Reich der Pragmatik aufgehoben, dort, wo man Gricesche Mechanismen am Werk vermutet. Trotz dieser sprachphilosophischen Unsicherheit bin ich zuversichtlich, daß sich der Nebel im Grenzbereich zwischen Semantik und Pragmatik für Konditionalsätze (im weitesten Sinne) nicht als schädlich herausstellen wird.[2] Daß ich einige ziemlich starke Bedingungen mit einem pragmatischen Anstrich in die Akzeptabilitätsbedingungen einverleibe, soll im Gegenteil dazu dienen, eine gemeinsame Analyse der verschiedenen Arten von wenn-Sätzen, von weil und obwohl erst zu ermöglichen. Eine Beurteilung oder Verurteilung meines Ansatzes sollte erst am Ende, nachdem die endgültigen, vereinfachten Versionen der Akzeptabilitätsbedingungen bekannt sind, ausgesprochen werden.

[2] Es gibt einen einzigen Fall, in dem ich zweifellos eher zu einer Behauptbarkeitsbedingung anstelle einer Akzeptabilitätsbedingung gelange, und dies ist der Fall des indikativischen wenn-dann möglicherweise (vgl. Fußnote 22). Es wäre lehrreich, die „Randbedingungen" als Präsuppositionen und nicht als Konjunkte in die Akzeptabilitätsbedingungen unten einzubauen und zu versuchen, dann Vereinfachungen ähnlich wie in Abschnitt 6.7 auszuführen. Da ich mir über das geeignete Konzept einer Präsupposition jedoch nicht im klaren bin, werde ich einen solchen Versuch hier nicht unternehmen.

6.2 Konditionalsätze und Erklärungen: der Ausgangspunkt der Analyse

Als Ausgangspunkt unserer Überlegungen betrachten wir beispielhafte, nicht formalisierte Ansätze der Interpretation von Konditionalsätzen und Erklärungen, welch letztere paradigmatisch als weil-Sätze formuliert werden. Wir streben an, am Ende zu einer vereinheitlichten Theorie von wenn-dann und weil zu kommen.

Den Ramsey-Test, der einen der wichtigsten Vorschläge zur Analyse von *Konditionalsätzen* darstellt, haben wir in Kapitel 4 schon in seiner formalisierten Version kennengelernt. Wie der Name es ausdrückt, geht er auf Ramsey (1931, S. 247, Fußnote 7) zurück. Für die Logik des wenn-dann wurde er zuerst von Stalnaker systematisch fruchtbar gemacht, welcher die folgende Formulierung gibt:

(6.2.1) This is how to evaluate a conditional:
First, add the antecedent (hypothetically) to your stock of beliefs; second, make whatever adjustments are required to maintain consistency (without modyfying the hypothetical belief in the antecedent); finally, consider whether or not the consequent is then true. (Stalnaker 1968, S. 102)

Ein Hauptproblem der Wissenschaftstheorie ist das Problem der (wissenschaftlichen) *Erklärung* singulärer Sachverhalte. Die Verdienste des folgenden Vorschlags von Gärdenfors kann man in der Originalarbeit (1980, S. 404; vgl. auch Gärdenfors 1988, S. 168) und in Stegmüller (1983, Kapitel XI) nachlesen. Hier kann nur die Idee wiedergegeben werden:

(6.2.2) The central criterion on an explanation is that the explanans in a non-trivial way increases the belief-value of the explanandum, where the belief-value of a sentence is determined from the given knowledge situation.

Weil-Sätze können auch Begründungen formulieren. Mir scheint, der Begriff der Erklärung ist schwieriger zu fassen als der des Grundes. Es ist nützlich, sich zuerst über den einfacheren Begriff Klarheit zu verschaffen. Weiter plädiert Spohn (1983) überzeugend dafür, daß eine Theorie der Kausalität auf dem Begriff des *Grundes* aufbauen sollte. Seine Explikation (1983, S. 372) von „Grund" ist ziemlich einfach:

(6.2.3) A is a reason for B for the person X at time s iff X's believing A at s would raise the epistemic rank of B for X at time s.

Natürlich kann man mit weil-Sätzen auch Kausalaussagen treffen, weshalb

an dieser Stelle eigentlich ein einfaches Modell der Kausalität fällig wäre. Ich glaube jedoch, daß wir damit unserem Ziel kaum näherkommen würden. Einerseits hat nämlich schon Spohn gezeigt, auf welche — durchaus nichttriviale — Weise man (direkte) Ursachen aus Gründen gewinnen kann.[3] Andererseits scheint Kausalität als ein in der Wirklichkeit anzusiedelndes Phänomen, das nicht allein an Überzeugungen und Theorien festzumachen ist, über die Grenzen hinauszugehen, die durch Konditionalsätze, Erklärungen und Gründe abgesteckt sind. Im folgenden gebe ich also keine Interpretation für das „kausale" weil, welches wohl eine asymmetrische, von epistemischen Zuständen prima facie unabhängige Kausalrelation in unserer wirklichen Welt voraussetzt. Ich werde mich auf das „informative" weil beschränken, welches nur Gründe oder Erklärungen spezifiziert. Möglicherweise ist jedoch dieses informative weil das verbreitetere und das allgemeinere weil, und das kausale weil kann zu guter Letzt durch die Angabe einiger zusätzlicher „realistischer" Bedingungen als ein Spezialfall des informativen charakterisiert werden.

Ein Vergleich von (6.2.1)–(6.2.3) nährt jedenfalls die Hoffnung auf eine Zusammenführung von wenn-dann und weil. Alle drei Formulierungen handeln von Überzeugungsänderungen, d.h. vom Theorienwandel in epistemischen Subjekten. Gärdenfors (1979, 1981, 1982, 1984) hat für die Analyse von Konditionalsätzen ein geeignetes Modell der Dynamik epistemischer Zustände entwickelt — eben jenes, das wir in seiner aktuellsten Version in Kapitel 3 kennengelernt haben. Für die folgenden Überlegungen benötigen wir das Modell nicht in seiner vollen Stärke, sondern können uns mit den sehr plausiblen *grundlegenden* Gärdenfors-Postulaten (T*1)–(T*6) für Revisionen und (T–1)–(T–6) für Kontraktionen begnügen.[4] Um uns die wenig instruktiven, formal meist gesondert zu behandelnden „Grenzfälle" zu ersparen, wollen wir davon ausgehen, daß wir es von nun an immer mit konsistenten Theorien $T \neq T_\perp$ und mit kontingenten (weder logisch wahren

[3] Ganz grob gesagt, lautet Spohns (1983, S. 388) Vorschlag so: Eine Proposition A ist eine *direkte Ursache* für eine Proposition B (in der wirklichen Welt w_o) genau dann, wenn A und B (in w_o) wahr sind, A zeitlich vor B liegt und A unter den (in w_o) obwaltenden Umständen ein Grund für B ist. Die obwaltenden Umstände sind dabei die ganze Vergangenheit der wirklichen Welt bis hin zur Referenzzeit von B, mit Ausnahme desjenigen „Faktors", zu dem A gehört. Die Definition von *Ursachen* schlechthin aus direkten Ursachen ist ein schwieriges, von Spohn bisher noch nicht vollständig gelöstes Problem.

[4] Es ist jedoch zu beachten, daß (6.2.1)–(6.2.3) eindeutige Revisionen voraussetzen — wie man sie über die Relation der theoretischen Wichtigkeit durch eine der Konstruktionen aus Kapitel 3 erhalten kann. Die Gärdenfors-Postulate reichen nicht hin, um für gegebenes T und gegebenes A ein eindeutiges T^*_A oder T^-_A festzulegen.

noch logisch falschen) Sätzen A, B, etc. zu tun haben.

6.3 Die Normalanalyse und der Starke Ramsey-Test

Gärdenfors hat sein Modell des Überzeugungs- oder Theorienwandels gerade zur Analyse von Konditionalsätzen gemäß dem Ramsey-Test entwickelt (vgl. Kapitel 4). Weil wir im Verlauf dieses Kapitels verschiedene Arten von Konditionalsätzen und verschiedene Ansätze zu ihrer Analyse betrachten werden, wollen wir für das natürlichsprachliche Explikandum Wenn A, dann B von nun an verschiedene Explikate in symbolischer Notation untersuchen. Wir fügen zu unserer Objektsprache zunächst das nicht wahrheitsfunktionale Konditional > hinzu, welches dem *Ramsey-Test* unterworfen ist:

(6.3.1) $A>B \in T \Leftrightarrow B \in T^*_A$.

Diese Formulierung des Ramsey-Tests liefert zusammen mit einer modifizierten Kollektion der Gärdenfors-Postulate (T*1)–(T*8) Lewis' (1973a) „offizielle" Logik **VC** für Konditionalsätze (vgl. Kapitel 4, Abschnitt 2). Insbesondere ergibt sich aus (T*3) und (T*4) das Stabilitätskriterium (T*S), d.h. daß $T^*_A=T$ gilt, falls A schon in $T(\neq T_\perp)$ ist. Deshalb folgt aus (6.3.1) als Spezialfall

(6.3.2) $A,B \in T \Rightarrow A>B \in T$.

Nun wird sofort klar, daß die Normalanalyse der Beziehung von weil und wenn-dann in diesem Rahmen scheitern muß. Wenn wir weil durch $^a>$ symbolisieren,[5] liest sie sich in der Version von Ramsey bzw. Blau folgendermaßen:

(6.3.3) $A \in T \Rightarrow (A^a>B \in T \Leftrightarrow A>B \in T)$,

(6.3.4) $A^a>B \in T \Leftrightarrow A \in T \wedge A>B \in T$.

Unter Zuhilfenahme von (6.3.2) erweisen sowohl (6.3.3) als auch (6.3.4) die folgende Bedingung als gültig:

(6.3.5) $A,B \in T \Rightarrow A^a>B \in T$.

[5] Der nun folgende Pfeilchenzoo sollte niemanden vom Weiterlesen abhalten. Die Systematik ist einfach. Pfeile > kennzeichnen vorläufige Analysen von natürlichsprachlichen Konditionalen, die in Abschnitt 6.5 eingeführten Pfeile $\rightarrow$ (bzw. $\prec$ und $\multimap$) kennzeichnen (Spezialisierungen von) universelle(n) Konditionale(n). Ein Superskript vor einem Pfeil soll den Akzeptanzstatus des Antezedens markieren: „^{a}" steht für „akzeptiert", „^{z}" steht für „zurückgewiesen" und „^{o}" steht für „offen". Schließlich werden einige Unterarten mit Hilfe des Quadrats □ und der Raute ◇ klassifiziert.

Dies ist offensichtlich absurd, wenn $^{a}>$ weil repräsentieren soll.

Was soll man jetzt tun? Im Moment möchte ich die Normalanalyse nicht ablehnen, sondern das Problem bei der Bedingung (6.3.2) anpacken. Rechtfertigt das bloße Akzeptieren von A und B tatsächlich das Akzeptieren von Wenn A, dann B? Man betrachte den Satz

Wenn Anton zur Party geht, dann geht Berta zur Party.

Würden wir diesen Konditionalsatz akzeptieren, wenn wir wüßten, daß sowohl Anton als auch Berta zur Party gehen, aber daß sich Bertas frühere Sympathie für Anton inzwischen in eine heftige Abneigung verwandelt hat, so daß Antons Anwesenheit schon beinahe Grund genug war für Berta, nicht zu kommen? Wohl kaum. Außer in den Logeleien von Wochenendausgaben anspruchsvoller Zeitungen drücken Konditionalsätze in der natürlichen Sprache eine *positive Verknüpfung* zwischen dem Antezedens und dem Konsequens aus, das Antezedens muß *relevant* für das Konsequens sein. (Später werde ich dafür plädieren, daß dies nicht der einzige Grund dafür ist, daß anstelle von wenn-dann im obigen Satz obwohl verwendet werden muß, wenn man die Geschichte von Anton und Berta richtig wiedergeben will.)

Will man das Gärdenforssche Modell nicht ganz aufgeben, dann sehe ich nur eine Möglichkeit, (6.3.2) zu unterbinden. Wir müssen (6.3.1) so abändern, daß die schiere Tatsache, daß A und B in einer Theorie enthalten sind, nicht ausreicht, um die rechte Seite wahr zu machen. Ich meine, die richtige Idee besteht darin, den Ramsey-Test zum *Starken Ramsey-Test* zu verschärfen, der uns als Bedingung (R2) aus Abschnitt 4.2 schon bekannt ist. Für Konditionalsätze, die dem Starken Ramsey-Test gehorchen, führen wir ein neues Symbol $\gg$ ein:

(6.3.6) $A \gg B \in T \Leftrightarrow B \in T^*_A \wedge B \notin T^*_{\neg A}$.

Mit $\gg$ anstelle von $>$ liefert die Normalanalyse eine ziemlich plausible Bedingung für die kritische Situation $A,B \in T$:

(6.3.7) Falls $A,B \in T$, gilt: $A^a{>}B \in T \Leftrightarrow B \notin T^*_{\neg A}$.

Angesichts der Tatsache, daß der Ramsey-Test in erster Linie für kontrafaktische Konditionalsätze gedacht war, ist es erfreulich, daß sich der Starke Ramsey-Test in typisch „kontrafaktischen“ (oder „kontratheoretischen“) Kontexten auf den normalen Ramsey-Test reduziert:

(6.3.8) Falls $B \notin T$, gilt: $A \gg B \in T \Leftrightarrow A > B \in T$.

Dies sieht man sofort durch Anwendung von Lemma 3.1.6 aus Kapitel 3. Insgesamt ist (6.3.6) ein sowohl für wenn-dann als auch für weil vielver-

sprechender Ansatz.[6] Das Konnektiv $\gg$ scheint der natürlichen Sprache angemessener zu sein als $>$, weil es explizit verlangt, daß das Antezedens für das Konsequens (positiv) relevant ist. Angenommen Anton ginge nicht zur Party. Dann würde Berta erst recht zur Party gehen. Also sagt der Starke Ramsey-Test, daß wir $A \gg B$ nicht akzeptieren — im Einklang mit unseren Intuitionen hinsichtlich des in Frage stehenden Konditionalsatzes.

Probieren wir noch alternative Ideen zur Verhinderung von (6.3.2) aus. Man macht sich leicht klar, daß der Begriff der Relevanz durch die Variation (R1) des Ramsey-Tests (s. Abschnitt 4.2) nicht richtig eingefangen werden kann, wenn man auch weil-Sätze analysieren will. Die Bedingung $B \in T^*_A \wedge B \notin T$ hätte einen gegenteiligen, aber nicht minder unerwünschten Effekt wie (6.3.5): Es wäre nach der Normalanalyse (6.3.3) unmöglich, $A >^a B$ zu akzeptieren, wenn man A akzeptiert, und gemäß der Normalanalyse (6.3.4) schon, wenn man A oder B akzeptiert. Eine andere Methode zu vermeiden, daß das Akzeptieren von A und B das Akzeptieren von Wenn A, dann B erzwingt, besteht in der Variation (R3) (s. Abschnitt 4.2), d.h. darin, die ursprüngliche Theorie vor der Anwendung des Ramsey-Tests zu kontrahieren. Eine Kontraktion bzgl. des Antezedens A würde nichts taugen, da $(T^-_A)^*_A = T^*_A$ (falls $A \in T$, gilt dies im wesentlichen wegen (T–5), falls $A \notin T$ wegen (T–3)). Aber eine Kontraktion bzgl. des Konsequens B ist interessant. Wir landen bei der Ramsey-Test-Variation (R3) (s. Abschnitt 4.2) mit der Bedingung $B \in (T^-_B)^*_A$, die — wenn sie auch bisher kaum motiviert ist — keine offensichtlichen Fehler hat. Die Idee von (R3) wird uns im nächsten Abschnitt in einem ganz anderen Zusammenhang wieder begegnen. (Die Variation (R4) wird in Abschnitt 6.8 begutachtet.)

Gärdenfors (1981, S. 209f; 1988, S. 154–156) analysiert auch might-conditionals im Geiste des Ramsey-Tests (vgl. auch Abschnitt 4.5). Schreiben wir $A \diamond\!\!> B$ für Wenn A, dann möglicherweise B, dann lautet sein Vorschlag so:

(6.3.9) $A \diamond\!\!> B \in T \Leftrightarrow \neg B \notin T^*_A$.

Eine dem Starken Ramsey-Test analoge, stärkere Version ist

(6.3.10) $A \ll\!\diamond\!\gg B \in T \Leftrightarrow \neg B \notin T^*_A \wedge \neg B \in T^*_{\neg A}$.

[6] Wegen der Stärke von (6.3.6) ist die Logik von $\gg$ ziemlich schwach. Aus später anzuführenden Gründen glaube ich, daß die Logiken von $^z\Box\!\!\rightarrow$, $^\circ\Box\!\!\rightarrow$, $^a\!\!\rightarrow$ usw. (s. Abschnitt 6.7) wichtiger und interessanter sind als die desjenigen Konnektivs, welches allein durch den Starken Ramsey-Test interpretiert wird. Es ist überraschend, daß die doppelte Pragmatisierung (Relevanz plus Randbedingungen) sich häufig am Ende auf Hauptbedingungen reduziert, die identisch mit bekannten semantischen Analysen sind.

Der Zusammenhang zwischen den bisher eingeführten Konditionalen wird durch folgende Verbindungen hergestellt:

(6.3.11) $A \diamond\!\!> B \in T \Leftrightarrow A > \neg B \notin T$,

(6.3.12) $A \ll\!\!\diamond\!\!\gg B \in T \Leftrightarrow \neg A \gg \neg B \in T$,

(6.3.13) $A \gg B \in T \Leftrightarrow A > B \in T \wedge \neg A \diamond\!\!> \neg B \in T$,

(6.3.14) $A \ll\!\!\diamond\!\!\gg B \in T \Leftrightarrow A \diamond\!\!> B \in T \wedge \neg A > \neg B \in T$.

Durch die Normalanalyse von weil-Sätzen sind wir auf den Starken Ramsey-Test (6.3.6) für Konditionalsätze gestoßen. Weiter unten, in den Abschnitten 6.6 und 6.7, werde ich die Normalanalyse aber zurückweisen. Außerdem argumentiere ich in den nächsten beiden Abschnitten, daß eine sorgfältige Untersuchung der Bedeutung von weil eine etwas kompliziertere Interpretation von Konditionalen erfordert als den Starken Ramsey-Test. Diese Interpretation ist — im Gegensatz zu (6.3.6) — auch ein natürliches Sprungbrett zu den Akzeptabilitätsbedingungen anderer Arten von Konditionalen und von obwohl.

6.4 Eine alternative Analyse von weil

Der Grundgedanke für meine Behandlung von weil-Sätzen ist, daß Weil A, B synonym ist mit A ist eine Erklärung für B.[7] Vorbereitend untersuchen wir die Idee, daß Weil A, B durch A ist ein Grund für B expliziert werden kann. Hingegen will ich die für weil-Sätze ebenfalls einschlägigen Begriffe der Kausalität, der Rechtfertigung, der Argumentation und der Erläuterung beiseite lassen, die alle Probleme anderer, je eigener Art aufwerfen. Gemäß der vorgegebenen Richtung wollen wir (6.2.2) und (6.2.3) diskutieren und eine allgemeine Akzeptabilitätsbedingung für weil-Sätze herausarbeiten. Um die weitere, allgemeinere Analyse nicht vorzeitig zu blockieren, werde ich in diesem Abschnitt nur notwendige Bedingungen angeben.

Beginnen wir mit *Gründen*. Wir beschäftigen uns mit Theorien — Theorien, die von gewissen Personen zu gewissen Zeitpunkten akzeptiert werden — und dürfen deshalb die Spohnschen Indices in (6.2.3) weglassen. „Epistemische Ränge" von Sätzen (oder Sachverhalten) werden durch das Enthaltensein oder Nichtenthaltensein dieser Sätze (bzw. von Sätzen, welche diese Sachverhalte ausdrücken) in den je betrachteten Theorien wi-

[7] Man könnte den Zusammenhang auch über warum-Fragen herstellen, an denen neuerdings wieder ein großes Interesse zu verzeichnen ist, vgl. Sintonen (1984), Koura (1988) und Temple (1988). Warum-Fragen sind in der Regel erklärungheischend und werden mit weil-Sätzen beantwortet.

dergespiegelt. So können wir, wenn wir das Symbol $^{a}>$ für weil beibehalten, (6.2.3) in einem ersten Versuch umschreiben zu

(6.4.1) $A^{a}>B \in T \Rightarrow (B \in T^{*}_{A} \wedge B \notin T) \vee (\neg B \notin T^{*}_{A} \wedge \neg B \in T)$.

Dies kann jedoch nicht richtig sein. Erstens kann ein Grund A Bestandteil der Theorie und trotzdem noch ein Grund sein. Aber nach (6.4.1) führt $A \in T = T^{*}_{A}$ zu $A^{a}>B \notin T$. Wir können diesen Mangel beseitigen, indem wir die Idee des Starken Ramsey-Tests (6.3.6) in Anspruch nehmen:

(6.4.2) $A^{a}>B \in T \Rightarrow (B \in T^{*}_{A} \wedge B \notin T^{*}_{\neg A}) \vee (\neg B \notin T^{*}_{A} \wedge \neg B \in T^{*}_{\neg A})$
$\Leftrightarrow A \gg B \in T \vee A \ll\gg B \in T$.[8]

Man beachte, daß sich die notwendige Bedingung für $A^{a}>B \in T$ nach (6.4.2) im Fall $A,B \in T$ (dies ist die für weil-Sätze zutreffende „Randbedingung") auf $\neg A>B \notin T$ reduziert.

Zweitens kann eine Folge B Bestandteil der Theorie und trotzdem noch eine Folge sein. Aber nach (6.4.1) führt $B \in T$ zu $A^{a}>B \notin T$. Eine naheliegende Möglichkeit, diesen Fehler von (6.4.1) auszubügeln, ist

(6.4.3) $A^{a}>B \in T \Rightarrow (B \in (T^{-}_{B})^{*}_{A} \wedge B \notin T^{-}_{B}) \vee (\neg B \notin (T^{-}_{B})^{*}_{A} \wedge \neg B \in T^{-}_{B})$.

Hier wird die Idee von (R3) benutzt. Man beachte, daß sich nach (6.4.3) im Fall $\neg B \notin T$ die notwendige Bedingung für $A^{a}>B \in T$ auf $A>B \in T^{-}_{B}$ reduziert.

Es ist leicht zu sehen, daß (6.4.2), obwohl für die Behandlung des ersten Problems vorgeschlagen, auch das zweite löst, und daß umgekehrt (6.4.3), obwohl für die Behandlung des zweiten Problems vorgeschlagen, auch das erste löst. Wir befinden uns also in der Verlegenheit, kein Kriterium dafür die Entscheidung zu haben, ob man (6.4.2) oder (6.4.3) als Analyse von weil bevorzugen soll. Glücklicherweise verhilft eine nähere Untersuchung von *Erklärungen* beiden Ideen zu ihrem Recht. Betrachten

[8] An dieser Stelle ist ein Vergleich mit dem interessant, was man David Lewis' Beitrag zur Analyse von weil nennen könnte. Mit Lewis könnte man vielleicht sagen, daß Weil A, B synonym sei mit B ist kausal abhängig von A, was wiederum folgendermaßen erklärt wäre (vgl. Lewis 1973b, S. 563, und seine Interpretation von kontrafaktischen Konditionalsätzen in 1973a):

$A^{a}>B \in T \Rightarrow A>B \in T \wedge \neg A>\neg B \in T$
$\Leftrightarrow B \in T^{*}_{A} \wedge \neg B \in T^{*}_{\neg A}$
$\Leftrightarrow A \gg B \in T \wedge A \ll\gg B \in T$.

Wenn man (6.2.3) als prinzipiell adäquat für eine Analyse von weil-Sätzen ansieht, dann ist die Explikation à la Lewis zu stark.

wir (6.2.2) und vergegenwärtigen uns, daß die „gegebene Wissenssituation" einer Erklärung die ist, wo das Explanandum gerade zur Kenntnis genommen, d.h. in unsere Theorie über die Welt aufgenommen wurde. Deshalb muß die Gärdenforssche Explikation von Erklärungen wohl so wiedergegeben werden (Gärdenfors 1980; 1988, Kap. 8; vgl. auch Stegmüller 1983, S. 957–1012):

$$(6.4.4) \qquad A^a{>}B \in T^*_B \;\Rightarrow\; (B \in T^*_A \wedge B \notin T) \vee (\neg B \notin T^*_A \wedge \neg B \in T).$$

Hier denken wir zuerst an ein bereits vertrautes Argument: Da wir die Möglichkeit $A \in T$ nicht ausschließen wollen, müssen wir T durch $T^*_{\neg A}$ ersetzen:

$$(6.4.5) \qquad A^a{>}B \in T^*_B \;\Rightarrow\; (B \in T^*_A \wedge B \notin T^*_{\neg A}) \vee (\neg B \notin T^*_A \wedge \neg B \in T^*_{\neg A})$$
$$\Leftrightarrow\; A \gg B \in T \vee A \ll\gg B \in T\,.$$

Es gibt aber ein wichtigeres Problem. Erstens ist es unschön, die Akzeptabilitätsbedingung für T^*_B und nicht für T formuliert zu haben. Zweitens wollen Gärdenfors und auch Stegmüller für T offenbar die unmittelbar zeitlich vorangehende Theorie vor dem überraschenden Auftreten des Explanandums hernehmen, d.h. die zeitlich letzte, reale Theorie vor der realen Situation T^*_B. Aber das ist eine unzulässige Vereinfachung. Man denke an einen Bomberpiloten, der sich sein Gehirn zermartert wegen der Frage, warum er damals — vor Jahrzehnten — auf den Knopf gedrückt hat — er erkennt jetzt die Normen jener Tage nicht mehr an. Astronomen können durchaus noch heute nach der Erklärung für ein sonderbares Phänomen suchen, welches aus Tycho Brahes Beobachtungen bekannt ist — nichtsdestotrotz würde keiner zum theoretischen Kontext des 16. Jahrhunderts zurückgehen wollen. Wissenschaftliche Revolutionen erzeugen oft das Bedürfnis nach einer Erklärung von Fakten, die bis dahin als selbstverständlich gegolten hatte.[9] Der Schritt von T^*_B zurück zu T kann also nicht immer durch den simplen Rückzug auf die Vorgängertheorie getan werden. Umso besser, daß wir mit diesem Problem bereits umgehen

[9] Vergleiche zum Beispiel Kuhn: „Through his [Hauksbee's] researches ... repulsion suddenly became *the* fundamental manifestation of electrification, and it was then attraction that needed to be *explained*." (1970, S. 117f), „Lavoisier's reform ... ended by depriving chemistry of some actual and much potential *explanatory power*." (1970, S. 107) und etwas allgemeiner: „Changes in the standards governing permissible problems, concepts, and *explanations* can transform a science." (1970, S. 106) (die letzten drei Hervorhebungen von mir)

können, und zwar durch die Verwendung von Theorienkontraktionen:[10]

$$(6.4.6) \quad A^{a}{>}B \in T \Rightarrow (B \in (T^{-}{}_{B})^{*}{}_{A} \wedge B \notin (T^{-}{}_{B})^{*}{}_{\neg A}) \vee (\neg B \notin (T^{-}{}_{B})^{*}{}_{A} \wedge \neg B \in (T^{-}{}_{B})^{*}{}_{\neg A}) \Leftrightarrow A \gg B \in T^{-}{}_{B} \vee A \ll\gg B \in T^{-}{}_{B} .$$

Ich meine nun, daß das die richtige Analyse von weil-Sätzen ist.[11] Man vergegenwärtige sich, daß (6.4.6) durch die Kombination der Ideen von (6.4.2) und (6.4.3) dafür sorgt, daß das Akzeptieren von A und/oder B weder das Akzeptieren oder das Zurückweisen von Weil A, B präjudiziert.[12]

6.5 Universelle Konditionale

Die Idee der Parallelität von Konditionalsätzen und Erklärungen wiederaufgreifend, wollen wir nun (6.4.6) als zentrale Bedingung einer Analyse des gesamten natürlichsprachlichen Bereichs zugrundelegen, der durch verschiedene wenn-dann-Kombinationen, durch weil und obwohl abgesteckt ist. Ich werde eine Basis von *universellen Konditionalen* postulieren, die nicht in der natürlichen Sprache realisiert sind. Für dieses Konzept, das sozusagen im Geiste, aber nicht im Wörterbuch zu finden ist, werde ich den Pfeil $\rightarrow$ anstelle von $>$ verwenden.[13]

Die Bedingung (6.4.6) ist kann direkt zur allgemeinen Charakterisierung einer „konditionalen Verknüpfung" dienen. Zum Zwecke der Einführung eines *universellen (Pro-)Konditionals* $\rightarrow$ kann die Implikation (6.4.6) zu einer Äquivalenz verstärkt werden:

$$(6.5.1) \quad A \rightarrow B \in T \Leftrightarrow A \gg B \in T^{-}{}_{B} \vee A \ll\gg B \in T^{-}{}_{B} .$$[14]

[10] Gärdenfors erwähnt den Gebrauch von Kontraktionen in Erklärungskontexten in (1980, S. 412; 1981, S. 205; 1982, Fußnote 4; 1984, S. 139). Erst später, in Gärdenfors (1988, Kapitel 8), spielen Kontraktionen eine entscheidende Rolle.

[11] Welche allerdings noch durch die Angabe der Randbedingungen A,B$\in$T zu ergänzen ist, siehe die Abschnitte 6.6 und 6.7.

[12] Die in diesem Abschnitt verfolgte Strategie, Erklärungen anstelle von Gründen zu betrachten und (6.4.6) anstelle der einfacheren Methoden (6.4.2) oder (6.4.3) auszuwählen, ist letztlich noch nicht zwingend motiviert. Es sei deshalb schon hier angemerkt, daß (6.4.2) und (6.4.3) keine allgemeine Analyse entlang der Richtlinien von Abschnitt 6.5 und 6.6 ermöglichen. Zum Beispiel würde (6.4.2) kein Akzeptieren von Obwohl A, B (des Kontrakonditionals mit A,B$\in$T) zulassen, und auf der Grundlage von (6.4.3) wären (6.5.2) und (6.5.3) unten nicht mehr äquivalent und indikativische wenn-dann möglicherweise-Sätze ganz ausgeschlossen.

[13] Man beachte im folgenden die streng auseinanderzuhaltenden Notationen: $\rightarrow$ ist das materiale Konditional, $\rightarrow$ das universelle (Pro-)Konditional.

[14] Man könnte noch weiter verallgemeinern, indem man das symmetrische

In den folgenden Diskussionen werde ich das erste Glied der Disjunktion auf der rechten Seite die $\Box$-*Version von* $\succ$ und das zweite Glied die $\Diamond$-*Version von* $\succ$ nennen.

Um Konnektive wie obwohl interpretieren zu können, werden wir ein Bild vom „Gegenteil" einer konditionalen Verknüpfung brauchen, d.h. eine Relation des Entgegenwirkens. Es gibt zwei naheliegende Möglichkeiten für die Definition des *universellen Kontrakonditionals* $\prec$:

(6.5.2) $A \prec B \in T \Leftrightarrow \neg A \gg B \in T^-{}_B \vee \neg A \ll\gg B \in T^-{}_B$.

(6.5.3) $A \prec B \in T \Leftrightarrow A \gg \neg B \in T^-{}_B \vee A \ll\gg \neg B \in T^-{}_B$.

Ich wüßte kein Kriterium, um zu entscheiden, welche dieser beiden Akzeptabilitätsbedingungen die bessere ist. Glücklicherweise sind (6.5.2) und (6.5.3) äquivalent, wovon man sich leicht (durch Anwenden der Definitionen (6.3.6) und (6.3.10)) überzeugt. Die jeweiligen $\Box$-Versionen und $\Diamond$-Versionen tauschen jedoch ihre Rollen. Per Konvention wähle ich (6.5.2) als die Standardformulierung, auf die ich mich beziehe, wenn ich später von den $\Box$-Versionen und $\Diamond$-Versionen von $\prec$ spreche.[15] Wir bemerken noch, daß sich die in (6.5.1) und (6.5.2) vorkommenden Bedingungen $A \gg B \in T^-{}_B$ und $\neg A \gg B \in T^-{}_B$ wegen (T-4) und (6.3.8) auf $A > B \in T^-{}_B$ bzw. $\neg A > B \in T^-{}_B$ reduzieren. Wie intuitiv wünschenswert, können $A \succ B$ und $A \prec B$ nicht gleichzeitig in einer Theorie T enthalten sein — es ist unmöglich, daß A zugleich positiv und negativ relevant für B ist. Natürlich gibt es aber eine dritte Möglichkeit: A braucht ja B weder zu fördern noch zu hemmen, A und B können schlicht voneinander unabhängig sein. Dieser Sachlage soll das *universelle Nichtkonditional* $\multimap$ Rechnung tragen, welches so formuliert werden kann:

(6.5.4) $A \multimap B \in T \Leftrightarrow A \succ B \notin T \wedge A \prec B \notin T$.

Damit ist die disjunkte und vollständige Menge der universellen Konditionale, die — so meine ich — der Vielfalt der tatsächlich realisierten natürlichsprachlichen Konditionale zugrunde liegen, vollständig.

$(T^-{}_B)^-{}_{\neg B}$ an die Stelle von $T^-{}_B$ in (6.5.1) einsetzt. Unter Zugrundelegung meiner Annahmen über natürlichsprachliche Konditionale (siehe Abschnitt 6.6) würde sich ein Unterschied nur hinsichtlich kontrafaktischer Konditionalsätze ergeben. Wie sich herausstellen wird, ist es (6.5.1), welches in befriedigender Weise mit den etablierten Interpretationen kontrafaktischer Konditionalsätze übereinstimmt.

[15] Die Unterscheidung von $\Box$-Versionen und $\Diamond$-Versionen bei Kontrakonditionalen wird sich in Abschnitt 6.7 am Ende als überflüssig herausstellen. Dies erklärt vielleicht noch einmal die Unsicherheit, ob man (6.5.2) und (6.5.3) nehmen soll.

6.6 Natürlichsprachliche Konditionale: Thesen

In die Menge der natürlichsprachlichen Konditionale möchte ich einschließen: wenn-dann, wenn-dann möglicherweise und auch wenn, jeweils mit Indikativ und mit Konjunktiv, obwohl und weil (beide nur mit Indikativ).[16] Ich beziehe mich im folgenden mit dem Ausdruck „Antezedens" auf den Nebensatz und mit dem Ausdruck „Konsequens" auf den Hauptsatz der entsprechenden Satzgefüge.

Meine Thesen über natürlichsprachliche Konditionale lauten folgendermaßen:

(6.6.1) Natürlichsprachliche Konditionalsätze gehorchen Akzeptabilitätsbedingungen, die auf den Wandel von Theorien Bezug nehmen.

(6.6.2) Die Akzeptabilitätsbedingung eines natürlichsprachlichen Konditionalsatzes setzt sich zusammen aus einer Bedingung für das Akzeptieren eines universellen Konditionals („Hauptbedingung") und einer Bedingung für das Akzeptieren von Antezedens und Konsequens („Randbedingung").

(6.6.3) Wenn und weil sind Realisierungen des universellen Prokonditionals; obwohl ist eine Realisierung des universellen Kontrakonditionals; auch wenn ist eine Realisierung des universellen Nichtkonditionals.[17]

(6.6.4) Realisierungen des universellen Prokonditionals werden nur dann akzeptiert, wenn der Akzeptanzstatus des Antezedens und des Konsequens gleich sind; Realisierungen des universellen Kontrakonditionals und Nichtkonditionals werden nur dann akzeptiert, wenn das Konsequens akzeptiert wird.

(6.6.5) Weil und obwohl werden nur dann akzeptiert, wenn das Antezedens akzeptiert wird; wenn-dann, wenn-dann möglicherweise und auch wenn werden nur dann akzeptiert, wenn das Antezedens nicht akzeptiert wird, wobei es der grammatische Verbmodus ist, welcher den genauen Akzeptanzstatus des Antezedens anzeigt.[18]

---

[16] Vielleicht könnte man auch Konjunktionen wie aber und deshalb mit einbeziehen, wobei sich die erstere zu obwohl verhält wie die letztere zu weil.

[17] Ich kenne keine Konjunktion, die als ein Nichtkonditional zu verstehen ist und bei der Antezedens und Konsequens akzeptiert sind. Betreffs der naheliegenden Vermutung, daß auch wenn eigentlich ein Kontrakonditional sein sollte, siehe Abschnitt 6.7.

[18] Diese verallgemeinernde Aussage vernachlässigt die Tatsache, daß auch wenn machmal wie wenn auch (mit akzeptierten Antezedens) zu deuten ist und umgekehrt. Ich

Es versteht sich, daß diese Thesen sehr stark abgefaßt sind und der Variabilität der natürlichen Sprache letztendlich nicht gerecht werden. Dennoch hoffe ich, im folgenden durch die engen Restriktionen interessante Aussagen über einen logischen Kernbereich natürlichsprachlicher Konditionalsätze machen zu können.

6.7 Natürlichsprachliche Konditionale: Akzeptabilitätsbedingungen

Der Einfachheit und Kürze willen sind (6.6.1)–(6.6.5) nicht als explizite Analyse einzelner natürlichsprachlicher Konditionale formuliert. Die Thesen bedürfen der Rechtfertigung. Ich werde versuchen, sie zu stützen, indem ich die nacheinander Akzeptabilitätsbedingungen angebe und ihre Konsequenzen prüfe.

Beginnen wir mit dem Konditional im engsten Sinne, mit wenn-dann. Ich habe hier drei ziemlich kontroverse Thesen zu verteidigen. Zuerst zu (6.6.5): Warum sollen wir nicht gleichzeitig A und Wenn A, dann B akzeptieren? Es erscheint zwingend, daß wir in dieser Situation auch B akzeptieren müssen. Meine Antwort lautet dann so: entweder wir nehmen eine „konditionale Verknüpfung" zwischen A und B an — dann müssen wir Weil A, B akzeptieren (da wir A und B in der Tat akzeptieren) —, oder A ist irrelevant für B — dann sollten wir nicht mehr als die Konjunktion A und B akzeptieren —, oder A wirkt B sogar entgegen — dann sollten wir allein Obwohl A, B gelten lassen. Wenn A, dann B ist nach meinem Sprachgefühl jedoch nicht akzeptabel, denn dies würde implizieren, daß wir A nicht in unserer Theorie enthalten haben und daß wir eine positive Verknüpfung zwischen A und B unterstellen. Es sei darauf aufmerksam gemacht, daß die in (6.6.5) markierten Unterschiede im Akzeptanzstatus des Antezedens das Argument bilden, dem die Normalanalyse von weil zum

nehme in Kauf, die funktionellen Unterschiede besonders zwischen Kontrakonditionalen und Nichtkonditionalen in gewissem Grade zu überbetonen. Daß auch wenn synonym durch wenn ... trotzdem (trotzdem im Hauptsatz) ersetzt werden kann, deutet auf die Möglichkeit grundsätzlicher Probleme bei der Behandlung von auch wenn als lexikalische Einheit hin. Ich unterdrücke hier auch die wohlbekannte Beobachtung, daß der grammatische Verbmodus irreführend sein kann, besonders wenn mit dem Konditional Aussagen über zukünftige Ereignisse gemacht werden sollen; vgl. z.B. für das Englische Thomson und Martinet (1980, S.188): „Sometimes, rather confusingly, type 2 [konjunktivische Konditionalsätze] can be used as an alternative to type 1 [indikativische Konditionalsätze] for perfectly possible plans and suggestions."

Opfer fällt.

Zweitens, noch einmal zu (6.6.5): Hat nicht Adams' berühmtes Kennedy-Beispiel ein für alle Mal gezeigt, daß eine gemeinsame Analyse von indikativischen und konjunktivischen Konditionalsätzen undurchführbar ist? Denn, so behaupten Adams und viele andere, wir akzeptieren doch alle

(6.7.1) Wenn Oswald Kennedy nicht getötet hat, dann war es jemand anderes.

und gleichzeitig

(6.7.2) Wenn Oswald Kennedy nicht getötet hätte, dann könnte Kennedy immer noch am Leben sein.

Und es ist offensichtlich unvereinbar, einerseits einen unbekannten Mörder und andererseits einen lebenden Kennedy anzunehmen. Aber gerade das simultane Akzeptieren von (6.7.1) und (6.7.2) möchte ich bestreiten. Wer (6.7.1) akzeptiert, der glaubt nicht, daß er den Mörder kennt. Wer aber (6.7.2) akzeptiert, der glaubt das schon: Nach seiner Ansicht war es Oswald. Und natürlich schließen sich diese beiden kriminalistischen „Theorien" aus. Nur unserer Unsicherheit, ob wir denn wissen sollten, daß Oswald Kennedy getötet hat, verdankt das Beispiel seine Überzeugungskraft: (6.7.1) induziert einen Kontext, in dem wir — wie der Polizeiinspektor am Beginn seiner Ermittlungen — den Attentäter nicht kennen, während (6.7.2) uns unvermittelt in die Situation versetzt, in der wir uns der Identität des Täters sicher zu sein glauben.[19]

Was drittens (6.6.4) anbetrifft, gibt es nicht Gegenbeispiele zu der Behauptung, daß das Antezedens und das Konsequens eines wenn-dann-Satzes vom gleichen Akzeptanzstatus sein müssen? Erinnern wir uns an Berta, die

[19] Ähnliche Beispiele sind schon in Ramsey (1931, S. 249) und Mackie (1962, S. 71) zu finden (vgl. Kapitel 4, Fußnote 10). Der durch solche Beispielspaare induzierte Prozeß ist wohl als ein Musterbeispiel dessen zu betrachten, was Lewis (1979b) „conversational scorekeeping" nennt. Obgleich ich glaube, daß die im Text gegebene Antwort ausreichend ist, um den Einwand gegen (6.6.5) zu entkräften, bin ich nach einer Diskussion mit Wolfgang Spohn zu der Überzeugung gekommen, daß es tatsächlich einen weiteren Unterschied zwischen indikativischen und konjunktivischen Konditionalsätzen gibt. (Dies ist nicht Spohns Ansicht.) Wie in Abschnitt 4.6.1, angedeutet, denke ich, daß beide derselben Form des (Starken) Ramsey-Tests gehorchen, daß sich die Revisionsmethoden jedoch unterscheiden. Beim indikativischen Konditional hat man sich vorzustellen, daß man das Antezedens *wirklich* als ein *neues, unbestreitbares Stück Information* bekommt; die Revision basiert dann wohl auf einer Relation der theoretischen Wichtigkeit, die Bestätigungsgrade widerspiegelt. Beim konjunktivischen Fall hingegen muß man das Antezedens *nur als Hypothese* annehmen, wobei in die theoretische Wichtigkeit Betrachtungen der von Lewis (1979a) diskutierten Art eingehen.

Anton nicht mehr mag. Der kritische Fall ist der folgende. Angenommen, wir wissen, daß Anton zur Party kommt, aber wir haben keine Ahnung, ob Berta auftauchen wird. Scheinbar können wir dann akzeptieren:

> Wenn Anton zu Hause bliebe, dann würde Berta zur Party gehen.

Ich halte es aber für sehr gut möglich, daß jemand gegen einen solchen Satz protestiert: „Nun, es ist doch überhaupt nicht ausgemacht, daß Berta nicht zur Party kommt!" Daraufhin müssen wir präziser formulieren, was wir eigentlich meinen:

> Wenn Anton zu Hause bliebe, dann würde Berta ganz sicher zur Party gehen.

Und damit haben wir ein Konsequens, dessen Negation wir in der Tat akzeptieren.[20]

Als Resultat der Thesen (6.6.3)–(6.6.5) ist wenn-dann (egal, mit welchem Verbmodus) das einzige natürlichsprachliche Konditional, welches ein nicht akzeptiertes Konsequens hat. Dies bringt es mit sich, daß wenn-dann das einzige natürlichsprachliche Konditional mit einer separat zu analysierenden $\Diamond$-Version ist. In allen anderen Fällen wäre es witzlos, einen möglicherweise-Operator auf das Konsequens anzuwenden — man akzeptiert es ja bereits. Sollte hier einmal tatsächlich ein möglicherweise im Konsequens zu finden sein, dann ist es als integraler Bestandteil dieses Konsequens aufzufassen.[21]

Wir sind jetzt in der Lage, die Akzeptabilitätsbedingungen von indikativischen und konjunktivischen Konditionalsätzen explizit anzugeben. Im folgenden benutze ich Varianten von $\rightarrow$ anstelle von $>$, um den Einschnitt zwischen der auf universellen Konditionalen basierenden Analyse und der vorläufigen Analyse aus den Abschnitten 6.3 und 6.4 kenntlich zu machen. Da die $\Diamond$-Versionen explizit angeschrieben werden, muß noch eine Zwischenebene eingezogen werden, bevor wir auf die natürlichsprachlichen Konnektive kommen:

[20] Die These (6.6.4) ist in der Anwendung auf Prokonditionale zugegebenermaßen ziemlich dogmatisch. Deshalb sind im Überblickschema am Ende von Abschnitt 6.7 sicherheitshalber die von (6.6.4) verbotenen Kombinationen mit aufgeführt. Vgl. auch die Diskussion des Beispiels eines „semifaktischen" Konditionalsatzes am Ende von Abschnitt 6.8.

[21] Vgl. aber den Versuch von Abschnitt 4.5, mit Lewis's would-be-possible-Lesart (WBP) alle wenn-dann möglicherweise-Sätze als Konditionalsätze mit einem modalisierten Konsequens aufzufassen.

(6.7.3) $A^{\circ}\!\!\rightarrow\!\!\!\!\succ B \in T \Leftrightarrow A\text{-}\!\!\rightarrow\!\!\!\!\succ B \in T \wedge A,\neg A,B,\neg B \notin T$.

(6.7.4) $A^{z}\!\!\rightarrow\!\!\!\!\succ B \in T \Leftrightarrow A\rightarrow\!\!\!\!\succ B \in T \wedge \neg A,\neg B \in T$.

Nun erreicht man das natürlichsprachliche wenn-dann, indem man die $\Box$-Version von $\rightarrow\!\!\!\!\succ$, und das natürlichsprachliche wenn-dann möglicherweise, indem man die $\Diamond$-Version von $\rightarrow\!\!\!\!\succ$ nimmt. $^{\circ}\!\!\rightarrow\!\!\!\!\succ$ bezeichnet den indikativischen und $^{z}\!\!\rightarrow\!\!\!\!\succ$ den konjunktivischen Fall:

(6.7.5) $A^{\circ}\Box\!\!\rightarrow B \in T \Leftrightarrow A \gg B \in T^{-}{}_{B} \wedge A,\neg A,B,\neg B \notin T$

$\Leftrightarrow A \rightarrow B \in T \wedge A,\neg A,B,\neg B \notin T$.

(6.7.6) $A^{\circ}\Diamond\!\!\rightarrow B \in T \Leftrightarrow A \ll\gg B \in T^{-}{}_{B} \wedge A,\neg A,B,\neg B \notin T$

$\Leftrightarrow \neg A \rightarrow \neg B \in T \wedge A,\neg A,B,\neg B \notin T$.

(6.7.7) $A^{z}\Box\!\!\rightarrow B \in T \Leftrightarrow A \gg B \in T^{-}{}_{B} \wedge \neg A,\neg B \in T$

$\Leftrightarrow B \in T^{*}{}_{A} \wedge \neg A,\neg B \in T$.

(6.7.8) $A^{z}\Diamond\!\!\rightarrow B \in T \Leftrightarrow A \ll\gg B \in T^{-}{}_{B} \wedge \neg A,\neg B \in T$

$\Leftrightarrow \neg B \notin T^{*}{}_{A} \wedge \neg A,\neg B \in T$.

Man kann leicht zeigen, daß sich die oberen Zeilen von (6.7.5)–(6.7.8) auf die endgültigen unteren Zeilen reduzieren (man verwendet insbesondere (T–3), (T*3), (T*4), (T*S) und die Tatsache, daß Theorien gegenüber tautologischer Folgerung abgeschlossen sind). Ich möchte diese Resultate hier nicht diskutieren; für (6.7.5), (6.7.7) und (6.7.8) kann man eine weitgehende Rechtfertigung in Gärdenfors (1981) finden.[22]

Es bleibt nur noch ein Prokonditional, das wir uns ansehen müssen, und das ist weil. Dem obigen Argument folgend, müssen wir hier die $\Box$-Version und die $\Diamond$-Version nicht auseinanderdividieren und können von $\rightarrow\!\!\!\!\succ$ in einem Schritt zum natürlichsprachlichen $^{a}\!\!\rightarrow\!\!\!\!\succ$ übergehen:

(6.7.9) $A^{a}\!\!\rightarrow\!\!\!\!\succ B \in T \Leftrightarrow A\rightarrow\!\!\!\!\succ B \in T \wedge A,B \in T$

$\Leftrightarrow (B \in (T^{-}{}_{B})^{*}{}_{A} \vee \neg B \in (T^{-}{}_{B})^{*}{}_{\neg A}) \wedge$
$A,B \in T.$

Die Vereinfachung geschieht mit Lemma 3.1.6 und (T–2). Es ist lehrreich, eine Fallunterscheidung bzgl. des Beibehaltens von A in $T^{-}{}_{B}$ durchzuführen. Wenn $A \in T^{-}{}_{B}$, sieht man sofort, daß die $\Box$-Version unmöglich ist (wegen (T*S) und (T–4)), man muß also nur die $\Diamond$-Version betrachten. Wenn $A \notin T^{-}{}_{B}$, ist andererseits die $\Diamond$-Version stets erfüllt (wegen (T*3), (T*4) und der Beziehung $\neg A \rightarrow \neg B \in T^{-}{}_{B} = T \cap T^{*}{}_{\neg B}$, die aus $A \in T$ und $\neg B \in T^{*}{}_{\neg B}$ folgt); deshalb sind weil-Sätze im Fall $A \notin T^{-}{}_{B}$ nur dann nichttrivial, wenn ihre $\Box$-Version intendiert ist, deren Akzeptabilitätsbe-

[22] Allerdings ist (6.7.6) wohl viel eher eine Behauptbarkeitsbedingung als eine Akzeptabilitätsbedingung, da nach dieser Bedingung zum Beispiel $A^{\circ}\Diamond\!\!\rightarrow B \in T$ und $A^{\circ}\Diamond\!\!\rightarrow \neg B \in T$ nicht gleichzeitig gelten kann. (Diese Beobachtung verdanke ich Peter Gärdenfors.)

dingung nun umgeformt werden kann zu $A \rightarrow B \in T^-_B$ (mit (T-2), (T*3) und (T*4)), oder äquivalent zu $\neg A \in T^*_{\neg B}$ (Kontraposition der materialen Implikation, Definition 3.1.2 und Levis These (L)).

Wir halten für einen Moment inne und sehen nach, ob unsere Analyse richtige Voraussagen liefert. Rufen wir uns wieder Anton und Berta in Gedächtnis, nehmen diesmal als gesichert an, daß beide zur Party gehen werden und daß auch Christoph da sein wird (Berta hat neuerdings ein Auge auf Christoph geworfen). Wir akzeptieren

> Es ist nicht der Fall, daß Berta zur Party geht, weil Anton zur Party geht.

Es dürfte unproblematisch sein, diesen Satz durch $\neg(A \,{}^{a}\!\!\rightarrow\!\!\succ B) \in T$ zu symbolisieren. Das impliziert für konsistentes T $A \,{}^{a}\!\!\rightarrow\!\!\succ B \notin T$, d.h. nach (6.7.9) müßte gelten:

(6.7.10) $(B \notin (T^-_B)^*_A \vee B \in (T^-_B)^*_{\neg A}) \wedge$
$\wedge (\neg B \notin (T^-_B)^*_{\neg A} \vee \neg B \in (T^-_B)^*_A)$.

Wir wollen die intuitive Adäquatheit dieser theoretischen Voraussage prüfen. Angenommen, wir wüßten nicht, daß sich Berta entschlossen hat, auf das Fest zu gehen. Würden wir deshalb unsere Überzeugung, daß Anton zur Party geht, fallen lassen? Sicherlich nicht, denn Antons Kommen ist bestenfalls irrelevant, wahrscheinlich aber sogar hinderlich für Bertas Zusage, das heißt Antons Kommen paßt sogar sehr gut zur Infragestellung von Bertas Kommen. Deshalb gilt $A \in T^-_B$, wegen (T*S) also $(T^-_B)^*_A = T^-_B$, und der erste Term von (6.7.10) gilt wegen (T-4). Der letzte Term hingegen widerspricht wegen (T-2) $\neg B \notin T \supseteq T^-_B$. Wir haben also zu zeigen, daß $\neg B \notin (T^-_B)^*_{\neg A}$. Aber wenn wir von unserer Unwissenheit bzgl. Bertas Plänen ausgehen, dann wird uns die Annahme, daß der ungeliebte Anton weg bleibt, auf keinen Fall zum Resultat führen, daß Berta auch weg bleibt, und wir haben gezeigt, daß (6.7.10) intuitiv tatsächlich erfüllt ist.

Man vergleiche dies mit dem wirklichen Grund für Bertas Anwesenheit:

> Weil Christoph zur Party geht, geht Berta zur Party.

Diese Antithese zum letzten Satz unseres Partygeplauders versteht man normalerweise so, daß das Antezedens das Konsequens irgendwie erzwinge. Zweifel an Bertas Anwesenheit hätten auch Zweifel an Christophs Anwesenheit zur Folge, und im Gegensatz zum Fall des armen Anton haben wir intuitiv $C \notin T^-_B$. Und das ist genau das, was (6.7.9) für die $\Box$-Version von $\,{}^{a}\!\!\rightarrow\!\!\succ$ vorausgesagt hat.

Wenden wir uns nun den Kontrakonditionalen zu. Wir bleiben beim alten Beispiel und nehmen an, daß Berta im allgemeinen sehr darauf bedacht ist, keinem enttäuschten Verehrer über den Weg zu laufen. Wir akzeptieren also

Obwohl Anton zur Party geht, geht Berta zur Party.

Wieder erwarten wir ganz bestimmt $A \in T^-_B$ — warum eine Überzeugung aus einer Kontraktion ausschließen, wenn genau diese Überzeugung die Kontraktion plausibler macht? Unsere Analyse sagt dies, wie wir gleich sehen werden, auch vorher. Wenn wir $\overset{a}{\dashv\!\prec}$ als Symbol für obwohl verwenden, dann sieht die Analyse gemäß (6.6.3)–(6.6.5) folgendermaßen aus:

$$\begin{aligned}(6.7.11)\quad A \overset{a}{\dashv\!\prec} B \in T \quad &\Leftrightarrow \quad A \dashv\!\prec B \in T \wedge A,B \in T \\ &\Leftrightarrow \quad (B \in (T^-_B)^*_{\neg A} \vee \neg B \in (T^-_B)^*_A) \wedge \\ &\qquad\qquad A,B \in T\,.\end{aligned}$$

Die Vereinfachung geschieht wieder mit Lemma 3.1.6 und (T–2). Angenommen nun, daß $A \notin T^-_B$. Dann läuft die $\Box$-Version auf $\neg A \rightarrow B \in T^-_B \subseteq T^*_{\neg B}$ hinaus; da $\neg B \in T^*_{\neg B}$, folgt $A \in T^*_{\neg B}$ und ebenso $A \in T^-_B = T \cap T^*_{\neg B}$, im Widerspruch zur Annahme. Die $\Diamond$-Version impliziert $\neg B \notin (T^-_B)^*_{\neg A}$ (andernfalls wäre nach 3.1.6 $\neg B \in T^-_B \subseteq T$, was mit $B \in T \neq T_\perp$ unverträglich ist). Gemäß unserer Annahme ist dies gleichwertig mit $\neg B \notin (T^-_B)^+_{\neg A}$, d.h. $\neg A \rightarrow \neg B \notin T^-_B$. Jetzt ist aber $A \in T$ und $\neg B \in T^-_B$, also nach der Harper-Identität $A \vee \neg B \in T \cap T^*_{\neg B} = T^-_B$, und wir haben einen Widerspruch. Damit haben wir das intuitiv gewünschte Resultat $A \in T^-_B$ bekommen. Aber nun sehen wir sofort, daß die $\Diamond$-Version von $\overset{a}{\dashv\!\prec}$ sich sozusagen selbst widerlegt: Sie reduziert sich auf $\neg B \in T^-_B$, was nach (T–2) und der Randbedingung $B \in T$ verboten ist. Also ist die einzige Hauptbedingung für obwohl $B \in (T^-_B)^*_{\neg A}$.

Man möchte meinen, daß auch wenn das Kontrakonditional für den Fall eines nicht akzeptierten Antezedens ist. Es stellt sich aber heraus, daß diese Vermutung trügt. Auf der Grundlage der vorangehenden Thesen und Vereinfachungen wird man keine Mühe haben, die folgenden Bedingungen nachzuprüfen. Ich benutze $\overset{\circ}{\dashv\!\prec}$ bzw. $\overset{z}{\dashv\!\prec}$ als Kandidaten für die Explikation des indikativischen und konjunktivischen auch wenn:

$$\begin{aligned}(6.7.12)\quad A \overset{\circ}{\dashv\!\prec} B \in T \quad &\Leftrightarrow \quad A \dashv\!\prec B \in T \wedge A,\neg A \notin T \wedge B \in T \\ &\Leftrightarrow \quad \neg A \rightarrow B \in T^-_B \wedge A,\neg A \notin T \wedge B \in T \\ &\Leftrightarrow \quad A \in T^*_{\neg B} \wedge A,\neg A \notin T \wedge B \in T\,.\end{aligned}$$

$$(6.7.13)\quad A^{z}\!\!-\!\!\prec\! B \in T \;\Leftrightarrow\; A\!-\!\!\prec\! B \in T \wedge \neg A,B \in T \;\Leftrightarrow\; (B \in (T^{-}{}_{B})^{*}{}_{\neg A} \vee \neg B \in (T^{-}{}_{B})^{*}{}_{A}) \wedge \neg A,B \in T\,.$$

Einige Kommentare sind angebracht. Bei $^{\circ}\!\!-\!\!\prec$ führt die $\Diamond$-Version zu einem Widerspruch, so daß die $\Box$-Version übrig bleibt; während die zweite Zeile von (6.7.12) ziemlich plausibel aussieht, ist die dritte, tatsächlich äquivalente Zeile intuitiv zu stark. Stellen wir uns vor, wir wissen nicht, ob Anton, wir wissen aber genau, daß Berta zur Party geht, und wir akzeptieren

> Auch wenn Anton zur Party geht, geht Berta zur Party.

Unter der hypothetischen Annahme, daß Berta nicht kommt, würde wir unsere Unwissenheit hinsichtlich Anton nicht gegen die Überzeugung eintauschen, daß er die Party besucht. Denn der auch wenn-Satz sagt ja gerade, daß Antons Anwesenheit Berta nicht davon abhalten könnte, zur Party zu gehen, also kann es nicht an Anton liegen, daß Berta nicht kommt.

Bei $^{z}\!\!-\!\!\prec$ liegt eine perfekte Symmetrie zur Situation von weil vor. Falls $\neg A \in T^{-}{}_{B}$, erkennen wir sofort, daß die $\Box$-Version unmöglich ist, und es bleibt die Bedingung für die $\Diamond$-Version. Falls $\neg A \notin T^{-}{}_{B}$, ist die $\Diamond$-Version offensichtlich erfüllt, und die $\Box$-Version reduziert sich auf $\neg A \rightarrow B \in T^{-}{}_{B}$, oder äquivalent auf $A \in T^{*}{}_{\neg B}$. Tatsächlich ergibt ein Vergleich von (6.7.13) mit (6.7.9), daß $A^{z}\!\!-\!\!\prec\! B \in T$ äquivalent mit $\neg A^{a}\!\!\rightarrow\! B \in T$ ist. Dies zeigt endgültig, daß ohne weitere Verfeinerungen $^{z}\!\!-\!\!\prec$ nicht einmal eine partielle Explikation des konjunktivischen auch wenn sein kann:

> Auch wenn Anton zur Party gehen würde, würde Berta zur Party gehen.

und

> Weil Anton nicht zur Party geht, geht Berta zur Party.

sind keineswegs synonym, sondern scheinen im Gegenteil einander ausschließende Akzeptabilitätsbedingungen zu haben.

Was können wir aus dieser Situation lernen? Es scheint, daß Kontrakonditionale nicht das einzige Mittel sind, eine entgegenwirkende Verknüpfung zwischen A und B mitzuteilen. Das natürlichsprachliche auch wenn ist, wenn es auch häufig eine solche entgegenwirkende Verknüpfung anzeigt, kein Kontrakonditional, denn die im Antezedens beschriebenen „widrigen Umstände" sind eben nicht wirksam genug, um das Konsequens aus der

Theorie zu eliminieren. Das Konsequens ist *unbedingt* akzeptiert, und auch wenn kann man demgemäß ein Nichtkonditional nennen. Zur Vorbereitung der Analyse von Nichtkonditionalen geben wir der Grundidee (6.5.4) eine handlichere Form:

$$\begin{aligned}(6.7.14)\quad A\text{—}\!\shortmid\! B \in T \;&\Leftrightarrow\; A\text{—}\!\succ\! B \notin T \wedge A\text{—}\!\prec\! B \notin T\\ &\Leftrightarrow\; (B \in (T^{-}{}_{B})^{*}{}_{A} \Leftrightarrow B \in (T^{-}{}_{B})^{*}{}_{\neg A}) \wedge\\ &\qquad (\neg B \in (T^{-}{}_{B})^{*}{}_{A} \Leftrightarrow \neg B \in (T^{-}{}_{B})^{*}{}_{\neg A})\\ &\Leftrightarrow\; B \notin (T^{-}{}_{B})^{*}{}_{A} \wedge B \notin (T^{-}{}_{B})^{*}{}_{\neg A} \wedge\\ &\qquad (\neg B \in (T^{-}{}_{B})^{*}{}_{A} \Leftrightarrow \neg B \in (T^{-}{}_{B})^{*}{}_{\neg A})\ .\end{aligned}$$

Die Thesen (6.6.3)–(6.6.5) liefern schließlich die folgenden Akzeptabilitätsbedingungen für indikativisches und konjunktivisches auch wenn, die wir durch $^{\circ}$—$\shortmid$ bzw. durch ^{z}—$\shortmid$ symbolisieren:

$$\begin{aligned}(6.7.15)\quad A^{\circ}\text{—}\!\shortmid\! B \in T \;&\Leftrightarrow\; A\text{—}\!\shortmid\! B \in T \wedge A,\neg A \notin T \wedge B \in T\\ &\Leftrightarrow\; A\text{—}\!\!>\! B,\neg A\text{—}\!\!>\! B \notin T^{-}{}_{B} \wedge A,\neg A \notin T \wedge\\ &\qquad B \in T\\ &\Leftrightarrow\; A,\neg A \notin T^{*}{}_{\neg B} \wedge A,\neg A \notin T \wedge B \in T\ .\\ (6.7.16)\quad A^{z}\text{—}\!\shortmid\! B \in T \;&\Leftrightarrow\; A\text{—}\!\shortmid\! B \in T \wedge \neg A,B \in T\\ &\Leftrightarrow\; B,\neg B \notin (T^{-}{}_{B})^{*}{}_{A} \wedge \neg A,B \in T\ .\end{aligned}$$

Man beachte, daß sich wegen der gemeinsamen Randbedingung $B \in T$ $A\text{—}\!\shortmid\! B$ in (6.7.15) und (6.7.16) auf $B,\neg B \notin (T^{-}{}_{B})^{*}{}_{A}, (T^{-}{}_{B})^{*}{}_{\neg A}$ reduziert. In (6.7.15) wird die Vereinfachung durch (T*3), (T*4) und Definition 3.1.2, die Kontraposition des materialen Konditionals, (T–5) und die Levi-Identität herbeigeführt. In (6.7.16) prüfe man nach, daß $\neg A \notin T^{-}{}_{B}$ zu einem Widerspruch führt. Die Folgerung $\neg A \in T^{-}{}_{B}$ ist, wenn auch nicht völlig unplausibel, sicherlich eine sehr starke Bedingung. Ich habe den Verdacht, daß wir uns zur Analyse von auch wenn reumütig wieder der Untersuchung von (6.7.13) werden zuwenden müssen, wobei die absurde Äquivalenz zu einem weil-Satz mit negiertem Antezedens zu vermeiden ist, vielleicht indem man normale und außergewöhnliche Fälle hinsichtlich des Akzeptanzstatus von A in $T^{-}{}_{B}$ unterscheidet. Diese sehr diffizile Angelegenheit soll hier aber nicht weiter verfolgt werden.

Ich kenne keine natürlichsprachliche Realisierung des Nichtkonditionals, in der das Antezedens und das Konsequens akzeptiert sind. Der Vollständigkeit halber möchte ich die entsprechende Akzeptabilitätsbedingung dennoch auflisten:

$$\begin{aligned}(6.7.17)\quad A\ ^{a}\text{—}\!\shortmid\! B \in T \;&\Leftrightarrow\; A\text{—}\!\shortmid\! B \in T \wedge A,B \in T\\ &\Leftrightarrow\; B,\neg B \notin (T^{-}{}_{B})^{*}{}_{\neg A} \wedge A,B \in T\ .\end{aligned}$$

Offenbar ist $A^{a}\text{—}\!\shortmid\! B \in T$ äquivalent mit $\neg A^{z}\text{—}\!\shortmid\! B \in T$. Dies gibt zu Spekulatio-

nen darüber Anlaß, warum $\overset{a}{\dashrightarrow}$ nicht realisiert ist: ich vermute, wegen der Ökonomie der natürlichen Sprache. Den eben erwähnten Verdacht einmal beiseite lassend, kann man annehmen, daß es ebenfalls Ökonomiegründe sind, die einer Realisierung von $\overset{z}{\prec}$ entgegenstehen (man erinnere sich, daß genau dann $A \overset{z}{\prec} B \in T$, wenn $\neg A \overset{a}{\rightarrow} B \in T$). Andererseits bin ich mir nicht sicher, daß in der natürlichen Sprache keine Ausdrucksmöglichkeit für $\overset{\circ}{\prec}$ existiert (man erinnere sich, daß (6.7.12) nur eine zu starke Akzeptabilitätsbedingung lieferte). Es deutet manches darauf hin, daß das indikativische auch wenn am besten durch eine disjunktive Kombination von $\overset{\circ}{\dashrightarrow}$ und $\overset{\circ}{\prec}$ erfaßt wird, was zu der Akzeptabilitätsbedingung $A \rightarrow B \notin T^-{}_B$, oder äquivalent zu $\neg A \notin K^*{}_{\neg B}$, führen würde.[23]

Das Endergebnis meiner Thesen ist, zusammen mit den Bedingungen für (vermutlich) nicht realisierte Kombinationen, auf Seite 236f tabellarisch zusammengefaßt.

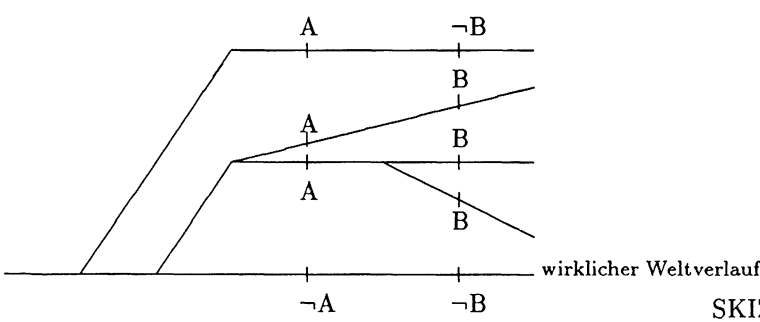
6.8 Vergleich mit der Theorie von McCall

Es gibt nur sehr wenig Literatur zum Thema dieses Kapitels. Eine sehr bemerkenswerter und meines Wissens der einzige neuere, explizite Ansatz zu einer Verbindung von Konditional- und Kausalsätzen stammt von Storrs McCall (1983; 1984; 1987). Sein Anwendungsbereich ist etwas enger als der unsere, denn er betrachtet nur die $\Box$-Versionen von Prokonditionalen.

[23] Hierzu vergleiche man Goodmans (1954, Fußnote 2) Zögern, ob auch wenn als kontradiktorisches oder als konträres Gegenteil von wenn aufzufassen sei. Goodman (1954, S. 15f) beobachtet, daß (das konjunktivische) Auch wenn A, B normalerweise als die Negation des kontrafaktischen Wenn A, $\neg$B gemeint ist (beachte, daß dies genau der Lewisschen 1973a, S. 2, Forderung für möglicherweise-Konditionalsätze entspricht). Dies kann jedoch als Ansatz für auch wenn nicht hinreichen, wenn die folgende Überlegung korrekt ist. Nach meiner Auffassung wird die Negation eines natürlichsprachlichen Konditionalsatzes akzeptiert, wenn seine Randbedingung akzeptiert, seine Hauptbedingung aber zurückgewiesen wird. Dies bedeutet für den in Frage stehenden Fall, daß $\neg(A \overset{z}{\Box\!\!\rightarrow} \neg B) \in T$ genau dann gilt, wenn $\neg B \notin T^*{}_A$ und $\neg A, B \in T$ (vgl. (6.7.7)). Aber die charakteristische Bedingung $\neg B \notin T^*{}_A$ ist sicherlich zu schwach für Auch wenn A, B (dies bestätigt eher Lewis). Wie dem auch sei, nach den Thesen (6.6.3)–(6.6.5) impliziert das Akzeptieren von Auch wenn A, B $A \notin T \wedge B \in T$, und dies paßt zu Goodmans Vorschlag, auch wenn das „semifaktische Konditional" zu nennen. Gärdenfors' (1981, S. 209; 1988, S. 153) Akzeptabilitätsbedingung $B \in T^*{}_A \cap T^*{}_{\neg A}$ $(= (T^-{}_A)^-{}_{\neg A}$, s. Lemma 3.1.6) für Auch wenn A, B erscheint ebenfalls zu schwach. Sie ist immer erfüllt, wenn $A, \neg A \notin T \wedge B \in T$, eine Situation, die für indikativisches auch wenn notwendig, aber wohl kaum hinreichend ist. Ich bin nicht geneigt, den Satz *Auch* wenn Herr Bonzmann 13 Rolls Royce besitzt, ist er reich zu akzeptieren; richtig muß es heißen Wenn Herr Bonzmann 13 Rolls Royce besitzt, ist er reich.

Ähnlich wie Goodman spricht McCall von „counterfactuals", „factuals" (since-Sätzen), „semifactuals" (Prokonditionale mit falschem Antezedens und wahrem Konsequens) und von „subjunctive and indicative conditionals", die er — im Gegensatz zum Vorgehen in dieser Arbeit — als stilistische Varianten von offenen Konditionalsätzen anzusehen scheint. Seine Semantik für all diese Konditionalsätze ist keine Semantik von Akzeptabilitätsbedingungen, sondern eine von Wahrheitsbedingungen, die auf baumartig „verzweigten mögliche-Welten-Strukturen" basiert. Gemeint sind hiermit mögliche Weltverläufe, die sich „smooth and lawlike" (McCall 1983, S. 311; 1984, S. 467; 1987, S. 2), ohne Sprünge und Diskontinuitäten nach der Zukunft hin verzweigen. Voraussetzung hierfür ist, daß McCall etwa im Gegensatz zu Lewis (1979a) ein indeterministisches Weltbild zugrundelegt, wonach sich bei einer gegebenen Vergangenheit immer wieder verschiedene Möglichkeiten für die Zukunft eröffnen. McCalls „objektiver" Maßstab für die Ähnlichkeit eines möglichen Weltverlaufs mit dem wirklichen Weltverlauf ist eindeutig festzulegen als die Länge der gemeinsamen Vergangenheit (McCall 1983, S. 312; 1984, S. 467f; 1987, S. 3).

Die Wahrheitsbedingung für kontrafaktische Konditionalsätze ist an Stalnaker und Lewis angelehnt. Ein kontrafaktischer Konditionalsatz Wenn A, (dann) B ist genau dann (auf nichttriviale Weise[24]) wahr, wenn jede nächste A-Welt eine B-Welt ist (McCall 1983, S. 312f; 1984, S. 467).



SKIZZE 6.1

Diese Bedingung bleibt nach McCall auch dann gültig, wenn (¬)B früher als (¬)A passiert.

Ein „faktischer Konditionalsatz" Weil A, (deshalb) B ist genau dann

[24] Wenn A, dann B heißt auf triviale Weise wahr, wenn es gar keine (erreichbaren) A-Welten gibt oder wenn alle (erreichbaren) Welten B-Welten sind (McCall 1987, S. 11f).

Akzeptabilitätsbedingungen von na-

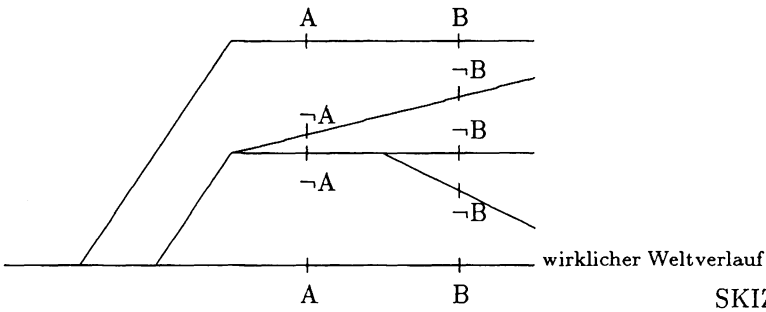
		Ohne explizites möglicherweise	
	Randbedingungen	$B \in T$	
W O E B I W L O H & L	$A \in T$	**obwohl** $A \ ^{a}\!\prec B \in T$	* $A \ ^{a}\!\rightarrowtail B \in T$
		□-Version $B \in (T^{-}{}_{B})^{*}{}_{\neg A}$ $(\Rightarrow A \in T^{-}{}_{B})$ ◇-Version unmöglich	$B, \neg B \notin (T^{-}{}_{B})^{*}{}_{\neg A}$ $(\Rightarrow A \in T^{-}{}_{B})$
W E	$A, \neg A \notin T$	* (ind. auch wenn?) $A \ ^{\circ}\!\prec B \in T$	**ind. auch wenn** $A \ ^{\circ}\!\rightarrowtail B \in T$
		□-Version $\neg A \rightarrow B \in T^{-}{}_{B}$ $(\Leftrightarrow A \in T^{*}{}_{\neg B})$ ◇-Version unmöglich	$A \rightarrow B, \neg A \rightarrow B \notin T^{-}{}_{B}$ $(\Leftrightarrow A, \neg A \notin T^{*}{}_{\neg B})$
N N	$\neg A \in T$	* (konj. auch wenn??) $A \ ^{z}\!\prec B \in T$	**konj. auch wenn** $A \ ^{z}\!\rightarrowtail B \in T$
		□-Version $\neg A \rightarrow B \in T^{-}{}_{B}$ $(\Leftrightarrow A \in T^{*}{}_{\neg B})$ $(\Rightarrow \neg A \notin T^{-}{}_{B})$ ◇-Version trivial, wenn $\neg A \notin T^{-}{}_{B}$	$B, \neg B \notin (T^{-}{}_{B})^{*}{}_{A}$ $(\Rightarrow \neg A \in T^{-}{}_{B})$
		KONTRA-KONDITIONALE	NICHT-KONDITIONALE

türlichsprachlichen Konditionalsätzen

	Mit explizitem möglicherweise	
	$B, \neg B \notin T$	$\neg B \in T$
weil $A \overset{a}{\rightarrow} B \in T$	* *	* *
□-Version $A \rightarrow B \in T^-{}_B$ $(\Leftrightarrow \neg A \in T^*{}_{\neg B})$ $(\Rightarrow A \notin T^-{}_B)$	□-Version unmöglich	□-Version unmöglich
◇-Version $\neg B \in (T^-{}_B)^*{}_{\neg A}$ trivial, wenn $A \notin T^-{}_B$	◇-Version $\neg B \in T^*{}_{\neg A}$	◇-Version unmöglich
* (ind. wenn?) *	**ind. wenn** $A \overset{\circ}{\Box\rightarrow} B \in T$	* (ind. wenn?) *
$A \rightarrow B \in T^-{}_B$ $(\Leftrightarrow \neg A \in T^*{}_{\neg B})$	$A \rightarrow B \in T$ $(\Leftrightarrow \neg A \in T^*{}_{\neg B})$	$A \rightarrow B \in T$
*(ind. wenn...mögl.?) *	**ind. wenn...mögl.** $A \overset{\circ}{\Diamond\rightarrow} B \in T$	*(ind. wenn...mögl.?) *
unmöglich	$\neg A \rightarrow \neg B \in T$ $(\Leftrightarrow B \rightarrow A \in T)$	unmöglich
* (konj. wenn?) *	* (konj. wenn?) *	**konj. wenn** $A \overset{z}{\Box\rightarrow} B \in T$
$B \in (T^-{}_B)^*{}_A$ $(\Rightarrow \neg A \in T^-{}_B)$	$B \in T^*{}_A$	$B \in T^*{}_A$ $(\Leftrightarrow B \in (T^-{}_B)^*{}_A)$
*(konj. wenn...mögl.?) *	*(konj. wenn...mögl.?) *	**konj. wenn...mögl.** $A \overset{z}{\Diamond\rightarrow} B \in T$
unmöglich	unmöglich	$\neg B \notin T^*{}_A$

PROKONDITIONALE

(auf nichttriviale Weise[25]) wahr, wenn es eine ¬A∧¬B-Welt gibt, die näher an der wirklichen Welt ist als jede A∧¬B-Welt (McCall 1983, S. 315), d.h. — vorausgesetzt es gibt überhaupt eine ¬B-Welt — wenn jede nächste ¬B-Welt eine ¬A-Welt ist.



SKIZZE 6.2

Ein „semifaktischer Konditionalsatz" Wenn A, (dann) B, der nicht von der Art eines (nicht- oder kontrakonditionalen)auch wenn-Satzes ist, sondern dieselbe (prokonditionale) Verknüpfung zwischen A und B ausdrückt wie kontrafaktische und faktische Konditionalsätze,[26] ist wahr, wenn jede nächste A-Welt eine B-Welt und wenn jede nächste ¬B-Welt eine ¬A-Welt ist (McCall 1983, S. 318). McCalls (1983, S. 315f) Wahrheitsbedingungen für von ihm so genannte konjunktivische und indikativische Konditionalsätze sind kompliziert und etwas undurchsichtig und brauchen hier nicht diskutiert werden.

Die drei angegebenen Fälle lassen sich verallgemeinern, wenn man die plausible „egozentrische" Annahme macht, daß der wirkliche Weltverlauf sich selbst der einzig nächste ist. Kontrafaktische, faktische und semifaktische Konditionalsätze sind dann Spezialfälle eines universelleren Konditionals, das der folgenden Bestimmung genügt (vgl. McCall 1987, S. 9–11):

6.8.1. Definition Ein Konditional A⇝B ist genau dann wahr in einer Welt w, wenn die an w nächstliegenden A-Welten B-Welten und wenn die an w nächstliegenden ¬B-Welten ¬A-Welten sind.

Die Idee hinter dieser Bedingung ist, daß die A∧¬B-Welten, die das

[25] Dies muß wohl analog hinzugefügt werden. Weil A, deshalb B heißt demgemäß auf triviale Weise wahr, wenn es gar keine (erreichbaren) ¬B-Welten gibt oder wenn alle (erreichbaren) Welten ¬A-Welten sind.

[26] Es ist die Frage, ob es solche überhaupt gibt. McCalls (1983, S. 317f; 1987, S. 5f) Beispiele sind nicht ganz überzeugend. Vgl. auch die Beispielsdiskussion am Ende dieses Abschnitts.

Konditional $A \rightsquigarrow B$ falsifizieren würden, relativ am weitesten von w entfernt sind. Wenn die Motivation von 6.8.1 auch nicht völlig klar ist, so soll es doch ein Explikat für die Verbindung („connection", „link") sein, die durch Konditionalsätze aller Art zum Ausdruck kommen kann. Formales Merkmal solcher Konditionalsätze, die intuitiv eine Verbindung zwischen Antezedens und Konsequens behaupten, ist, so McCall (1983, S. 318–320; 1987, S. 1, 11), ihre Kontraponierbarkeit:[27] die Regel der doppelten Negation vorausgesetzt, ist 6.8.1 invariant gegenüber der gleichzeitigen Ersetzung von A durch $\neg B$ und von B durch $\neg A$.

McCall hat versucht, mit seinen verzweigten Weltverläufen anstelle der problematischen Lewisschen Ähnlichkeitsrelation zu einer physikalisch-objektiven Grundlage der Beurteilung von Konditionalsätzen zu kommen. Dies ist ein ehrgeizigeres Unterfangen als eine eher subjektive, jedenfalls aber theorienrelative Beurteilung à la Gärdenfors. Es ist aber auch von viel eingeschränkterer Anwendbarkeit. Zunächst gibt es natürlich sehr viele Konditionalsätze, deren Antezedenzien (und Konsequenzen) man keine festen Zeitpunkte oder -räume zuordnen kann (vgl. Lewis 1979a, S. 464), und insbesondere werden die uns interessierenden Revisionen von wissenschaftlichen Theorien typischerweise durch solche „zeitlosen" kontrafaktischen Annahmen vorgenommen (s. die beiden Beispiele von Kapitel 8). Aber selbst wenn man deutliche Referenzzeiten hat, ist es nicht immer richtig, als nächstliegende Weltverläufe diejenigen anzusehen, die am längsten mit dem wirklichen Weltverlauf übereinstimmen. Wir mischen uns ein letztes Mal in die Affäre zwischen Anton und Berta und überlegen nach der Party noch einmal, ob der Satz

> Wenn Anton zu Hause geblieben wäre, wäre Berta zur Party gegangen.

richtig sein kann. Die Situation war die, daß Anton die Party besuchte, Berta aber nicht, weil sie wußte, daß Anton kommen würde und er ihr inzwischen doch reichlich auf die Nerven geht. Ansonsten wäre Berta aber gerne gekommen. Unter diesen Umständen werden wir den angegebenen Konditionalsatz ohne Zweifel akzeptieren. Die nach dem McCallschen Modell wirklichkeitsnächste Möglichkeit, das Antezedens zu realisieren, ist aber die, daß sich Anton in letzter Minute, als er eigentlich schon losgehen wollte, die Sache anders überlegt und auf den Partybesuch verzichtet. In

[27] Natürlich wird durch die Kontraposition aus einem faktischen Konditionalsatz ein kontrafaktischer und umgekehrt.

diesem Fall wäre es für Berta aber schon zu spät zum Umdisponieren gewesen, sie hätte gar nicht rechtzeitig erfahren, daß die Bahn für sie (und Christoph) frei sein würde. Wenn wir den obigen Konditionalsatz also für wahr halten, dann deshalb, weil Anton *normalerweise* frühzeitig weiß, was er will, und mit seinen Plänen nicht hinter dem Berg hält, weil es sich mithin schnell bis zu Berta herumgesprochen hätte, daß Anton nicht kommt. Nun hat McCall (1984, S. 471–473), im Gegensatz zu Bennett und Lewis, expressis verbis anerkannt, daß Konditionalsätze in der Regel ein gewisses „Backtracking" verlangen und daß die „last minute deviation objection" stichhaltig ist.[28] Aber er hat nicht genügend gewürdigt, daß seine eigene Theorie davon bedroht ist. Naheliegend sind zumeist *nicht* die Verläufe, die am längsten mit dem wirklichen Verlauf übereinstimmen. Anstelle der McCallschen Verzweigungen, die nur durch den mehr oder weniger objektiven Begriff des physikalischen Gesetzes reglementiert sind, benötigt man eine genaue Vorstellung darüber, wie der *normale* Gang der Dinge, die zu A führen, aussehen müßte. dies ist nicht mehr loszulösen von epistemischen Subjekten oder Alltagstheoretikern, und deshalb ist die flexiblere, inhaltlich nicht festzunagelnde Relation der theoretischen Wichtigkeit angemessener als das starre Zeit-und-Gesetz-Modell McCalls.

Unabhängig von der nicht ganz glücklichen objektivistischen Grundlage seiner Theorie erscheint mir die Idee von (6.8.1) sehr interessant und nach Funktion und Zielsetzung — in meinen Worten: die Analyse der □-Versionen von Prokonditionalen — eine echte Alternative zur □-Version von (6.5.1) darzustellen. Um eine gute Vergleichsgrundlage zu haben, wollen wir (6.8.1) jetzt auf die offensichtliche Art in das bisher verwandte Schema von Akzeptabilitätsbedingungen übertragen:

(6.8.2) $A \rightsquigarrow B \in T \Leftrightarrow B \in T^*_A \wedge \neg A \in T^*_{\neg B}$.[29]

Dies ist die aus Kapitel 4 bekannte Ramsey-Test-Variante (R4). Demge-

[28] Lewis argumentiert, wie gesagt, von einer deterministischen Position aus und kann deshalb Verzweigungen von Weltverläufen nur durch mehr oder weniger große Wunder erklären. Sein „System von Gewichten und Prioritäten" (Lewis 1979a, S. 472), wonach die großen Verletzungen von Naturgesetzen dringend zu vermeiden sind, und sei es auf Kosten des „raumzeitlichen Gebiets, in dem eine völlige Übereinstimmung der Einzeltatsachen vorherrscht", scheint in vielen Fällen ebenfalls ein Backtracking zu erzwingen, was nicht recht zu Lewis' (1979a, S. 456f) Ansicht paßt, die „Standardauflösung" von kontrafaktischen Konditionalsätzen würde ohne Backtracking geschehen.

[29] Diese einfache Übersetzung beseitigt unter der Hand schon die von McCall (1987, S. 11) angeführten logischen „Anomalien" seiner Idee. Eigentlich müßte man in (6.8.2) noch hinzufügen, daß A und ¬B von T aus „erreichbar" sein müssen, d.h. nach (T*5), daß weder $\vdash \neg A$ noch $\vdash B$ gilt. Aber wir sind ja stets von kontingenten Sätzen ausgegangen.

genüber hat das oben klassifizierte $A\Box\!\!\rightarrow B$ die Akzeptabilitätsbedingung $A \gg B \in T^-_B$, was sich, wie bemerkt, wegen Lemma 3.1.6 reduziert auf

(6.8.3) $\quad A\Box\!\!\rightarrow B \in T \Leftrightarrow B \in (T^-_B)^*_A$.

Aus $B \in (T^-_B)^*_A$ folgt $\neg A \in T^*_{\neg B}$: Denn angenommen $\neg A \notin T^*_{\neg B} \supseteq T^-_B$, dann heißt $B \in (T^-_B)^*_A$ nichts anderes als $A \rightarrow B \in T^-_B$, was gleichbedeutend mit $\neg A \in T^*_{\neg B}$ ist, im Widerspruch zur Voraussetzung. Dies scheint jedoch der einzige, höchst partielle Zusammenhang zwischen (6.8.2) und (6.8.3) zu sein. Das Erstaunliche ist nun, daß sich unter den neun möglichen Kombinationen der Randbedingungen A und B nur in einem einzigen Fall ein Unterschied zwischen (6.8.2) und (6.8.3) ergibt. Falls A in T ist, muß nach beiden Bedingungen auch B in T sein, d.h. es liegt der rein faktische Fall vor, und sowohl nach McCalls als auch nach meinem Vorschlag bleibt $\neg A \in T^*_{\neg B}$, oder äquivalent $A \rightarrow B \in T^-_B$. Falls weder A noch $\neg A$ in T ist, gibt es zwei verschiedene Fälle: Wenn darüber hinaus $B \in T$, erhalten wir dasselbe Ergebnis wie eben; wenn aber zusätzlich $B \notin T$, folgt aus (6.8.2) ebenso wie aus (6.8.3) die Bedingung $A \rightarrow B \in T$. Falls schließlich $\neg A$ in T ist, sind wieder zwei Möglichkeiten zu unterscheiden. Wenn darüber hinaus $B \notin T$, dann liegt der rein kontrafaktische Fall vor und die beiden Akzeptabilitätsbedingungen ergeben übereinstimmend $B \in T^*_A$ oder, anders gesagt, $A \rightarrow B \in T^-_{\neg A}$. Der einzig übrige Fall ist der semifaktische und durch $\neg A \in T$ und $B \in T$ gegeben, und hier finden wir die einzige Differenz. Nach (6.8.2) (und Levis These) ist $A \rightarrow B$ sowohl in $T^-_{\neg A}$ als auch in T^-_B, mit (T-5) gilt daneben $B \rightarrow \neg A \in T^-_B$ und $\neg A \rightarrow B \in T^-_{\neg A}$, d.h. zusammengenommen $\neg A \in T^-_B$ und $B \in T^-_{\neg A}$. Durch diese kleine Vereinfachung ist aber kaum etwas gewonnen, denn die Vergleichbarkeit mit der Bedingung $B \in (T^-_B)^*_A$ besteht allein in der Tatsache, daß aus letzterem $\neg A \in T^-_B$ folgt.

Kann man herausfinden, welches von (6.8.2) und (6.8.3) die bessere Idee darstellt, indem man Situationen mit $\neg A \in T$ und $B \in T$ nennt, in denen eines der beiden Kriterien angemessene, das andere aber unangemessene Resultate liefert? Ich führe ein Beispiel an, welches vielleicht als Entscheidungshilfe dienen kann. Stellen wir uns vor, Berta fährt in den Urlaub nach Afrika, in ein Gebiet, wo die Malariamücke Anopheles so verbreitet ist, daß man ohne vorbeugende Maßnahmen praktisch sicher Malaria bekommt. Nehmen wir weiter an, es gibt genau zwei Mittel gegen Malaria, M_1 und M_2, die beide zuverlässig vor der Infektion schützen, die aber miteinander unverträglich sind und kombiniert zu schweren Vergiftungen führen. Es soll also folgendes vereinfachtes „Gesetz" gelten: Wenn man gar nichts gegen Malaria oder wenn man M_1 und M_2 einnimmt, dann wird man krank, wenn man aber genau eines der beiden Mittel nimmt, bleibt

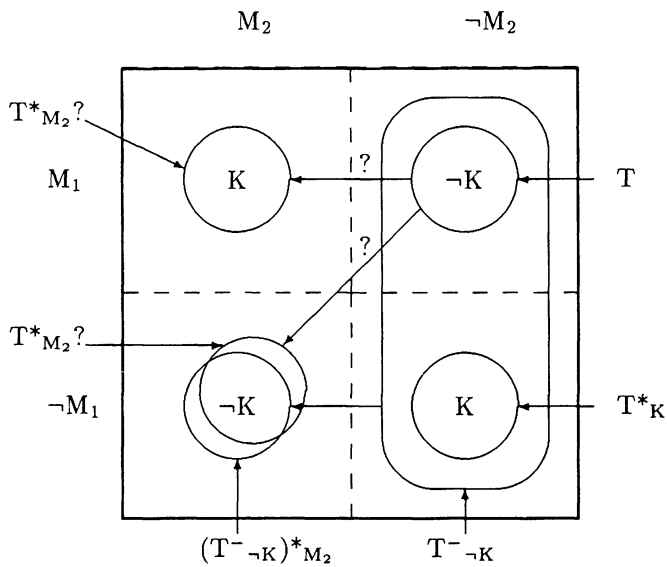
man gesund, kurz und in Zeichen $(M_1 \leftrightarrow M_2) \leftrightarrow K$. Mir sei nun außerdem bekannt, daß Berta sich für das Mittel M_1 entschieden hat und einige schöne Wochen ohne gesundheitliche Beschwerden verlebte. Also, so schließe ich, hat sie M_2 nicht eingenommen, und meine kleine Theorie T über Bertas Urlaub enthält $(M_1 \leftrightarrow M_2) \leftrightarrow K$, M_1, $\neg M_2$ und $\neg K$. Ich würde sogar sagen, daß Berta gesund blieb, *weil* sie M_1 eingenommen hatte, was sowohl nach McCalls als auch nach meiner Analyse heißt, daß $\neg M_1 \in T^*_K$. Damit ist klar, daß T^*_K das (theoretisch besonders wichtige) Gesetz $(M_1 \leftrightarrow M_2) \leftrightarrow K$ und die kontingenten Aussagen $\neg M_1$, $\neg M_2$ und K enthalten muß. Wir betrachten nun den „semifaktischen" Konditionalsatz

(6.8.4) Wenn Berta M_2 genommen hätte, wäre sie (auch) nicht krank geworden.[30]

Dieser Satz ist intuitiv wohl akzeptabel, denn unwillkürlich kontrastiert man die hypothetische Einnahme von M_2 mit der tatsächlichen Einnahme von M_1. Dagegen ist es fraglich, ob man wirklich behaupten kann, daß der Ramsey-Test in seiner ursprünglichen Fassung, wie er auch in (6.8.2) enthalten ist, ein richtiges Bild liefert. Gilt also $\neg K \in T^*_{M_2}$? Um die Gesundheit von Berta zu garantieren, müßten wir nachweisen, daß M_1 in $T^*_{M_2}$ gelöscht wird, da die Einnahme beider Mittel ja zu Vergiftungen führt. Die naheliegende Idee ist es, zu sagen, Berta sei wohl so schlau, nicht zwei verschiedene Medikamente gegen dieselbe Krankheit gleichzeitig zu nehmen, also sei insbesondere $M_1 \rightarrow \neg M_2$ wichtiger als M_1, also wäre $M_1 \rightarrow \neg M_2$, nicht aber M_1 in $T^*_{M_2}$ aufzufinden. Das Problem ist aber, daß wir M_1 sozusagen als Axiom von T gegeben hatten und daß die größere theoretische Wichtigkeit von $M_1 \rightarrow \neg M_2$ gegenüber M_1 postuliert werden muß. Man kann sich, so scheint es, sich durchaus auf den Standpunkt stellen, es sei sehr wichtig, daß Berta M_1 genommen hat (M_1 ist vielleicht auch noch für andere Dinge gut). Damit würde man formal $K \in T^*_{M_2}$ erhalten, ohne deshalb unbedingt den semifaktischen Konditionalsatz (6.8.4) aufgeben zu wollen. Dann wäre die Akzeptabilitätsbedingung (6.8.2) verletzt. Anders sieht die Sache aus, wenn wir gemäß (6.8.3) $\neg K \in (T^-_{\neg K})^*_{M_2}$ betrachten. $T^-_{\neg K}$ ist nach der Harper-Identität (s. Abschnitt 3.1) gleich dem Durchschnitt von T und T^*_K. Wie oben gesehen, enthält T^*_K die Sätze $\neg M_1$ und $M_1 \rightarrow \neg M_2$. Also ist in $T^-_{\neg K}$ zwar $M_1 \rightarrow \neg M_2$ enthalten, nicht aber M_1. Der Konflikt zwi-

[30] Es ist gar nicht einfach zu sagen, ob man diesen Satz als prokonditional oder als nichtkonditional ansehen soll, da einerseits zwar M_2 gegen Malaria *wirkt*, andererseits aber Berta de facto ja *ohnehin* nicht krank geworden ist. Da es McCall nur um die Verknüpfung geht, die auch in kontrafaktischen und faktischen Konditionalsätzen zum Ausdruck kommt, wollen wir (6.8.4) als Prokonditional behandeln.

schen den beiden Sätzen, der bei der McCallschen Ramsey-Test-Bedingung nicht eindeutig auflösbar war, kann nun gar nicht erst entstehen.[31] Deshalb bleibt $M_1 \rightarrow \neg M_2$ in $(T^{-}{}_{\neg K})^*{}_{M_2}$ erhalten, es folgt nach (T*2) und (T*1) $\neg M_1 \in (T^{-}{}_{\neg K})^*{}_{M_2}$, und da das Gesetz $(M_1 \leftrightarrow M_2) \leftrightarrow K$ als wichtigster Bestandteil der Theorie T erhalten bleibt, gilt $\neg K \in (T^{-}{}_{\neg K})^*{}_{M_2}$. Das folgende Diagramm faßt die Verhältnisse in diesem Beispiel anschaulich zusammen:



SKIZZE 6.3

Das Beispiel scheint ein Indiz dafür zu sein, daß der Vorschlag (6.8.3) besser ist als der Vorschlag (6.8.2). Natürlich kann es durchaus sein, daß sich auch Beispiele für die umgekehrte Richtung finden lassen. In den meisten Fällen werden die Akzeptabilitätsbedingungen (6.8.2) und (6.8.3) aber auch für semifaktische Konditionalsätze die gleichen Ergebnisse liefern.

Zusammenfassend bleibt festzuhalten, daß die Unterschiede der beiden formalen Ansätze für die □-Versionen von Prokonditionalen gegenüber ihren Ähnlichkeiten verschwindend klein sind. Besonders interessant und

[31] Formal entsteht freilich ein analoger Konflikt mit dem in $T^{-}{}_{\neg K} = T \cap T^*{}_K$ enthaltenen Satz $M_2 \rightarrow M_1$. Dieser Satz wird aber nach den Voraussetzungen des Beispiels sicher eine geringere theoretische Wichtigkeit haben als $M_1 \rightarrow \neg M_2$.

befriedigend erscheint es mir, daß McCall wie Goodman und ich zu der Auffassung kommt, daß der Satz Weil A, (deshalb) B mit seiner kontrafaktischen „Kontraposition" Wenn ¬B, (dann) ¬A äquivalent ist. Wir werden genau diese Beziehung im nächsten Kapitel brauchen, wenn wir wieder unsere ursprüngliche wissenschaftstheoretische Problemstellung aufgreifen und daran gehen, zu sagen, welche Eigenschaften gute und überlegene Nachfolgertheorien haben sollen.

Kapitel 7

Intertheoretische Erklärungen, gute und überlegene Nachfolgertheorien

7.1 Kontinuität und Widerspruch

Wenden wir uns nun einem Lösungsvorschlag zu den Problemen zu, die von der Reduktionsdiskussion am Ende von Kapitel 2 aufgeworfen wurden. Wir werden aber nicht unmittelbar an diese Diskussion anschließen, sondern von vornherein mit den Begriffen und Ergebnissen des Revisionsmodells arbeiten. Um dies auch terminologisch kenntlich zu machen, sprechen wir im folgenden nicht mehr von „Reduktionen", sondern von „intertheoretischen Erklärungen". Obgleich es in der Bedeutung dieser Termini sicherlich feine Unterschiede gibt, wollen wir diese vernachlässigen und der weitverbreiteten Praxis folgen, „Reduktion" und „(intertheoretische) Erklärung" als im wesentlichen synonyme Begriffe zu verwenden.

Eine weitere, strenggenommen ebenfalls problematische Laxheit, die wir uns erlauben werden, besteht darin, daß wir Reduktionen bzw. intertheoretische Erklärungen stets als Relationen zwischen einer Vorgängertheorie T_1 und ihrer Nachfolgertheorie T_2 auffassen. Rationaler, fortschrittlicher

Theorienwandel soll umgekehrt stets durch eine solche Relation explizierbar sein.[1] Es muß gesagt werden, was es heißt, daß eine Theorie T_2 *besser* ist als eine Theorie T_1, denn bei progressivem Theorienwandel darf T_1 nur durch eine bessere Theorie T_2 ersetzt werden. Diese Redeweise präsupponiert eine Art Kontinuität im Übergang von von T_1 nach T_2. Eine Idee, Kontinuität zu explizieren, ist eben der Begriff der intertheoretischen Erklärung:

(7.1.1) T_2 erklärt T_1 .

Wenn nun Theorien als Satzmengen verstanden werden und wenn man das einfache deduktiv-nomologische Modell der Erklärung verwendet, kann man (7.1.1) in einem ersten Versuch so formalisieren:

(7.1.2) $T_2 \vdash T_1$.

Ich werde für die folgenden Überlegungen voraussetzen, daß T_1 und T_2 deduktiv abgeschlossen (d.h. Theorien im Sinne von Definition 3.1.1) und konsistent sind (d.h. $T_1 \neq T_\perp$ und $T_2 \neq T_\perp$). Dies ist, wissenschaftstheoretisch gesehen, eine stark idealisierende Voraussetzung.

Schema (7.1.2) ist im allgemeinen leider inadäquat. Wenn T_2 allgemeiner ist als T_1 — und dies ist es in den meisten Beispielfällen —, dann muß man T_2 noch durch gewisse Zusatzbedingungen ergänzen, um die Ableitung von T_1 ermöglichen. Nennen wir dies die *Anwendungsbedingungen für* T_1 vom Standpunkt von T_2 aus; der Buchstabe „A" soll in diesem Kapitel immer diese Anwendungsbedingungen bezeichnen. Es ist plausibel anzunehmen, daß A eine Menge von *nichttheoretischen* („nicht gesetzesartigen", „empirischen") Sätzen ist, welche die Anfangs- oder Randbedingungen beschreiben, die den Anwendungsbereich von T_1 charakterisieren. Jedenfalls aber soll A *in der Sprache von* T_2 *ausdrückbar* sein. Ich will mich hier nicht darauf festlegen, daß die Menge A notwendigerweise eindeutig bestimmt ist. Trotzdem werde ich, um lästige, komplizierende Existenzquantifikationen zu vermeiden, im folgenden so tun, als ob sie das wäre.[2] Wir haben dann also

[1] Was diese Formulierungen anbetrifft, so gibt es eine weitgehende Übereinstimmung zwischen der in der Einleitung genannten universellen Reduktionsthese Stegmüllers und dem hier vorgelegten Ansatz. In der Explikation ergeben sich dann aber sehr große Unterschiede, vgl. die Kapitel 1 und 7.

[2] Formal kann zur Disambiguierung bei verschiedenen möglichen Kandidaten A von Anwendungsbedingungen (die A's als je ein Satz aufgefaßt) in den meisten Fällen wohl die Disjunktion dieser A's dienen. Vgl. unten Fußnote 11. Praktisch wird es aber ein wichtiges und nichttriviales Problem bei Beispielsanalysen sein, die richtigen Anwendungsbedingungen für T_1 zu finden.

(7.1.3) $T_2, A \vdash T_1$.

(7.1.3) ist identisch mit der Bedingung (2.1.3), entspricht also genau dem klassischen D-Konzept der Reduktion. Aus den in Kapitel 2 angeführten Gründen lasse ich hier das Problem der Inkommensurabilität beiseite. T_1 soll bereits eine adäquate Übersetzung der Vorgänger- in die Sprache der Nachfolgertheorie bezeichnen, so daß man „Brückenprinzipien" oder ähnliches nicht mehr in A erwähnen muß. Nur wenn aus T_2 schon folgt, d.h. wenn man in T_2 schon „weiß", daß A der Fall ist, dann reduziert sich (7.1.3) auf (7.1.2).[3]

Eine Nachfolgertheorie T_2 ist noch besser, wenn sie außer T_1 auch einige Anomalien von T_1, d.h. das Scheitern von T_1 erklären kann, wenn sie — in Sklars (1967, S. 112) Worten — T_1 „wegerklären" kann. In diesem Fall gibt es ein empirisches Explanandum E und Anfangsbedingungen I derart, daß sich aus T_1 und I $\neg$E, aus T_2 und I aber E ableiten läßt. Also gilt auch

(7.1.4) $T_2, I \vdash \neg T_1$.[4]

Die Konjunktion von (7.1.3) und (7.1.4) würde keine Probleme bereiten, wenn wir einfach ausschließen könnten, daß I zum Anwendungsbereich von T_1 gehört, d.h. wenn wir einfach $A \vdash \neg I$ festsetzen könnten. Das Problem mit dieser „Lösung" ist aber, daß weder der „anomale" Charakter von empirischen Befunden noch die Grenzen des Anwendungsbereichs von T_1 präzise festgemacht werden können. Auch innerhalb dessen, was man plausiblerweise den Anwendungsbereich von T_1 nennen könnte, erweisen sich die Voraussagen auf der Grundlage von T_1 häufig als strenggenommen falsch, wenn man sie (im Lichte von T_2) genau betrachtet, sie stellen also strenggenommen Anomalien dar. Dies ist die Motivation von Duhem und Feyerabend, die die These vertraten, daß aufeinander folgende Theorien im allgemeinen *an sich schon inkonsistent* sind:

(7.1.5) $T_2 \vdash \neg T_1$.

(7.1.2) und (7.1.5) machen das Problem am augenfälligsten: Wie kann T_2 gleichzeitig für und gegen T_1 sprechen? Wie ist Kontinuität trotz Wider-

[3] Das strukturalistische Gegenstück zu (7.1.3) ist im Adams-Sneedschen ($K2^s$) zu finden, und die Entsprechung zu (7.1.2) erhielte man durch die Zusatzbedingung $M^o_{p2}=M_{p2}$.

[4] Es erscheint nicht ganz einfach, das soeben skizzierte Argument in strukturalistische Begriffe zu übertragen. Man kann aber auch das Mayrsche Anomalienkriterium hernehmen, um zu einer Entsprechung zu (7.1.4) zu gelangen. Nach ($K7^s$) gilt insbesondere $\mathbf{F}[M_2] \cap \mathcal{C}M_1 \neq \emptyset$, weshalb man einen T_2-Satz I mit $||I|| = \mathbf{F}^{-1}[\mathcal{C}M_1]$ konsistent zu T_2 hinzufügen kann und (7.1.4) bzw. die strukturalistische Entsprechung $\mathbf{F}[M_2 \cap ||I||] \subseteq \mathcal{C}M_1$ erhält.

spruch möglich?[5] Wenn T_2 mit A und I verträglich ist, dann steht (7.1.2) im Widerspruch zu (7.1.4) und (7.1.3) im Widerspruch zu (7.1.5). Schließlich sind sogar (7.1.3) und (7.1.4) inkompatibel, wenn man nicht apodiktisch ausschließen will, daß Anomalien von T_1 auch — und gerade — im intendierten Anwendungsbereich von T_1 auftreten. In meinen Augen geben (7.1.3) und (7.1.4) letztlich doch ein korrekteres Bild des Theorienwandels als (7.1.2) und (7.1.5). Insbesondere deuten die erstgenannten Kriterien auf eine andere Ausdrucksweise hin als die letzten beiden: Anstatt zu sagen, daß T_2 (das Scheitern von) T_1 erklärt, finde ich es natürlicher und genauer, von A und I als den *Explanantia* der *Explananda* T_1 und $\neg T_1$ *relativ zu* (oder einfach *in*) T_2 zu sprechen. Ich lege diese Redeweise zugrunde, wenn ich nun versuchen werde, eine neue Explikation des Verhältnisses zwischen aufeinander folgenden Theorien zu skizzieren.

7.2 Definitionen

Der Kern meiner Idee besteht darin, den Anwendungsbedingungen eine zentrale Rolle in intertheoretischen Erklärungen zuzuteilen. Die Funktion von A in (7.1.3) erscheint klar, aber wozu soll A gut sein, wenn T_1 aus T_2 alleine ableitbar oder mit T_2 alleine unverträglich ist? Mein Vorschlag lautet, daß A im Fall (7.1.2) eine *faktische* (oder *reale*) Erklärung von T_1 in T_2 liefert, während A im Fall (7.1.5) eine *kontrafaktische* (oder *idealisierende*) Erklärung von T_1 in T_2 liefert. Da A in (7.1.3) normalerweise nicht als Teil von T_2 aufgefaßt werden soll, kann man hier sagen, daß A eine *potentielle* (oder *bedingte*) Erklärung von T_1 in T_2 darstellt.[6] Ich glaube nicht, daß eine ganz bestimmte logische Relation zwischen T_1 und T_2 (ohne Zusatzbedingungen) notwendig ist, um den Übergang von T_1 zu T_2 rechtfertigen zu können. Jede Art von Erklärung kann diesen Übergang zu einem rationalen Schritt machen. Entsprechend wollen wir drei Fälle unterscheiden:

[5] Scheibe (1975; 1976; 1982; 1984; 1988) hat eine ganze Reihe sehr interessanter Artikel zu genau dieser Frage geschrieben. Nach einer Stelle aus Boltzmanns *Populären Schriften* nennt er die These, daß im wissenschaftlichen Theorienwandel gleichzeitig Kontinuität und Widerspruch gegeben sind, die „Boltzmannsche Dialektik". Vgl. auch Abschnitt 1.9.

[6] Es ist vielleicht nicht unnötig, darauf hinzuweisen, daß diese Begriffe mit Scheibes (1983, S. 76f) „potentieller" und „faktischer" Erklärung nichts zu tun haben. Vgl. Abschnitt 1.9. Mehr Gemeinsamkeiten bestehen mit Hempels (1965, S. 338) Unterscheidung zwischen „potentieller" und „akzeptierter" Erklärung.

7.2.1. Definition T_2 ist genau dann eine *gute (konservative) Nachfolgertheorie für* T_1, wenn
(a) A T_1 in T_2 faktisch erklärt *oder*
(b) A T_1 in T_2 potentiell erklärt *oder*
(c) A T_1 in T_2 kontrafaktisch erklärt .

Eine Bemerkung noch zur Redeweise. Wenn ich im folgenden davon spreche, daß A T_1 in T_2 „als Idealisierung" oder „durch eine Idealisierung" erklärt, intendiere ich immer den in Definition 7.2.1(c) angegebenen Fall. Das tue ich auch, wenn ich kurz und etwas schlampig sage, daß T_2 T_1 „als/durch eine Idealisierung" erklärt; diese letztere Sprechweise führt meistens noch die Konnotation mit sich, daß es auf den ersten Blick nicht klar ist, wie die Anwendungsbedingungen A für T_1 genau aussehen.

Eine wirklich überlegene, progressive Nachfolgertheorie T_2 sollte strikt besser als ihre Vorgängertheorie T_1 sein und auch deren (kontrafaktisches, potentielles oder faktisches) Scheitern erklären, sofern T_1 (nirgends, teilweise oder völlig) scheitert. Welche Zusatzbedingung kann aber die Rolle von I in (7.1.4) übernehmen und als Explanans dienen für das Scheitern von T_1? Ein einleuchtender Vorschlag ist es, einfach die Verletzung der Anwendungsbedingungen von T_1 als das, was T_1 „wegerklärt", in Anschlag zu bringen.[7] Das heißt, wir setzen einfach ¬A an die Stelle von I:

7.2.2. Definition T_2 ist genau dann eine *überlegene (progressive) Nachfolgertheorie für* T_1, wenn T_2 eine gute (konservative) Nachfolgertheorie für T_1 ist und
(a) ¬A das Scheitern von T_1 in T_2 kontrafaktisch erklärt *bzw.*
(b) ¬A das Scheitern von T_1 in T_2 potentiell erklärt *bzw.*
(c) ¬A das Scheitern von T_1 in T_2 faktisch erklärt .

Wenn man Anschluß an die Reduktionsdiskussion herstellen will, kann man die durch die Definitionen 7.2.1 und 7.2.2 eingeführten Begriffe auch als *schwache* und *starke Reduzierbarkeit von* T_1 *auf* T_2 bezeichnen.

Was heißt dies alles aber? Welcher Erklärungsbegriff hat Aussichten, die gestellte zweischneidige Aufgabe bewältigen, und zwar in der faktischen, der potentiellen und der kontrafaktischen Variante? Man vergegenwärtige sich noch einmal, daß von T_2 eine *doppelte Erklärungsleistung* gefordert ist, eine Erklärung sowohl ihrer Vorgängertheorie T_1 als auch des Scheiterns derselben. Eine Lösung wird erkennbar, wenn wir die verschiedenen Erklärungstypen durch natürlichsprachliche Formulierungen charakterisieren, nämlich durch weil-Sätze und ihre nahen Verwandten, die indikativi-

[7] Eine allgemeinere Idee wäre es, irgendein B mit B⊢¬A herzunehmen.

schen und konjunktivischen wenn-dann-Sätze:

7.2.3. Definition (a) A *erklärt* T_1 *in* T_2 (*faktisch*) genau dann, wenn der Satz Weil A der Fall ist, ist T_1 wahr in T_2 ist;
(b) A *kann* T_1 *in* T_2 *erklären* (*erklärt* T_1 *in* T_2 *potentiell*) genau dann, wenn der Satz Wenn A der Fall ist, ist T_1 wahr in T_2 ist;
(c) A *würde* T_1 *in* T_2 *erklären* (*erklärt* T_1 *in* T_2 *kontrafaktisch*) genau dann, wenn der Satz Wenn A der Fall wäre, wäre T_1 wahr in T_2 ist.

Nach der Analyse des letzten Kapitels behaupten diese Formulierungen einen je verschiedenen Status der Anwendungsbedingungen A von T_1 in T_2. In (a) muß schon T_2 beinhalten, daß A tatsächlich („faktisch") erfüllt ist. In (b) sagt T_2 entweder nichts darüber aus, ob A erfüllt ist, oder, daß A in manchen Fällen, unter bestimmten Umständen erfüllt ist. In (c) ist A, nach T_2 zu urteilen, nicht erfüllt oder nicht erfüllbar, d.h. die Anwendungsbedingungen für T_1 müssen Idealisierungen sein. Der Fall (c) ist eine Antwort auf die Unverträglichkeitsthese (7.1.5). Es sei noch einmal betont, daß die Anwendungsbedingungen für T_1 nach der vorliegenden Definition im allgemeinen nicht wahr und nicht einmal mit T_2 kompatibel sein müssen. In A können durchaus — und sind auch häufig — vereinfachende, idealisierende kontrafaktische Annahmen enthalten sein, die ein T_2-Theoretiker machen muß, damit T_1 für ihn akzeptabel wird. (Insofern sind die Anwendungsbedingungen für T_1, obgleich in einer eher theorieneutralen Sprache formuliert, natürlich von der Nachfolgertheorie T_2 abhängig.) Wir nennen kontrafaktische Erklärungen deshalb auch *idealisierende Erklärungen.*

„Das Scheitern von T_1 erkären" schließlich kann man plausiblerweise mit „$\neg T_1$ erklären" gleichsetzen.[8] Wir machen wieder von der im Anschluß an Definition 7.2.1 erwähnten genaueren Redeweise Gebrauch und erhalten sofort die

7.2.4. Definition (a) B *erklärt das (faktische) Scheitern von* T_1 *in* T_2 genau dann, wenn B $\neg T_1$ in T_2 erklärt;
(b) B *kann das Scheitern von* T_1 *in* T_2 *erklären* (*erklärt das potentielle Scheitern von* T_1 *in* T_2) genau dann, wenn B $\neg T_1$ in T_2 erklären kann;
(c) B *würde das Scheitern von* T_1 *in* T_2 *erklären* (*erklärt das kontrafaktische Scheitern von* T_1 *in* T_2) genau dann, wenn B $\neg T_1$ in T_2 erkären würde.

Zur Illustration und Erleichterung des Verständnisses seien jetzt noch einmal die natürlichsprachlichen Formulierungen für den wichtigen Spezialfall angeführt, daß T_2 erstens eine überlegene Nachfolgertheorie für T_1 ist

[8] Eine allgemeinere Idee wäre es, irgendein $T_1{}^*$ mit $T_1{}^* \vdash \neg T_1$ herzunehmen.

und zweitens im Widerspruch zu T_1 steht. Es ist offensichtlich, daß dann die Theorie T_2 ihre Vorgängertheorie nur kontrafaktisch, d.h. als Idealisierung erklären kann. In diesem Fall müßte also — wenn mein Ansatz richtig ist — ein T_2-Theoretiker folgende Formulierung als zutreffend akzeptieren: Wenn A der Fall wäre, dann wäre T_1 wahr; weil aber A tatsächlich nicht der Fall ist, ist T_1 nicht wahr.

7.3 Literaturverweise

In diesem Abschnitt möchte ich zwei Beiträge aus der wissenschaftstheoretischen Literatur vorstellen, die ganz ähnliche Gedanken erkennen lassen wie die eben präsentierten. Dabei sei kurz auf ein paar feine Unterschiede hingewiesen.

Einige der obigen Ideen kann man, wie bereits am Ende von Kapitel 2 angedeutet, bei Glymour (1970) finden, abstrakt auf S. 341:

> Inter-theoretical explanation is an exercise in the presentation of counterfactuals. One does not explain one theory from another by showing why the first is true; a theory is explained by showing under what conditions it *would be* true, and by contrasting those conditions with the conditions which actually obtain.

konkreter und vielleicht am besten formuliert auf S. 345:

> Thus Galileo's law is an approximation which *would* approach the Newtonian truth as a falling body comes arbitrarily close to the surface of the earth, *if* all forces other than the gravitational attraction of the earth were negligible and if the earth were spherical. Galileo's law fails in fact *because* the earth is not spherical and because forces other than the gravity of the earth are not zero and because the gravitational force is a function of distance. In the explanation of why Galileo's law fails one is not simply committing the fallacy of denying the antecedent. Rather, one is implicitly contrasting a contrary-to-fact situation in which Galileo's law would hold with the real situation, in which Newton's laws entail the denial of Galileo's law — or at least the denial of a formal analogue of that law.

(Hervorhebungen im Original.) Das einzige, was ich an der zweiten Belegstelle auszusetzen hätte, steht im ersten Satz. Hier werden Ideen der Ap-

proximation („approach", „comes arbitrarily close") mit Ideen der Idealisierung („if the earth were spherical") vermischt, zusammen mit dem Begriff „negligible", der immer dann verwendet wird, wenn man nicht recht weiß, ob man eher an Approximation oder an Idealisierung denken soll. Überhaupt scheint Glymour in seinem Aufsatz Approximationen und Idealisierungen nebeneinander als unabhängige Methoden der intertheoretischen Erklärung zu betrachten, wohingegen ich im nächsten Kapitel dafür plädieren werde, daß man danach streben sollte, zumindest in manchen Fällen wie etwa dem Kepler-Newton-Fall Idealisierungen *anstatt* Approximationen anzusetzen. Außerdem ist sich Glymour etwas unschlüssig darüber, was denn eigentlich von — in meinem Sinne — überlegenen Theorien erklärt wird (vgl. dazu Abschnitt 2.4.2). Im zweiten Zitat oben sagt er korrekt, daß NTG *das Scheitern* von Galileis Fallgesetz erklärt, während er im ersten Zitat meint, daß die Vorgängertheorie selbst erklärt wird, und wieder an anderer Stelle, daß erklärt wird, warum die Vorgängertheorie (speziell: das ideale Gasgesetz) „works where it works and fails where it fails" (S. 344). Nach meiner Ansicht erklären sowohl Approximationen als auch Idealisierungen in T_2 tatsächlich das Scheitern von T_1. Nur Idealisierungen, nicht aber Approximationen (im ersten Sinn[9]) erklären (kontrafaktisch) auch T_1 selbst, was man etwas vage, aber suggestiv auch dadurch ausdrücken kann, daß man sagt, T_2 erkläre *die theoretische Signifikanz von* T_1. Approximationen erklären, warum T_1 ungefähr richtig ist (vom Standpunkt von T_2 aus gesehen), oder — nach Sklar (1964, S. 81) — den scheinbaren Erfolg von T_1. Man könnte auch sagen, sie erklären *die praktische Signifikanz von* T_1 in T_2.

Den oben entwickelten Vorstellungen ebenfalls sehr nahe kommt der folgende Absatz von Yoshida (1977, S. 4):

> ... there are no good reasons for insisting on the *truth* of the auxiliary assumptions in the analysis of „incompatibility" and there are good scientific reasons for not so insisting. By allowing them to be false we can see the conditions under which the less comprehensive theory would appear to be true (in terms of the more comprehensive theory), and so we have the basis for an explanation of the success or failure and the limited or unlimited validity of the less comprehensive theory. When we know why the auxiliary assumptions are strictly speaking false,

[9] S. Abschnitt 2.4.1. Über Approximationen im zweiten Sinn wird in diesem Kapitel nichts gesagt.

> we know why the less comprehensive theory failed when it did and where it is not valid. When we know why the assumptions are approximately true, we know why the less comprehensive theory was as successful as it was and just where it is valid. This is also reason for allowing the auxiliary assumptions to be false on theoretical as well as empirical grounds. That is, it is reason for allowing auxiliary assumptions to be logically incompatible with certain postulates of the more comprehensive theory.

Bis auf eine unnötig verschwommene Formulierung („would appear to be true" statt einfach „were true") würde ich den ersten beiden Sätzen vorbehaltlos zustimmen, ebenso wie den beiden letzten Sätzen. Schwer zu verstehen finde ich hingegen die beiden Sätze dazwischen. Handelt es sich hier um eine Fallunterscheidung, mit der eine Opposition zwischen „strictly speaking false" und „approximately true" aufgebaut werden soll? Das wäre sonderbar, werden doch die beiden Prädikate im allgemeinen beinahe synonym verwendet. Yoshidas Ausführungen betreffs des Scheiterns bzw. Erfolgs der „umfassenderen Theorie" legen allerdings eine Opposition nahe. Schließlich wäre interessant zu erfahren, was es hier heißen soll, daß man weiß, *warum* die Hilfsannahmen strenggenommen falsch bzw. approximativ wahr sind. Yoshida läßt es uns leider nicht wissen.

7.4 Analyse

Der Grund dafür, warum es bei Glymour und Yoshida nicht recht weitergeht und ihre Betrachtungen letztendlich doch ungenau bleiben, ist, daß sie über keine ausgefeilte Theorie des kontrafaktischen Schließens verfügten. Wir dagegen haben inzwischen ein gut entwickeltes Modell für die Analyse von konjunktivischen und indikativischen Konditionalsätzen kennengelernt, welches auf Revisionen (minimalen Änderungen) von Theorien oder Überzeugungen basiert. In Kapitel 6 habe ich Gärdenfors' Kollektion der grundlegenden Rationalitätspostulate übernommen ((T*7) und (T*8) aus Kapitel 3 wurden nicht benötigt), seine Akzeptabilitätsbedingungen für Konditionalsätze jedoch modifiziert und den „Starken Ramsey-Test" anstelle des üblichen Ramsey-Tests für Konditionalsätze benutzt. Dies geschah mit dem Ziel, die Verwandtschaft zwischen Konditionalsätzen und weil-Sätzen erkennbar zu machen, wobei die letzteren als Standardformulierungen von Erklärungen betrachtet werden. Da intertheoretische Erklärungen offenbar stets warum notwendig- und nie nur wie möglich-Erklärungen sind (vgl.

Stegmüller [2]1983, S. 429ff, 998f), wird nur die □-Version von weil (vgl. die Abschnitte 6.5 und 6.7) zur Anwendung kommen.

Im folgenden bezeichne $T_2{}^*_C$ wieder die minimale Revision von T_2, die nötig ist, um C in T_2 „einzuverleiben", d.h. um C von der Warte der Theorie T_2 aus zu akzeptieren. Der Einfachheit halber setzen wir nun voraus, daß sowohl die Vorgängertheorie T_1 als auch ihre Anwendungsbedingungen A durch eine endliche Anzahl von Sätzen repräsentiert (axiomatisiert) werden kann, so daß wir T_1 und A je als einzelnen Satz, nämlich als die Konjunktion der repräsentierenden Sätze, behandeln dürfen.

7.4.1. Theorem T_2 ist genau dann eine gute Nachfolgertheorie für T_1, wenn

(a) $A, T_1 \in T_2$ und $\neg A \in T_2{}^*_{\neg T_1}$ *oder*
(b) $A, \neg A, T_1, \neg T_1 \notin T_2$ und $A \rightarrow T_1 \in T_2$ *oder*
(c) $\neg A, \neg T \in T_2$ und $T_1 \in T_2{}^*_A$.

Beweis Nach den Definitionen 7.2.1 und 7.2.3 und mit der Symbolisierung natürlichsprachlicher Konditionalsätze aus Kapitel 6 ist T_2 genau dann eine gute Nachfolgertheorie für T_1, wenn entweder $A^a\Box\!\!\rightarrow T_1$ oder $A^\circ\Box\!\!\rightarrow T_1$ oder $A^z\Box\!\!\rightarrow T_1$ in T_2 ist. Beachtet man, daß bei intertheoretischen Erklärungen nur die □-Version von weil eine Rolle spielen soll, so folgt die Behauptung aus den Akzeptabilitätsbedingungen (6.7.9) (beachte auch die Bemerkungen im Anschluß an (6.7.9)), (6.7.5) und (6.7.7). □

Man erinnnere sich, daß die □-Version des Satzes Weil C der Fall ist, ist auch D der Fall nach der Analyse im vorigen Kapitel äquivalent ist mit seiner konjunktivischen „Kontraposition" Wenn ¬D der Fall wäre, dann wäre auch ¬C der Fall. Dies ist dafür verantwortlich, daß A und T_1 im Fall (a) von Satz 1 an etwas unerwarteten Stellen auftauchen. Mutatis mutandis trifft diese Bemerkung auch auf den nächsten Satz zu:

7.4.2. Theorem T_2 ist genau dann eine überlegene Nachfolgertheorie für T_1, wenn T_2 eine gute Nachfolgertheorie für T_1 ist und

(a) $\neg T_1 \in T_2{}^*_{\neg A}$ *bzw.*
(b) $\neg A \rightarrow \neg T_1 \in T_2$ *bzw.*
(c) $A \in T_2{}^*_{T_1}$.

Beweis Nach den Definitionen 7.2.1–7.2.4 und mit der Symbolisierung natürlichsprachlicher Konditionalsätze aus Kapitel 6 ist T_2 genau dann eine überlegene Nachfolgertheorie für T_1, wenn eine der drei folgenden Bedingungen erfüllt ist:

(a) Sowohl $A^a\Box\!\!\rightarrow T_1$ als auch $\neg A^z\Box\!\!\rightarrow \neg T_1$ ist in T_2;
(b) sowohl $A^\circ\Box\!\!\rightarrow T_1$ als auch $\neg A^\circ\Box\!\!\rightarrow \neg T_1$ ist in T_2;
(c) sowohl $A^z\Box\!\!\rightarrow T_1$ als auch $\neg A^a\Box\!\!\rightarrow \neg T_1$ ist in T_2.

Die Behauptung folgt jetzt genauso wie im Beweis von Theorem 7.4.1. □

Ist T_2 eine gute oder überlegene Nachfolgertheorie für T_1 nach Fall (b) oder (c) der eben angeführten Theoreme, dann ist der Übergang von T_1 auf T_2 *nichtmonoton* in dem einfachen Sinn, daß $T_1 \notin T_2$ gilt. Nicht alles, was man als T_1-Vertreter „wußte", will man als T_2-Vertreter noch gelten lassen.

Wir wollen nun zwei Sätze C und D genau dann *bezüglich einer Theorie T theoretisch völlig äquivalent* nennen, wenn sowohl $T^*_C = T^*_D$ als auch $T^*_{\neg C} = T^*_{\neg D}$ gilt. Sei T_2 wieder eine überlegene Nachfolgertheorie für T_1. Wenn man die obigen Definitionen und Gärdenfors' komplette Kollektion von Rationalitätspostulaten für Revisionen (inklusive (T*7) und (T*8)) verwendet, dann kann man zeigen, daß die (nichttheoretischen) Anwendungsbedingungen A für T_1 bezüglich T_2 theoretisch völlig äquivalent mit der Vorgängertheorie T_1 selbst sind, und zwar unabhängig davon, ob der Fall (a), (b) oder (c) vorliegt. Dies ist der Inhalt des nächsten Satzes:

7.4.3. Theorem T_2 ist genau dann eine überlegene Nachfolgertheorie für T_1, wenn $T_2{}^*_A = T_2{}^*_{T_1}$ und $T_2{}^*_{\neg A} = T_2{}^*_{\neg T_1}$.

Beweis Bei Voraussetzung aller Gärdenforsscher Postulate für Revisionen gilt die Beziehung

(T*I) $(C \in T_2{}^*_D$ und $D \in T_2{}^*_C)$ genau dann, wenn $T_2{}^*_C = T_2{}^*_D$.

(s. Abschnitt 3.1). Mit (T*I) kann man aus den Theoremen 7.4.1 und 7.4.2 schließen, daß T_2 genau dann eine überlegene Nachfolgertheorie für T_1 ist, wenn eine der drei folgenden Bedingungen erfüllt ist:

(a) $A, T_1 \in T_2$ und $T_2{}^*_{\neg A} = T_2{}^*_{\neg T_1}$;

(b) $A, \neg A, T_1, \neg T_1 \notin T_2$ und $A \leftrightarrow T_1 \in T_2$;

(c) $\neg A, \neg T_1 \in T_2$ und $T_2{}^*_A = T_2{}^*_{T_1}$.

Sei T_2 nun eine überlegene Nachfolgertheorie für T_1, d.h. es gilt einer der Fälle (a), (b) oder (c). Im Fall (a) gilt wegen $A, T_1 \in T_2$ auch $T_2{}^*_A = T_2 = T_2{}^*_{T_1}$; im Fall (b) gilt wegen $A, \neg A, T_1, \neg T_1 \notin T_2$ und der materiale Äquivalenz von A und T_1 in T_2 sowohl $T_2{}^*_A = T_2{}^+_A = \{C: A \rightarrow C \in T_2\} = \{C: T_1 \rightarrow C \in T_2\} = T_2{}^+_{T1} = T_2{}^*_{T_1}$ als auch völlig analog $T_2{}^*_{\neg A} = T_2{}^*_{\neg T1}$; im Fall (c) gilt wegen $\neg A, \neg T_1 \in T_2$ auch $T_2{}^*_{\neg A} = T_2 = T_2{}^*_{\neg T_1}$. Damit gilt in allen Fällen $T_2{}^*_A = T_2{}^*_{T_1}$ und $T_2{}^*_{\neg A} = T_2{}^*_{\neg T_1}$, und die eine Richtung der Behauptung ist bewiesen.

Sei umgekehrt $T_2{}^*_A = T_2{}^*_{T_1} \wedge T_2{}^*_{\neg A} = T_2{}^*_{\neg T_1}$. Daraus folgt, daß A genau dann in T_2 ist, wenn T_1 in T_2 ist, und daß $\neg A$ genau dann in T_2 ist, wenn $\neg T_1$ in T_2 ist. Dies macht deutlich, daß die ersten Hälften (a) $A, T_1 \in T_2$, (b) $A, \neg A, T_1, \neg T_1 \notin T_2$ und (c) $\neg A, \neg T_1 \in T_2$ alle Kombinationsmöglichkeiten der Elementschaft von A und T_1 in T_2 abdecken (T_2 war

ja als konsistent vorausgesetzt). Die zweiten Hälften der Fälle (a) und (c) folgen direkt aus der Voraussetzung, während für Fall (b) noch die Gleichung $A \in T_2^*{}_A = T_2^*{}_{T_1} = T_2^+{}_{T_1} = \{C\colon T_1 \rightarrow C \in T_2\}$ und die analoge Gleichung für $\neg A$ und $\neg T_1$ zu berücksichtigen sind. Also gilt einer der Fälle (a), (b) oder (c), d.h. T_2 ist eine überlegene Nachfolgertheorie für T_1, und auch die andere Richtung der Behauptung ist bewiesen. □

Obwohl die Beweise alle ziemlich trivial sind, habe ich sie hier wiedergegeben, um besser sichtbar zu machen, woran die doch sehr starke volle T_2-theoretische Äquivalenz von T_1 mit ihren Anwendungsbedingungen A liegt. Die Richtung $(\neg)T_1 \in T_2^*{}_{(\neg)A}$ kommt vielleicht nicht mehr unerwartet, überraschend und kritisch erscheint aber die Richtung $(\neg)A \in T_2^*{}_{(\neg)T_1}$. Im Fall (b) der bedingten Erklärungen erlaubt einfach die Kontraponierbarkeit des materialen Konditionals z.B. den folgenden direkten Schluß: Wenn $\neg A$ das potentielle Scheitern von T_1 in T_2 erklärt, dann muß man, um T_1 „wahr zu machen", das Erfülltsein der Anwendungsbedingungen A für T_1 annehmen. In den Fällen (a) und (c), wo jeweils eine faktische und eine kontrafaktische (idealisierende) Erklärung zur Anwendung kamen, ist aufgrund der Analysen von Kapitel 6 z.B. folgende Argumentation möglich: Wenn A T_1 in T_2 kontrafaktisch erklärt, d.h. wenn in T_2 der Satz Wenn A der Fall wäre, dann wäre T_1 korrekt enthalten ist, dann erklärt faktisch $\neg T_1$ umgekehrt $\neg A$ in T_2: Weil T_1 nicht korrekt ist, sind auch die Anwendungsbedingungen A von T_1 nicht erfüllt.[10]

Man vergleiche hiermit Yoshidas (1977, S. 71) zweite Einschränkung der Wahlmöglichkeit von „Hilfsannahmen" A:

> An auxiliary assumption must be true, or if false, must be approximately true though strictly speaking false; and in a successful reduction or theoretical transformation if the more comprehensive theory is true, the validity of the approximately true assumption must be identical with that of the less comprehensive theory.

Der Teil nach dem Strichpunkt hat — abgesehen von dem Attribut „approximately true" — große Ähnlichkeit mit der in Theorem 7.4.3 genannten Bedingung. Allerdings finde ich bei Yoshida keine überzeugende Motivation. Den Teil vor dem Strichpunkt würde ich wieder kritisieren: Ist die

[10] Besser würde man dies so ausdrücken: Weil T_1 nicht korrekt ist, können die Anwendungsbedingungen A von T_1 nicht erfüllt sein. Die „Modalisierung" von gegenüber der Standardrichtung umgekehrten Konditional- und Kausalsätzen ist ein oft beobachtetes Phänomen.

Annahme, daß die Lichtgeschwindigkeit unendlich ist, approximativ wahr? Ist es die Annahme, daß Planeten Teile von Zweikörpersystemen mit der Sonne sind? (Vgl. Kapitel 8.1).

Trotz der abstrakten Argumente zugunsten überlegener Nachfolgertheorien wird der realistischste und häufigste Fall wohl der sein, der durch Theorem 7.4.1, Teil (c), charakterisiert wird. Die dort angeführte Bedingung besagt insbesondere

(7.4.4) $T_2{}^*{}_A \vdash T_1$.

Die Bedingung (7.4.4) erhält die Intentionen von (7.1.2) und (7.1.3) und stimmt dennoch mit (7.1.5) überein; sie ist, glaube ich, eine gute Explikation von (7.1.1).[11] Wir haben nun endlich die am Ende von Abschnitt 2.4.1 versprochene Alternative zu den Ideen der Bereichseinschränkung und der Approximation — siehe die Schemata (2.4.1)-(2.4.5) — formuliert.[12]

7.5 Eine intuitive Beschränkung für minimale Revisionen

Auf eine weitere, über Kapitel 3 hinausgehende Diskussion formaler Kriterien für minimale Revisionen kann im jetztigen wissenschaftstheoretischen Zusammenhang verzichtet werden. Nicht aber auf eine Diskussion dessen, was man mit dem Begriff „minimale Revision" intuitiv verbindet. Sei zum Beispiel T die Menge der logischen Konsequenzen des einzigen Axioms C. Formal wäre nun nichts dagegen einzuwenden, ja es ist sogar einleuchtend, wenn man $T^*{}_{\neg C}$ mit der Menge aller Konsequenzen von $\neg C$ identifiziert. Hat dann aber die Redeweise von einer minimalen Revision *von T* nicht etwas sehr Irreführendes an sich, weil von T ja „gar nichts mehr übrig gelassen wird"? Formal nein, intuitiv ja. Und wir sollten den intuitiven Begriff einer minimalen Revision nicht aus den Augen verlieren, wenn wir

[11] Mit (7.4.4) wird nun auch verständlich, wie die in Fußnote 2 erwähnte Disambiguierung bei mehreren möglichen Anwendungsbedingungen durch Disjunktionsbildung funktionieren kann. Seien A_1 und A_2 zwei Anwendungsbedingungen für T_1, die $T_2{}^*{}_{A_1} \vdash T_1$ und $T_2{}^*{}_{A_2} \vdash T_1$ erfüllen. Da $T_2{}^*{}_{A_1}$ und $T_2{}^*{}_{A_2}$ Theorien im Sinne von Definition 3.1.1 sind, gilt $T_1 \in T_2{}^*{}_{A_1} \cap T_2{}^*{}_{A_2}$. Aus der Gesamtheit der Gärdenfors-Postulate folgt aber $T_2{}^*{}_{A_1} \cap T_2{}^*{}_{A_2} \subseteq T_2{}^*{}_{A_1 \vee A_2}$ (siehe Lemma 3.1.3), also folgt $T_1 \in T_2{}^*{}_{A_1 \vee A_2}$, und es gilt $T_2{}^*{}_{A_1 \vee A_2} \vdash T_1$. Mithin ist die Disjunktion $A_1 \vee A_2$ gemäß (7.4.4) eine geeignete Anwendungsbedingung für T_1 (vom Standpunkt von T_2 aus gesehen).

[12] Es soll aber nicht verschwiegen werden, daß das Schema (7.4.4) noch mit Problemen behaftet ist. Einer der wichtigsten Punkte, die noch genauerer Untersuchung bedürfen, wurde in Kapitel 1, Fußnote 80, angesprochen.

Nachfolgerrelationen nach obigem Muster analysieren wollen. Meine These ist, daß eine intertheoretische Erklärung in T_2 nur dann zustande kommen kann, wenn die Sätze, bezüglich derer eine Revision auszuführen wäre, also A, $\neg$A, T_1 und/oder $\neg T_1$, nicht *mit den grundlegendsten Axiomen* von T_2 in Konflikt stehen. So kann etwa T_2 nur dann eine überlegene Nachfolgertheorie für T_1 sein, wenn weder A noch T_1 noch $\neg$A noch $\neg T_1$ die fundamentalen Axiome von T_2 über den Haufen wirft.

Zu erläutern ist noch, was es denn heißen soll, daß z.B. A „mit den grundlegendsten Axiomen von T_2 in Konflikt steht". Genügt es, wenn A nur *einem* Axiom aus dem Kern von T_2 widerspricht? Meiner Ansicht nach: Ja. Verdeutlichen wir uns dies kurz an einem Beispiel. Ein ganz fundamentale Einsicht der speziellen Relativitätstheorie (**SRT**) ist es, daß die maximale Ausbreitungsgeschwindigkeit für Signale c endlich ist. Deshalb muß jede adäquate Rekonstruktion von **SRT** den Satz „$c<\infty$" als Axiom oder unmittelbare Folgerung eines wichtigen Axioms enthalten. Davon zu sprechen, die mit „dem Geiste von **SRT**" durchaus unverträgliche Annahme $c=\infty$ könne **SRT** durch eine „minimale Revision" einverleibt werden, wäre deshalb ziemlich unangemessen. Dennoch, überraschenderweise wird **SRT** mit dieser Annahme fertig. Ein Vertreter von **SRT** kann genau sagen, was der Fall *wäre*, *wenn* c gleich ∞ wäre: Dann wäre die Newtonsche Mechanik richtig. (Weil er damit aber den ganzen Witz seiner Errungenschaften opfern müßte, würde ein wirklicher Vertreter von SRT wohl gleich hinzufügen: „Aber um Himmels Willen, die maximale Ausbreitungsgeschwindigkeit *ist* doch nicht unendlich".) Der **SRT**-Vertreter hat also ein sehr klares Bild davon, wie man **SRT** revidieren müßte, um die Annahme $c=\infty$ in **SRT** einzupassen. Es wäre zwar noch in einer detaillierten Untersuchung nachzuweisen, ist aber plausibel anzunehmen, daß sich die Gesetze einer **SRT**-Standardaxiomatisierung intuitiv adäquat und eindeutig nach ihrer theoretischen Wichtigkeit ordnen lassen, so daß die kontratheoretische Annahme $c=\infty$ in einer Revision à la Gärdenfors aufgenommen werden kann.[13]

[13] Felix Mühlhölzer hat mir plausibel gemacht, daß $c<\infty$ für **SRT** wichtiger ist als etwa $F=m_0/\sqrt{1-(v/c)^2}$ (wobei m_0 die Ruhemasse sei). Auf den ersten Blick sieht es so aus, als ob dann die in Kapitel 3 angegebenen Konstruktionen von Revisionen nicht mehr passen. Denn aus Theorem 3.5.13 kann man $\mathbf{SRT}^*_{c=\infty}$ dadurch erhalten, daß man diejenigen **SRT**-Sätze beibehält, die echt wichtiger sind als $c<\infty$, $c=\infty$ „addiert" und den deduktiven Abschluß bildet. Verlieren wir also die relativistische Kraftgleichung? Das muß nicht sein, denn wenn die Disjunktion $c<\infty \vee F=m_0/\sqrt{1-(v/c)^2}$ wichtiger ist als $c<\infty$, dann bleibt sie in $\mathbf{SRT}^*_{c=\infty}$ erhalten (und wir können nach Einsetzen von $c=\infty$ das Zweite Newtonsche Gesetz bekommen). Die Frage ist aber: Wieso soll $c<\infty \vee F=m_0/\sqrt{1-(v/c)^2}$ echt wichtiger sein als $c<\infty$, und wie sollten wir das wissen?

Wie dem auch sei, ich werde zwei „Grade der Fundamentalität" von Konflikten mit T_2 unterscheiden. Der erste, absolut gravierende, aber relativ leicht zu bewältigende Konflikt von A mit T_2 besteht dann, wenn A mit mindestens einem grundlegenden Axiom von T_2 unverträglich ist, wenn aber T_2 nichtsdestotrotz auf natürliche und eindeutige Weise eine Antwort auf die Frage geben kann, *was wäre, wenn A der Fall wäre.* Einen solchen Konflikt nennen wir im folgenden *auflösbar.* Die zweite, absolut und relativ *unauflösbare* Art von Konflikt ist dann gegeben, wenn T_2 diese Antwort nicht mehr zu geben vermag. Das Eintreten dieses Falls kann man sich so vorstellen, daß A mit noch grundlegenderen, also in der vorausgesetzten Ordnung der theoretischen Wichtigkeit noch höher eingestuften Axiomen und Sätzen von T_2 unvereinbar ist als im ersten Fall. Wenn A — bildlich gesprochen — von T_2 nichts mehr übrig läßt und dadurch jede Anwendbarkeit von T_2 zunichte macht, sollte keinesfalls mehr eine intertheoretische Erklärung zustande kommen können. Natürlich sind diese Bemerkungen reichlich dunkel und verschwommen, aber wir werden sehen, daß die angedeutete Unterscheidung für die Diskussion des Kepler-Newton-Falls (in Kapitel 8.1) relevant ist.

Auf eine besondere Art von auflösbarem Konflikt möchte ich noch kurz hinweisen. Sie ist immer dann gegeben, wenn man A intuitiv ganz deutlich kontra*faktisch*, nicht aber kontra*theoretisch* zu nennen geneigt ist. Es sei noch einmal daran erinnert, daß ich den Terminus „Theorie" in einem ganz umfassenden Sinn verstehe: Theorien sollen auch das Wissen über die Anfangs- oder Randbedingungen, über die in der wirklichen Welt obwaltenden „Umstände" beinhalten. Dies ist unter anderem dadurch begründet, daß es sehr schwer ist, einen klaren systematischen Unterschied zwischen theoretischem Wissen und bloß empirischem Wissen zu machen. Beispielsweise werden wir bei der Diskussion des Kepler-Newton-Beispiels sehen, daß die dortige Annahme von Einkörpersystemen, die prima facie harmlos kontrafaktisch aussieht, in Wirklichkeit kontratheoretisch in Bezug auf Newtons Gravitationstheorie ist. Oder es kann sich die Annahme irgendeines sehr kleinen Molvolumens für Kohlendioxid, eine nach dem idealen Gasgesetz mögliche Anfangsbedingung, die technisch scheinbar zufällig noch nicht realisiert war, im Lichte des van der Waalsschen Gasgesetzes als kontratheoretisch herausstellen. Selbstverständlich ist Kontratheoretizität ein theorienrelativer Begriff, aber wissenschaftliche Theorien sind untereinander auf so komplizierte Weise vernetzt und voneinander abhängig, daß es zu-

Ich kenne im Augenblick keine gute Antwort.

mindest sehr großer Mühen bedarf, um eine klare, philosophisch tragfähige Unterscheidung zwischen bloß kontrafaktischen und sogar kontratheoretischen Bedingungen zu finden.

Es ist fraglich, ob sich diese Mühe lohnt. Denn im Gärdenforsschen Revisionsmodell, welches ich meinen Analysen zugrundegelegt habe, wird Gesetzesartigkeit partiell durch die hohe Resistenz gegenüber Revisionen, d.h. durch eine große theoretische Wichtigkeit (oder feste „epistemische Verankerung") in umfassend aufgefaßten Theorien expliziert.[14] Dies ist meines Erachtens eine sehr gute Methode, „theoretische" Generalisierungen gegenüber „bloß empirischen" Generalisierungen auszuzeichnen. Unangenehmerweise wird hierbei wegen der prima facie undifferenzierenden Auffassung von „Theorie" formal auch das kontratheoretisch, was man eigentlich gerne als nur kontrafaktisch bezeichnen wollte. Dies muß man im Gedächtnis behalten. Obgleich intuitiv kontrafaktische Annahmen wohl stets zu auflösbaren Konflikten führen, gibt meine Dichotomie „auflösbarer vs. unauflösbarer Konflikt" kein exaktes Abbild der angedeuteten Dichotomie „kontrafaktisch vs. kontratheoretisch": Es gibt viele echt kontratheoretische Annahmen, die zu auflösbaren Konfliken führen, wie etwa die Annahme von Einkörpersystemen im Kepler-Newton-Beispiel oder die Annahme einer unendlichen Lichtgeschwindigkeit im Newton-**SRT**-Beispiel.

7.6 Idealisierung und Approximation

Es gibt einen alternativen Ansatz zur Bewältigung des in Abschnitt 7.1 skizzierten Problems. Er ist uns bereits aus Abschnitt 4.2, bekannt, und seine Hauptthese lautet, daß T_2 die Vorgängertheorie T_1 nur approximativ erklärt. Natürlich schließt die in diesem Kapitel dargebotene Idee nicht aus, daß Approximationsverfahren im kontrafaktischen Schließen des öfteren eine große Rolle spielen. Man beachte aber, daß in der Definition von (realer, bedingter und idealisierender) Erklärung keinerlei Approximationsterminologie auftaucht; insbesondere meint das Wort „wahr" in Definition 7.2.3 strikte, ganz genaue, strenggenommene Wahrheit. Das hier bereitgestellte begriffliche Instrumentarium ist allgemeiner als das der Approximation, indem es offenbar sinnvolle Idealisierungen im Zusammenhang von Theorien, die man am natürlichsten als *nichtquantitativ* auffaßt, und

[14] Da auch Beobachtungssätze unempfindlich gegenüber Revisionen durch faktische Neuinformation (nicht durch hypothetische Annahmen) sind, ist es für eine vollständige Explikation wohl nötig, Kriterien wie den Allgemeinheitsgrad eines Satzes mit zu berücksichtigen.

sinnvolle Idealisierungen *ohne approximative Gültigkeit* zu behandeln vermag. Laymon (1980) und Nowak (1980, S. 79–81) diskutieren einschlägige Beispiele: Newtons Idealisierung in seinem Experimentum crucis und die von der speziellen Relativitätstheorie angesetzte Idealisierung im Experiment von Michelson und Morley führen zu qualitativ falschen Voraussagen, während die idealisierenden Gesetze, welche die Brownsche Bewegung oder die Einflußgeschwindigkeit einer Flüssigkeit durch ein kleines Loch beschreiben, sich als weit entfernt von der (approximativen) Wahrheit erweisen. Je nachdem, wie man die Standards setzt für den Begriff der approximativen Gültigkeit, entscheidet sich, ob eine Idealisierung *nur* Idealisierung oder auch Approximation (im ersten Sinn) ist.[15] Es ist zum Beispiel gar nicht absurd, den Keplerschen Gesetzen abzusprechen, daß sie eine gute Annäherung an die Wahrheit sind. Scriven (1963, S. 123) zum Beispiel meint, wir müßten uns mit einer „unbequemen Tatsache" abfinden: „the planets do not obey Kepler's laws *by a long shot.*" (Hervorhebung von mir)[16]

Natürlich müssen Anwendbarkeit und Konsequenzen der hier vorgelegten Analyse vorurteilslos an Beispielen getestet werden. Ich will in Kapitel 8.1 zeigen, wie die Beziehung zwischen den Keplerschen Gesetzen der Planetenbewegung und der Newtonschen Gravitationstheorie mit dem oben entwickelten Instrumentarium analysiert werden kann, wobei Newtons Theorie eine überlegene Nachfolgertheorie für Keplers Gesetze sein und sie gleichzeitig als Idealisierung erweisen soll. Analoges versuche ich in Kapitel 8.2 am Beispiel des idealen und des van der Waalsschen Gasgesetzes. Die obigen Ideen allein haben wenig Wert, wenn man nicht weiß, *wie genau* Revisionen praktisch durchgeführt werden können. Es wäre eigentlich en détail aufzuweisen, daß und wie sich die in Kapitel 3 vorgestellte Relation der theoretischen Wichtigkeit in verschiedenen Fallbeispielen anwenden läßt.[17] Dieses hochgesteckte Ziel werden wir in diesem Buch nicht

[15] Umgekehrt kann es auch Approximationen geben, die nicht als Idealisierungen rekonstruierbar sind. Siehe etwa die in Kapitel 8.1 genannten Alternativen zu den Keplerschen Gesetzen im Rahmen der Cartesianischen Physik. Siehe auch Kapitel 8, Fußnote 39, zum Galilei-Newton-Beispiel.

[16] Genaugenommen heißt dies nicht, daß die Keplerschen Gesetze (**KGP**) weit entfernt sind von dem, was aus Newtons Gravitationstheorie (**NTG**) für das Planetensystem folgt. Normalerweise haben wir in diesem Buch nur die intertheoretische Relation zwischen **KGP** und **NTG** im Auge. Da jedoch der Unterschied zwischen **NTG** und **KGP** „den Löwenanteil" am Unterschied zwischen der Realität im Planetensystem und **KGP** ausmacht, ist Scrivens Bemerkung auch für die intertheoretische Lesart von „Approximation" einschlägig.

[17] Für einen ersten Eindruck der zu erwartenden Schwierigkeiten vgl. Fußnote 13.

mehr erreichen können, jedoch in Kapitel 8 erste Schritte in diese Richtung unternehmen. Ich werden zu zeigen versuchen, daß der Ansatz der idealisierenden Erklärung eine interessante und lehrreiche Alternative zum Ansatz der approximativen Erklärung ist.

Kapitel 8

Approximation versus Idealisierung: zwei Fallbeispiele

Im allgemeinen sind wissenschaftliche Gesetze und Theorien nicht ganz genau wahr. Soweit besteht unter den Wissenschaftstheoretikern weitgehende Einigkeit. Es gibt aber einen Unterschied in der Emphase. Normalerweise legt man die Betonung auf „genau“: wissenschaftliche Gesetze sind nicht *genau* wahr, aber sie weichen auch nicht allzu weit von den beobachteten Phänomenen ab (oder zumindest sollten sie es nicht). Das heißt, wissenschaftliche Gesetze sind *Annäherungen* oder *Approximationen* an die Wahrheit. Doch kann man, scheinbar widersinnig, im ersten Satz dieses Abschnitts die Hervorhebung auch so ansetzen: Wissenschaftliche Gesetze sind nicht genau wahr, also insbesondere *nicht wahr*. Was kann hinter dieser Betonung stecken? Sie basiert auf der Idee, daß wissenschaftliche Gesetze uns — genau — sagen, was der Fall wäre, wenn man gewisse vereinfachende Annahmen macht (machen könnte). Häufig geht es in diesen Annahmen um das Ausblenden der vielen „Störfaktoren“, die leider mehr oder weniger allgegenwärtig sind, und deshalb sind die Anfangs- oder Randbedingungen, die mit diesen Annahmen beschrieben werden, in der Wirklichkeit nicht erfüllt. Sie sind *kontrafaktisch*, unter Umständen *kontratheoretisch*, und können somit in der Regel verantwortlich gemacht werden für die Falschheit wissenschaftlicher Gesetze. In diesem Sinne werde ich von nun an

wissenschaftliche Gesetze und Theorien als *Idealisierungen* bezeichnen.

Die beiden skizzierten Auffassungen von wissenschaftlichen Gesetzen und Theorien, die ja vorderhand überhaupt nichts miteinander zu tun haben, werden in der Literatur leider nicht so sauber auseinandergehalten, wie man es sich wünschen würde. Mit wenigen Ausnahmen[1] haben Wissenschaftstheoretiker die Neigung, Argumente der Approximationsperspektive mit denen der Idealisierungsperspektive zu vermengen. Natürlich besteht der heuristische und praktische Wert vieler Idealisierungen in ihrer Eigenschaft, in gewissen Bereichen approximativ richtige Ergebnisse zu liefern. Das ist aber noch keine Garantie dafür, daß alle oder auch nur die Mehrzahl der idealisierenden Annahmen zu Voraussagen führen, die der Wahrheit nahekommen; ebensowenig darf man davon ausgehen, daß alle oder auch nur die Mehrzahl der approximativ gültigen Gesetze als auf kontrafaktischen Bedingungen basierend rekonstruiert werden können. Das begriffliche Verhältnis von Idealisierungen und Approximationen ist eine sehr heikle Angelegenheit, und zu einer wirklich befriedigenden Klärung bedarf es sicherlich noch vieler Arbeit.

Ich will in diesem Kapitel nicht daran gehen, abstrakte Argumente für mein Verständnis von „Idealisierung" zu präsentieren, noch werde ich allgemein-philosophische Gründe und Folgen der Abkoppelung dieses Begriffs vom Approximationsbegriff anführen. In groben Zügen werde ich dies in Kapitel 9 versuchen. Aber es ist vielleicht nützlich, schon einmal vorauszuschicken, welches Bild hinter meiner Vorstellung von Idealisierungen steckt. Idealisierung ist, um es mit einem Schlagwort zu sagen, *Dekomposition*. Aus der Vielzahl wirkender und wechselwirkender Faktoren, die man *praktisch* nie ganz ausschalten kann, werden *in Gedanken* ganz wenige, zuweilen nur ein einziger Faktor ausgesondert, um einfache und rechnerisch handhabbare Gleichungen für genau diese wenigen Faktoren zu erhalten.[2] Es ergibt sich als natürliches Forschungsprogramm die Aufgabe, zu untersuchen, welche zusätzlichen Auswirkungen die zahlreichen anderen, zunächst vernachlässigten Faktoren haben. Quantitative Übereinstimmung mit der Erfahrung spielt dabei zumindest am Anfang überhaupt keine Rolle, da

[1] Als positive Beispiele kann man M. Scriven, D. Shapere, F. Suppe und R. Laymon nennen. Nancy Cartwright ist bekannt für ihre Doktrin, daß die grundlegenden Gesetze der Physik wie zum Beispiel Newtons Gravitationsgesetz und Coulombs Gesetz schlichtweg falsch sind: „These two laws are not true; worse, they are not even approximately true." (Cartwright 1983, S.57)

[2] Der Begriff des Faktors soll hier nicht weiter eingegrenzt werden. Auch Objekte können als „Faktoren" interpretiert werden, wie die Idealisierungsanalyse des Kepler-Newton-Beispiels durch Ein- und Zweikörpersysteme zeigen wird.

man ja von vornherein weiß, daß noch viele andere Faktoren hinzugerechnet werden müßten. In diesem Sinne kann man den Lakatosschen Wissenschaftler verstehen, der sich, wenn er erst einmal eine positive Heuristik hat, „auf die Couch legt, die Augen schließt und die Daten vergißt" (Lakatos 1970, S. 135). Zu Beginn ist er zufrieden, wenn er weiß, was der Fall *wäre, wenn nur* der So-und-so-Faktor am Werk *wäre*. Dann versucht er, Schritt für Schritt die Effekte einiger anderer Faktoren „aufzuaddieren" oder zu kombinieren. Er träumt davon, am Ende einmal eine Gleichung zu finden, die ihm sagt, wie die Gesamtheit der verschiedenen Faktoren zusammenwirkt (weiß aber auch, daß dies ein Wunschtraum bleiben wird).

Mein Ziel in diesem Kapitel ist es, ausgehend von den im letzten Kapitel entwickelten Begriffen und Ergebnissen den Unterschied zwischen Approximationen und Idealisierungen anhand zweier einfacher und wohlbekannter Beispiele zu illustrieren: des Verhältnisses zwischen Keplers Gesetzen für die Planetenbewegung und Newtons Theorie der Gravitation und des Verhältnisses zwischen dem idealen und dem van der Waalsschen Gasgesetz. Wir werden hier also statt „Approximation" und „Idealisierung" simpliciter die intertheoretischen Versionen dieser Begriffe untersuchen (was auch schon deshalb angezeigt scheint, weil Wissenschaftler nicht über „die Wahrheit" verfügen, sondern höchstens über immer bessere Theorien), wobei Approximationen im interessanteren zweiten Sinn verstanden werden (vgl. Abschnitt 2.4). Mein Augenmerk in den Fallbeispielen werde ich ganz besonders auf die Frage legen, inwiefern die „Vorgängertheorie" aus ihrer „Nachfolgertheorie" approximativ bzw. kontrafaktisch ableitbar ist.

8.1 Die Keplerschen Gesetze und Newtons Gravitationstheorie

8.1.1 Die Inkonsistenz zwischen den Keplerschen Gesetzen und Newtons Gravitationstheorie

Wohl seit den Arbeiten von Duhem (1906) und Popper (1957), spätestens jedoch seit Feyerabend und Lakatos, wird allgemein anerkannt, daß Keplers Gesetze der Planetenbewegung (**KGP**) und Newtons Theorie der Gravitation (**NTG**) *strenggenommen unvereinbar* sind. Es lohnt sich, die explizitesten Referenzstellen für diese Behauptung einmal Revue passieren zu lassen. Duhem (1906/1978, S. 257) läßt es in der Zusammenfassung seines Arguments gegen eine weitverbreitete Deutung des Kepler-Newton-Falls an

Klarheit nicht fehlen:

> Das Prinzip der allgemeinen Gravitation kann daher keineswegs durch Generalisation und Induktion aus den Beobachtungstatsachen, die Kepler formuliert hatte, abgeleitet werden, es widerspricht vielmehr in aller Form diesen Gesetzen. Wenn die Theorie von Newton richtig ist, sind die Keplerschen Gesetze notwendigerweise falsch.[3]

Popper (1957, S. 28f) verweist auf Duhem, seine Sichtweise ist aber etwas differenzierter (und deutet in die geiche Richtung wie die unten durchgeführte Idealisierungsanalyse):

> Es ist offensichtlich, daß, nach Newtons Theorie, Keplers Gesetze nur annähernd gültig sind — das heißt ungültig —, wenn wir die gegenseitige Anziehung zwischen den Planeten berücksichtigen. Aber es gibt tiefer liegende Widersprüche zwischen den beiden Theorien als diese. ... [Denn es] widerspricht, selbst wenn wir alles über die gegenseitige Anziehung zwischen den Planeten vergessen, Keplers drittes Gesetz [a^3/T^2=konstant] Newtons Theorie, die [$a^3/T^2=m_0+m_1$] ergibt.[4]

Was Feyerabend anbetrifft, so wird meist Feyerabend (1962, S. 45 und 92f) als Referenzstelle angegeben, wo jedoch nichts Neues steht. Lehrreicher ist Feyerabend (1965a, S. 155):

---

[3] Was Newton in der Tat aus den Keplerschen Gesetzen „ableiten" konnte, war laut Duhem (1906/78, S. 255f) „in sehr präziser Form" folgender Lehrsatz: „Wenn die Sonne der Beziehungspunkt für alle Kräfte ist, ist jeder Planet einer gegen die Sonne gerichteten Kraft unterworfen, die proportional der Masse des Planeten und umgekehrt proportional dem Quadrat seines Abstandes von der Sonne ist. Was dieses Gestirn betrifft, so ist es, indem es als Beziehungspunkt gewählt wird, keiner Kraft unterworfen." Natürlich kann man auch diesen Lehrsatz, der die Begriffe „Kraft" und „Masse" enthält und der insbesondere die Abhängigkeit der Kräfte von der Wahl eines festen Bezugspunktes berücksichtigt, *nicht* wirklich (auch nicht induktiv) aus Keplers kinematischen Gesetzen ableiten, sondern bedarf hierzu des Rahmens der Newtonschen Dynamik.

[4] Popper hatte gegenüber Duhem den Vorteil, auf die Darstellung bei Born (1949) — laut Lakatos (1978/82b, S. 78) „der erste Autor in der Geschichte der Wissenschaft, der Newtons Deduktion rekonstruiert hat" — aufbauen zu können. Deshalb gibt Popper (1957, S. 26, Fußnote 7) auch korrekt an, was man aus Keplers Gesetzen *allein* ableiten kann, nämlich im wesentlichen die unten angeführte Gleichung (8.1.7) (ohne die ganz linke Seite mit „**F**", „m" gekürzt). — Übrigens erschien gleichzeitig mit Borns Rekonstruktion eine ausführlichere und sehr schöne Abhandlung des formal-deduktiven Verhältnisses zwischen **KGP** und **NTG** in Toeplitz (1949, S. 142–164).

> Newton introduces his law of gravitation in a manner which suggests that it was a direct consequence of his own inductivist rules of procedure. ... So convincing was his presentation that many later thinkers, including Drude and Born, believed the law of gravitation to be nothing but a mathematical transcription of Kepler's laws and that Hegel denied to Newton the discovery of anything over and above Kepler. Since Duhem's investigations, and especially since the general theory of relativity and the publication of Einstein's papers on the theory of knowledge, it has become clear that this is a travesty of the actual situation. Not only does Newton's theory transcend the domain of observation; it also contradicts the observational laws that were available when the theory was first suggested. It is therefore quite impossible to obtain it by inductive generalization, which leaves the „facts" unchanged; and if it *did* seem possible to obtain it in this fashion, then this was due to the omission from the argument of some essential premises.[5]

Auch bei Lakatos denkt man immer gleich an seine bekannteste Arbeit, obwohl die einschlägigen Bemerkungen dort (Lakatos 1970, S. 135f, 147Fn, 152Fn, 158) nicht besonders aufschlußreich für den Kepler-Newton-Fall sind. Interessantere Stellen findet man in erst posthum veröffentlichten Aufsätzen:

> Newtons Geist mit seinen gegeneinander abgeschotteten Abteilungen läßt sich gar nicht besser kennzeichnen als durch die Gegenüberstellung des Methodologen Newton, der seine Gesetze aus Keplers 'Erscheinungen' *abgeleitet* haben wollte, und des Wissenschaftlers Newton, der genau wußte, daß seine Gesetze

[5] Wie kann das Auslassen einer (wenn auch wesentlichen) Prämisse zu einer Inkonsistenz führen? Hat Feyerabend etwa die Vision einer nichtmonotonen Logik? Nein. Was Feyerabend meint, mag man aus dem Argumentationsgang in seiner Fußnote 44 ersehen: „In the mathematical argument that is usually presented, the premises are Kepler's laws and the conclusion is the formula kMm/r^2. The mathematics is, of course, faultless. Not so the interpretation of the premises and the conclusion. If we interpret the premises as expressing Kepler's laws, then we cannot base an inductive argument on them because they are *false*. If, on the other hand, we interpret these formulae as describing the behavior of a single mass point in the close neighborhood of a much larger mass, then their truth cannot be denied. However, in this case the formulae describe an imaginary process, and not what is going on in the real solar system. The conclusion will then also describe such an imaginary process."

> diesen Erscheinungen *geradewegs widersprachen*. ... Newton hat die *Negation* seiner neuen Theorie zu deren eigener Grundlage gemacht. (Lakatos 1978/82a, S. 225f)[6]

„Geradewegs" widerspricht **NTG KGP** aber nur dann, wenn man **NTG** auf die Planetenbewegungen in unserem Sonnensystem bezieht. Das heißt, wenn man die tatsächlich gegebenen *Anfangs- oder Randbedingungen* im Sonnensystem zu **NTG** dazunimmt, um von der Abstraktheit von **NTG** zu so etwas wie „Newtons Gesetzen der Planetenbewegung" zu kommen, die auf derselben konkreten Ebene wie **KGP** anzusiedeln sind. Die entscheidende Rolle von solchen Hilfsbedingungen („auxiliary statements") in Erklärungen und Voraussagen wurde für den Kepler-Newton-Fall ganz besonders von Putnam (1974) betont.[7] Faßt man aber, wie Scheibe (1973) und die in seinem Gefolge schreibenden Autoren, auch **KGP** als Gesetze über die Bewegungen von allgemeinen n-Körpersystemen auf, dann ist **NTG** — wie die sog. Karussellmodelle klar machen (s. Scheibe 1973) — nur dann unvereinbar mit **KGP**, wenn **NTG** zusammen mit irgendeiner der überwältigenden Mehrheit aller denkbaren Randbedingungen genommen wird. Zu dieser Mehrheit gehören selbstverständlich auch die Randbedingungen in unserem Planetensystem. Ohne diese Vorsichtsmaßnahme hat Scheibe recht mit seiner Ablehnung der Duhemschen Unverträglichkeitsthese hinsichtlich **KGP** und **NTG**.

---

[6] In Lakatos (1978/82b, S. 78) macht er aber folgendes beachtliche Zugeständnis: „Newton hat seine Theorien aus den Tatsachen *beinahe abgeleitet*"; vgl. auch die eingehende Analyse in Lakatos (1978/82b, S. 95–97).

[7] Putnam allerdings konzentriert sich auf die vermeintliche Ableitung von **KGP** aus **NTG** (und nicht auf den Widerspruch zwischen ihnen). Immerhin ist er einer der ganz wenigen, die versucht haben, die Hilfsbedingungen für diese Ableitung *explizit* anzugeben. Sein Vorschlag, gedacht „as a first approximation" (Putnam 1974, S. 225), lautet:

(I) No bodies exist except the sun and the earth.
(II) The sun and the earth exist in a hard vacuum.
(III) The sun and the earth are subject to no forces except mutually induced gravitational forces.

Daraus bekommt man aber, wie im Laufe dieses Kapitels klar werden wird, keineswegs alle Keplerschen Gesetze; es folgen die ersten beiden, und wenn das dritte überhaupt in irgendeiner Form folgt, dann nur in eingeschränkter Anwendung auf die Erde in theoretisch möglichen Bahnen oder auf theoretisch denkbare Planeten mit gleichen Massen (und nur in einer modifizierten Form, was sich hier allerdings gleich bleibt — siehe (KGP3) und (KGP3*) unten). (I)-(III) sind außerdem „known to be false" (S. 226, 228); generell sind Hilfsbedingungen nach Putnam „far more subject to revision than the theory" (S. 226), „highly risky" (S. 228) und „not fixed ..., but depend[ent] upon the context" (S. 236).

Wegen der — richtig verstandenen — Inkonsistenz von **KGP** und **NTG** können die verschiedentlich in populären Darstellungen der Wissenschaftsgeschichte und in Lehrbucheinleitungen aufgestellten Behauptungen, daß **NTG** aus **KGP** „abgeleitet" wurde oder daß umgekehrt **KGP** aus **NTG** „folgen", nicht richtig sein. Wenn **NTG** wahr ist, dann sind **KGP** falsch, und umgekehrt. Aber natürlich besteht eine enge Verbindung zwischen **KGP** und **NTG**, und es ist deshalb naheliegend anzunehmen, daß wir es hier mit einem Fall von intertheoretischer Approximation und/oder intertheoretischer Idealisierung zu tun haben. In der Tat wird sich herausstellen, daß **KGP** dem, was aus **NTG** unter Verwendung real vorliegender Randbedingungen für das Planetensystem folgt, sehr nahe kommen und daß Keplers Gesetze — mindestens in einer modifizierten Form — eine Idealisierung darstellen, wenn man sie aus dem Newtonschen Blickwinkel betrachtet. Ein Hauptzweck dieses Teilkapitels ist es, die Approximationsperspektive und die Idealisierungsperspektive in ihrem spezifischen Bezug auf das Kepler-Newton-Beispiel genau auseinanderzuhalten und gegenüberzustellen.

Im nächsten Abschnitt werde ich einige Vorschläge diskutieren und kritisieren, die alle in der Nachfolge von Scheibes grundlegender Arbeit (1973a) entstanden und darauf abzielen, die Relation zwischen **KGP** und **NTG** als ein paradigmatisches Beispiel von intertheoretischer Approximation im zweiten Sinn zu erweisen. Dann wird das Verhältnis zwischen **KGP** und **NTG** als ein typischer Fall von Idealisierung analysiert. Dabei stütze ich mich auf den in Kapitel 7 entwickelten qualitativen Ansatz, der verständlich macht, auf welche Weise idealisierende Annahmen eine zentrale Rolle spielen bei der gleichzeitigen Erklärung sowohl der theoretischen Signifikanz als auch des Scheiterns einer Vorgängertheorie durch eine überlegene Nachfolgertheorie. Zwei Versuche unternehme ich, um nachzuweisen, daß **NTG** wirklich eine überlegene Nachfolgertheorie von **KGP** ist. Die erste Idealisierung bezieht sich auf sog. „Einkörpersysteme", wird aber als mangelhaft erkannt, da sie vom Standpunkt von **NTG** aus kaum zu interpretieren ist. Die „Zweikörpersysteme" in der zweiten Idealisierung hingegen werden uns gute Dienste leisten, zugleich aber zeigen, daß Keplers drittes Gesetz modifiziert werden muß. Abschließend werde ich dafür plädieren, daß die Idealisierungsperspektive zur Approximationsperspektive gewissermaßen komplementär und dieser vielleicht sogar vorzuziehen ist.

8.1.2 Das Kepler-Newton-Beispiel als ein Fall von Approximation

Erhard Scheibe (1973a) ist der Autor einer wegweisenden Untersuchung zum Kepler-Newton-Beispiel. Er betrachtet Zustandsbeschreibungen von n-Körper-Systemen ($n \geq 2$), zusammen mit Konstanten $m_i \in \mathbb{R}^+$ ($i=1,\ldots,n$) oder $k \in \mathbb{R}^+$, welche die folgenden kinematischen Differentialgleichungen erfüllen:

(NTG) $\ddot{\mathbf{r}}_i = - \sum_{j=1,\ldots,n,\, j\neq i} (m_j/|\mathbf{r}_i-\mathbf{r}_j|^3)\cdot(\mathbf{r}_i-\mathbf{r}_j)$, $i=1,\ldots,n$;

beziehungsweise

(KGP) (a) $\ddot{\mathbf{r}}_i = - (k/|\mathbf{r}_i-\mathbf{r}_1|^3)\cdot(\mathbf{r}_i-\mathbf{r}_1)$, $i=2,\ldots,n$,

(b) $\ddot{\mathbf{r}}_1 = 0$,

(c) $\frac{1}{2}\,|\dot{\mathbf{r}}_i-\dot{\mathbf{r}}_1|^2 - k/|\mathbf{r}_i-\mathbf{r}_1| < 0$, $i=2,\ldots,n$.

Hierbei sind die $\mathbf{r}_i$'s die auf ein Inertialsystem bezogenen Ortsfunktionen der betrachteten Körper (Vektoren sind immer fettgedruckt), die Punktnotation bezeichnet Ableitungen nach der Zeit, und die m_j's sind die Massen der Körper. Man beachte, daß in der Formulierung von (NTG) die Einheiten als so gewählt vorausgesetzt werden, daß die Gravitationskonstante G den Wert 1 erhalten kann. Ich setze etwas Vertrautheit mit **NTG** und **KGP** voraus. Es sollte also nicht unerwartet kommen, wenn Systeme, die (NTG) erfüllen, *Newtonsche Systeme* genannt werden. Weniger selbstverständlich ist es, daß (KGP)(a)–(c) erfüllende Systeme als *Keplersche Systeme* bezeichnet werden, da Kepler seine Gesetze ja auf völlig andere Weise formulierte und da (KGP)(a), (b) und (c) natürlich überhaupt nicht mit Keplers erstem, zweitem und drittem Gesetz übereinstimmen. Die Bedingungen (a) und (b) kann man (erst) im Lichte von (NTG) leicht interpretieren, (c) ist eine Bedingung der totalen Energie eines Körpers im System, welche geschlossene Bahnen (im Gegensatz zu offenen parabolischen oder hyperbolischen Bahnen) garantiert. In diesem Abschnitt werden wir Scheibes Schachzug, von (KGP)(a)–(c) als der modernen, Galilei-invarianten Formulierung der Keplerschen Gesetze auszugehen, noch unhinterfragt lassen. Es ist wichtig, darauf hinzuweisen, daß (KGP)(a)–(c) keine Gesetze über Planetenbewegungen sind, sondern Gesetze über die Bewegungen in beliebigen n-Körper-Systemen, wobei der Körper i=1 als „Sonne" ausgezeichnet ist. Eine kinematische Zustandsbeschreibung (d.h. eine Menge von n Ortsfunktionen) heißt *Newtonsch* (*Keplersch*) genau dann, wenn es geeignete Konstanten m_i (bzw. k) gibt, so daß die Zustandsbeschreibung zusammen mit den (der) Konstanten ein Newtonsches (Keplersches) System bildet.

Scheibes wichtigste Behauptungen, welche er für n=2 als zutreffend beweisen, für n>2 jedoch nur vermuten kann, sind grob gesagt die folgenden. Sei $\varepsilon>0$ beliebig vorgegeben. Für jede Keplersche Zustandsbeschreibung gibt es dann eine Newtonsche Zustandsbeschreibung, welche von der ersteren „um höchstens ε" abweicht, d.h. $|\mathbf{r}_i^N-\mathbf{r}_i^K|<\varepsilon$ für alle i=1,...,n,[8] wobei $\mathbf{r}_i^N$ ($\mathbf{r}_i^K$) die Ortsfunktion des i-ten Körpers in der Newtonschen (Keplerschen) Zustandsbeschreibung ist. Umgekehrt gibt es eine „Spezialisierung" von **NTG** (d.h. **NTG** plus eine Menge von Zusatzbedingungen), so daß es für jede Zustandsbeschreibung, die mit dieser Spezialisierung verträglich ist, eine Keplersche Zustandsbeschreibung gibt, welche von der ersteren um höchstens ε abweicht.

Es ist ein großer Vorzug der Scheibeschen Darstellung, daß er für Zweikörpersysteme die Übergänge von Kepler zu Newton und umgekehrt konstruktiv angibt (Scheibe 1973, S. 114f). Die Idee soll nun in groben Zügen geschildert werden. Wir können hierzu gleich auf ganze Systeme statt nur auf Zustandsbeschreibungen Bezug nehmen. Seien nun $\mathbf{R}^N$ ($\mathbf{R}^K$) und $\mathbf{r}^N$ ($\mathbf{r}^K$) die Ortsfunktionen der „Sonne" bzw. des „Planeten" im jeweiligen Newtonschen (bzw. Keplerschen) System, M und m die (Newtonschen) Sonnen- und Planetenmassen und k die Kepler-Konstante. Von jedem gegebenen Kepler-System mit den Bestimmungsstücken $\mathbf{R}^K$, $\mathbf{r}^K$ und k kommt man zu einem ε-nahen Newton-System, indem man zunächst M und m so wählt, daß sie die Bedingungen

(8.1.1) $\quad m/M\ |\mathbf{r}^K-\mathbf{R}^K| < \varepsilon$

und

(8.1.2) $\quad M^3/(M+m)^2 = k$

erfüllen (man überlegt sich leicht, daß solche M und m existieren). Danach läßt man die Planetenbahn $\mathbf{r}^N=\mathbf{r}^K$ unverändert, „verwackelt" jedoch die Sonnenbahn mittels

(8.1.3) $\quad \mathbf{R}^N = \mathbf{R}^K - (m/M)(\mathbf{r}^K-\mathbf{R}^K)$

ein bißchen. Umgekehrt kommt man von jedem gegebenen Newton-System mit den Bestimmungsstücken $\mathbf{R}^N$, $\mathbf{r}^N$, M und m, d.h. mit Schwerpunkt $\mathbf{s} = (M\mathbf{R}+m\mathbf{r})/(M+m))$, welches als „Axiome" der Spezialisiserung die Zusatzbedingungen

(8.1.4) $\quad \frac{1}{2}\ |\dot{\mathbf{r}}^N-\dot{\mathbf{s}}|^2 - M^3/((M+m)^2|\mathbf{r}^N-\mathbf{s}|) < 0$

und

[8] „$|\mathbf{r}_i^N-\mathbf{r}_i^K|<\varepsilon$" ist natürlich die übliche Abkürzung für „$|\mathbf{r}_i^N(t)-\mathbf{r}_i^K(t)|<\varepsilon$ für alle $t\in\mathbb{R}$". Diese Schreibweise wird im folgenden öfter verwendet.

(8.1.5) $m/M\ |\mathbf{r}^N - \mathbf{R}^N| < \varepsilon$

erfüllt, zu einem ε-nahen Kepler-System grob gesagt durch die Umkehrung des ersten Prozesses: Man setzt $k = M^3/(M+m)^2$, beläßt die Planetenbahn mit $\mathbf{r}^K = \mathbf{r}^N$ und verwackelt die Sonne durch

(8.1.6) $\mathbf{R}^K = \mathbf{R}^N + (m/(M+m))(\mathbf{r}^N - \mathbf{R}^N)$

diesmal in die andere Richtung (Auflösung von (8.1.3) nach $\mathbf{R}^K$).

Die zwei für eine passende Spezialisierung nötigen Zusatzannahmen bestehen beim allgemeinen n-Körpersystem also in einem Gegenstück zu (KGP)(c) und in einer Bedingung des Inhalts, daß die Verhältniszahlen m_i/m_1 (i=2,...,n) klein sein sollen, klein genug zumindest, um eine gewisse Ungleichung zu erfüllen (vgl. (8.1.5)). Wir werden die erste Forderung im folgenden vernachlässigen.[9] Betreffs der zweiten Forderung, die bei fixer Sonnenmasse impliziert, daß die Massen m_i (beliebig) klein sind (ε kann ja beliebig klein vorgeben werden), sollte man beachten, daß sie wesentlich von ε abhängt.

Indem er approximative Erklärung mit der Inklusionsrelation zwischen „verschmierten" Modellmengen identifiziert, schließt Scheibe, daß es überraschenderweise eine direkte approximative Erklärung von **NTG** durch **KGP** gibt, während die erwartete approximative Erklärung von **KGP** durch **NTG** nur bedingt ist, d.h. auf zusätzliche Prämissen zurückgreifen muß.[10] Die intuitive Überlegenheit von **NTG** kommt erst zum Vorschein, wenn man die „Anwendungsbereiche" (Scheibe 1973b, S. 938) betrachtet: Die Spezialisierungen von **NTG** haben immer noch genügend Reichweite, um jede Keplersche Zustandsbeschreibung abzudecken (wobei der Genauigkeitsgrad ε beliebig gewählt werden darf), aber **KGP** deckt nur einen winzigen Bruchteil (das genaue Ausmaß ist abhängig von ε) der Newtonschen Zustandsbeschreibungen ab. Nach Scheibes (1973a, S. 117) Definitionsvorschlag besteht die *approximative Erklärung* einer Theorie T_1 durch eine Theorie T_2 (im Grade ε) aus zwei Teilen. Erstens müssen die Modelle einer Spezialisierung T_2' von T_2 (auch T_2 zählt als Spezialisierung

[9] Natürlich braucht man zur Herstellung einer intertheoretischen Relation zwischen **NTG** und **KGP** eigentlich Randbedingungen, die besagen, daß sich „die Planeten" nicht unendlich weit von der Sonne entfernen und daß sie nicht in die Sonne stürzen. Man kann Bedingungen solcher Art jedoch weglassen, indem man sich daran erinnert, daß **KGP** nur Aussagen *über Planeten* machen, und diese beiden Bedingungen als *definierende* Eigenschaften von Planeten auffaßt. Ebenfalls unterschlagen werde ich die theoretisch uninteressanten Annahmen, daß die Planeten nicht miteinander kollidieren oder durch sonstige Katastrophen aus ihrer Bahn geworfen werden. Es sei im übrigen angemerkt, daß keine dieser Annahmen kontrafaktisch zu sein scheint (s. aber Moser 1977–79).

[10] Vgl. hierzu auch Käsbauer (1976, S. 267–270).

ihrer selbst) eine Teilmenge der (im Grade ε) verschmierten Menge der T_1-Modelle ausmachen, und zweitens muß es für jedes T_1-Modell möglich sein, es (im Grade ε) so zu verschmieren, daß das verschmierte System ein Modell von T_2' (und damit auch von T_2 selbst) ist.

Ich will in diesem Abschnitt noch vier weitere Bearbeitungen des Kepler-Newton-Beispiels innerhalb des Approximationsansatzes betrachten, nämlich Moulines (1980), Mayr (1981b), Pearce und Rantala (1984b) und Balzer, Moulines und Sneed (1987, Abschnitt VII.3). Alle diese Aufsätze gehen von den Scheibeschen Formulierungen (NTG) und (KGP)(a)–(c) aus und sind im wesentlichen Versuche, Scheibes eher informelle Darstellung in einen strengen mathematischen und metatheoretischen Rahmen einzupassen. Insofern scheinen sie von geringerer Bedeutung für das Ziel dieses Teilkapitels zu sein. Aber sie beinhalten auch einige interessante Unterschiede gegenüber Scheibe, und um einen aktuellen Überblick über die Approximationsperspektive auf das Kepler-Newton-Beispiel zu bekommen, sollten wir sie ohnehin zur Kenntnis nehmen.

C.U. Moulines (1980) war der erste, der Scheibes Analysen in einen präzisen technischen Hintergrund eingebettet hat. Er ist ein Vertreter des Sneed-Stegmüllerschen „Strukturalismus" in der Wissenschaftstheorie, und er wählt uniforme Strukturen, oder einfach „Uniformitäten", als zentralen topologischen Begriff. Während Scheibe an der Idee einer deduktiv-nomologischen Erklärung à la Hempel orientiert war, ist Moulines bestrebt zu zeigen, daß auf das Kepler-Newton-Beispiel das Muster einer intertheoretischen Approximation auf der Grundlage des Adams-Sneedschen Reduktionsbegriffs paßt.[11] Er zeigt als Theorem, daß sich drei zentrale Ergebnisse von Scheibe automatisch ergeben, wenn man annimmt, daß wir es (a) nur mit Zweikörpersystemen zu tun haben, daß (b) Newtons Theorie korrekt ist und daß (c) das Kepler-Newton-Beispiel ein Fall von „strikter ϱ_2^1-Approximation" ist (vgl. die Bedingungen P1–P3 in Moulines 1980, S. 406–411). Dies kann als gute Stützung von Moulines' These dienen, daß **KGP** und **NTG** tatsächlich in einer Relation der strikten ϱ_2^1-Approximation stehen.[12]

[11] Wie wir in Kapitel 1 gesehen haben, blieb dieser Begriff auch innerhalb der strukturalistischen Schule nicht ohne Widerspruch. vgl. insbesondere die Diskussion der Beziehung zwischen Adams-Sneedscher Reduktion und deduktiv-nomologischer Erklärung in Abschnitt 1.2 einerseits, andererseits Fußnote 11 von Kapitel 1, aus der hervorgeht, daß Moulines' (1980, S. 400–403) Approximation in gewissem Sinn genau auf die Umkehrung dieser Reduktionsrelation hinausläuft.

[12] Natürlich *folgt* Moulines' These nicht — wie er behauptet (1980, S. 411) — aus dem erwähnten Theorem.

Leider gibt es einige Ungenauigkeiten, die die Präsentation von Moulines' Idee beeinträchtigen. Obgleich Moulines (1980, S. 395) beansprucht, eine „Reformulierung“[13] von Scheibes Analyse zu liefern, bedient er sich offensichtlich einer anderen Reduktionsrelation für Zweikörpersysteme (Mehrkörpersysteme werden nicht berücksichtigt). Zwar verwendet er die von Scheibe benutzte Gleichung (8.1.2), „verwackelt“ aber weder Planet noch Sonne, d.h. er setzt einfach $\mathbf{r}^{\mathrm{N}}=\mathbf{r}^{\mathrm{K}}$ und $\mathbf{R}^{\mathrm{N}}=\mathbf{R}^{\mathrm{K}}$ für zwei in der Reduktionsrelation stehende Keplersche bzw. Newtonsche Systeme. Damit aber müßte Moulines erst noch zeigen, daß auf diese Weise tatsächlich Kepler- und Newton-*Modelle* gepaart werden,[14] und außerdem weiß man nun nicht mehr recht, warum eigentlich die Reduktion eine *approximative* genannt wird. Unabhängig von diesen Fragen kann man, da Moulines Scheibes topologische Basis für die Fallanalyse übernimmt, ziemlich leicht nachweisen, daß das Kepler-Newton-Beispiel kein Fall von *strikter* Approximation im Sinne von Moulines ist.[15] Weiter ist seine zweite Reformulierung von Scheibes Ergebnissen (Moulines 1980, S. 407, (S2); vgl. Scheibe 1973a, S. 114f) nicht korrekt.

Wir wollen uns hier nicht auf die Details einlassen, doch das, was der Grund für Moulines' Irrtümer sein könnte, ist vom Blickwinkel dieses Kapitels aus interessant: Es gibt keine ausgezeichnete Spezialisierung von Newtons Theorie, die wir für *alle* beliebigen Genauigkeitsgrade ε gleichzeitig hernehmen können. Zwar kann man für beliebig gewähltes ε Zusatzbedingungen finden, so daß alle Systeme, welche **NTG** und diese Bedingungen erfüllen, sich nur im Grad ε von einem Kepler-System unterscheiden. Wie wir aber oben gesehen haben, manipuliert eine dieser Bedingungen das Verhältnis von Planeten- und Sonnenmassen und hängt dabei wesentlich von dem von außen vorgegebenen ε ab. Aus diesem Grunde muß die durch diese Bedingungen bewirkte „Spezialisierung“ von **NTG** als eine Ad-hoc-Theorie angesehen werden.

D. Mayrs (1981b) dichtbepackter Aufsatz ist der nächste, der der von Scheibe eingeschlagenen Richtung folgt. Seine allgemeine metatheoretische Grundlage ist G. Ludwigs Wissenschaftstheorie mit der intertheoretischen Relation der „verschmierten Einbettung“, und die von ihm benutzten ma-

[13] Stegmüller (1986, S. 8, 247) spricht von einer strukturalistischen „Übersetzung“.

[14] Zu erwarten sind Schwierigkeiten mit dem dritten Keplerschen Gesetz.

[15] Diese Behauptung setzt den Begriff der strikten Approximation voraus, wie er von Moulines im Beweis seines Theorems 2 (1980, S. 410, Zeile 1–3) benutzt wird. Seine Definition der strikten Approximation (S. 403), die durch viele Abkürzungen hochkondensiert und sehr schwer zu entschlüsseln ist, scheint mir nicht ganz äquivalent zu sein.

thematischen Konzepte sind (separierte) uniforme Räume und deren Vervollständigungen. Wir brauchen uns wieder nicht um die Feinheiten des Mayrschen Ansatzes zu kümmern und werden nur die entscheidende Abweichung von Scheibes Resultaten beachten. Mayr bemerkt, daß es im Falle von n-Körper-Systemen mit n>2 Keplersche Modelle gibt, die von keinem Newtonschen Modell für alle Zeiten approximiert werden können.[16] Es ist also vernünftig (und im Einklang mit der Praxis des Physikers), sich auf kompakte Zeitintervalle zu beschränken, d.h. die Ungleichung $|\mathbf{r}_i^N(t)-\mathbf{r}_i^K(t)|<\varepsilon$ für alle i=1,...,n und alle t in einem *kompakten* Intervall $T\subseteq\mathbb{R}$ als Basis für die Topologie im Raum der kinematischen Zustandsbeschreibungen zu verwenden. Sei n>2 und $\varepsilon>0$ gegeben. Für jedes kompakte Zeitintervall T, so sagt Mayrs zentrales Theorem, können dann alle Keplerschen Zustandsbeschreibungen von n Körpern bis auf ε durch Newtonsche Zustandsbeschreibungen *während T* approximiert werden. Da schon Dreikörpersysteme nur äußerst schwer — wenn überhaupt — explizit auszurechnen sind, ist Mayrs Beweis im Gegensatz zu Scheibes Behandlung des Zweikörperproblems inkonstruktiv. Die Approximation wird aber wieder dadurch erreicht, daß man die Massen „klein werden läßt" („by letting the masses become small", Mayr 1981b, S. 68).

D. Pearce und V. Rantala (1984b) haben eine eigene allgemeine Metatheorie entwickelt und machen dabei Gebrauch von abstrakter Logik und Kategorientheorie. Die allgemeinste intertheoretische Relation, welche sie einführen, ist die Relation der „Korrespondenz", die sich zusammensetzt aus einer strukturellen Korrelation und einer mit dieser auf präzise Weise zusammenstimmenden Übersetzung (s. Abschnitt 1.8). Eine sehr wichtige Unterart ist die Grenzfallkorrespondenz, die das Schema darstellt, welches für **NTG** und **KGP** passen soll. Die Rolle von Moulines' und Mayrs topologischem Instrumentarium wird bei Pearce und Rantala von der Nonstandard-Analysis übernommen. Sehr verkürzt ausgedrückt, sind

[16] Die durch **NTG** vorauszusagende tatsächliche Abweichung des Planetensystems (damals war n=7) von Keplers Gesetzen war sehr bald klar. Man bemerkte, daß die gegenseitigen Störungen von Jupiter und Saturn, deren Umlaufzeiten sich ziemlich genau wie 2 zu 5 verhalten, nicht vernachlässigt werden dürfen. Es erhob sich eine heftige Kontroverse darüber, ob das Planetensystem deshalb auf lange Sicht überhaupt stabil sein könne. Diese Kontroverse, welche Newton zur Einführung Gottes in die Himmelsmechanik veranlaßte, galt erst 1799 als durch Laplace (ohne die „Hypothese Gott") entschieden: Das Planetensystem, so hatte er errechnet, sei für alle Zeiten stabil, alle Schwankungen seien periodisch (vgl. aber Moser 1977–79). Kann man deshalb sagen, Scheibe hätte seine Fehleinschätzung schon aufgrund dieser historischen Debatte bemerken müssen? Nein, denn im wirklichen Weltall kann man die Massen der Planeten eben nicht beliebig klein machen.

ihre Ergebnisse hinsichtlich **NTG** und **KGP** die folgenden: Für jedes Nonstandardmodell x von **NTG** (mit geschlossenen Planetenbahnen) derart, daß die Sonne eine *endliche, aber nicht-infinitesimale* und die Planeten eine *infinitesimale* Masse haben, ist der Standardteil von x ein (Standard-)Modell von **KGP**. Umgekehrt zeigen sie als eine Folgerung von Mayrs (1981b) Theorem, daß solch ein Nonstandardmodell x von **NTG** für jedes Standardmodell von **KGP** existiert.[17] Im wesentlichen verbürgen diese beiden Resultate zusammen mit der Existenz einer geeigneten Übersetzung, den Prozeß der Bildung von Standardapproximationen beschreibt, daß es in der Tat eine Grenzfallkorrespondenz von **KGP** auf **NTG** (relativ zu einer bestimmten unendlichen Logik) im Sinne von Pearce und Rantala gibt.

Bis jetzt sahen wir uns stets mit der Schwierigkeit konfrontiert, daß es keine einzelne, unabhängig bestimmbare Menge von Zusatzannahmen für **NTG** gibt, mit der man die je erwünschte intertheoretische Relation zwischen **NTG** und **KGP** für jeden beliebigen Genauigkeitsgrad erhalten kann. Der Ansatz von Pearce und Rantala ist hier keine echte Ausnahme: Die Verwendung der Nonstandard-Analysis kann überall in der Mathematik die ε-δ-Methode ersetzen und eliminieren und macht keinen inhaltlichen Unterschied für das spezielle Problem im Kepler-Newton-Beispiel. Die Situation ändert sich jedoch drastisch, wenn wir uns den neuesten Vorschlag von W. Balzer, C.U. Moulines und J.D. Sneed (1987) ansehen. Auch sie beginnen mit Scheibes Formulierungen (NTG) und (KGP).[18] Doch ihre Darstellung beansprucht, *eine* (exakte) Spezialisierung von **NTG** gefunden zu haben, eine Theorie namens „Spezielle gravitationelle klassische Partikelmechanik" (**GCPM***) (1987, S. 378f), die für jedes ε gleichermaßen geeignet ist. Darüber hinaus ist ihre Reduktionsrelation äußerst einfach. Sie beschränken sich auf Zweikörpersysteme und verknüpfen — anders als Scheibe, ebenso wie Moulines — Keplersche und Newtonsche Systeme, deren kinematische Zustandsbeschreibungen identisch sind, während sie — anders als Scheibe und Moulines — die Kepler-Konstante mit der Masse der „Sonne" gleichsetzen. Bezüglich der Masse des „Planeten" in Newtonschen Systemen ist in ihrer Reduktionsrelation nichts gefordert. Schließlich findet man bei Überprüfung des Beweises des zentralen Theorems (1987, S.

[17] Pearce und Rantala haben allerdings vergessen, Mayrs Kompaktheitsbedingung für Zeitintervalle mit anzuschreiben.

[18] Die Autoren behaupten (1987, S. 375f), daß (KGP) nur die ersten beiden von Keplers Gesetzen repräsentiert. Das ist nicht richtig. Keplers drittes Gesetz ist wesentlich für die Ableitung eines konstanten k (statt planetenabhängiger k_i) in (KGP), und es wird auch von (KGP) impliziert. Siehe z.B. Toeplitz (1949, S. 148, 152–154).

379–381), daß die Autoren sogar über eine *exakte* Reduktion ϱ zu verfügen scheinen, und zwar in dem Sinn, daß das (eindeutige) ϱ-Korrelat jedes **NTG**-Modells schon ein Modell von **GCPM*** ist — ohne daß irgendwelche Verschmierungen nötig wären.

Leider ist das zu schön, um wahr zu sein. Die Theorie **GCPM*** der Autoren ist so, wie sie dasteht, keine Spezialisierung von **NTG**.[19] Und für Zweikörpersysteme ist die dritte Bestimmung von **GCPM***, daß nämlich die Sonne sich in einer Trägheitsbahn bewege, inkonsistent mit (NTG) — es sei denn, der Planet im Zweikörpersystem hat die Masse null (was zu Recht durch die Definitionen der Autoren ausgeschlossen wird). Es ist eine unmittelbare Folgerung der Impulserhaltung (welche sich ihrerseits aus Newtons drittem Gesetz ergibt), daß in geschlossenen Newtonschen Zweikörpersystemen nicht die Sonne, sondern der Schwerpunkt mit dem Ortsvektor $\mathbf{s}=(M\mathbf{R}+m\mathbf{r})/(M+m)$ eine gleichförmige geradlinige Bewegung ausführt.[20] Daher kann kein Zweikörpersystem ein Modell sowohl von **NTG** als auch von **GCPM*** sein.[21] Wir schließen unsere Übersicht ab mit der Feststellung, daß der Beitrag von Balzer, Moulines und Sneed die Approximationsperspektive auf den Kepler-Newton-Fall nicht weiterbringt.

Nachdem wir gesehen haben, auf welche Weisen die Beziehung zwischen den Theorien von Kepler und Newton von den Approximationstheoretikern behandelt wird, sind wir jetzt in der Lage, diese Sichtweise insgesamt zu beurteilen. Ich möchte drei Argumente zur Erläuterung vorbringen, warum ich die Approximationsperspektive als nicht ganz zufriedenstellend und nicht ganz vollständig empfinde.

Zunächst war es eines der hervorstechendsten Charakteristika von Approximationen, daß man die exakten Formulierungen von **NTG** und **KGP** nicht direkt aufeinander beziehen konnte. Ja, wir konnten nicht einmal durch die Erweiterung von **NTG** um eine einzige geeignete Menge von vernünftigen Zusatzbedingungen eine exakte Relation nachweisen. Jede

[19] Es scheint sich hier um einen Druckfehler zu handeln. Statt „$M_p(\mathbf{GCPM})$" muß wohl „$M(\mathbf{GCPM})$" in der Definition von $M(\mathbf{GCPM}^*)(a)(1)$ bei Balzer, Moulines und Sneed (1987, S. 378) stehen.

[20] vgl. Kittel u.a. (1979, S. 52f, 118f).

[21] Eine technische Randbemerkung: Wie gelingt Balzer, Sneed und Moulines (1987, S. 379–381) der Nachweis, daß $T_1=\mathbf{KGP}$ auf $T_2=\mathbf{GCPM}^*$ reduzierbar ist? Mit den Symbolen von Kapitel 1 ausgedrückt, müssen sie zeigen, daß $\mathbf{F}$ surjektiv ist und $\mathbf{F}[M_2]\subseteq M_1^{\approx}$ gilt, wobei $M_1^{\approx}\supseteq M_1$ eine „Verschmierung" von M_1 ist. Sie erreichen dies, indem sie eine Reduktionsfunktion $\mathbf{F}$ mit $\mathbf{F}^{-1}[M_1]\cap M_2=\emptyset$, ja sogar mit $M_{p2}^{\circ}\cap M_2=\emptyset$ konstruieren (denn M_{p2}° enthält nur Zweikörpersysteme). So ein $\mathbf{F}$ ist aber offenkundig inadäquat, da es die Ableitbarkeitsbedingung (K2) aus Kapitel 1 leerlaufen läßt.

Menge von Zusatzannahmen für **NTG** war nur für einen gegebenenen Genauigkeitsgrad ε und ein gegebenes Zeitintervall T tauglich, und im allgemeinen müssen wir die Wahl dieser Menge für ein neues $\varepsilon' < \varepsilon$ oder ein neues $T' \supseteq T$ sofort wieder korrigieren. Da aber weder der Genauigkeitsgrad ε noch das Zeitintervall T ein den Theorien **KGP** und **NTG** selbst innewohnendes Charakteristikum ist, können die Zusatzannahmen nicht dahin gehend interpretiert werden, daß sie — von **NTG** aus gesehen — irgendwie „natürliche" Anwendungsbereiche von **KGP** beschreiben.

Zweitens entfernen wir uns mit den Annahmen der Approximationisten offenbar von der Wahrheit. Sie alle verlangen, daß die Massen der Planeten immer kleiner werden (je nachdem, wie groß ε und T sind). Aber man muß immer im Gedächtnis behalten, daß **KGP** Gesetze über *unser* Sonnensystem sind, und daß **NTG** unter anderem zu dem Zweck erdacht wurde, *diese* Gesetze über *dieses* Sonnensystem zu erklären, zu reduzieren oder zumindest zu ersetzen. Schon vor diesen beiden Theorien gab es bereits Beobachtungen der Bahnen von Sternen, Planeten, Satelliten usw. Tycho de Brahes exzellente Messungen beispielsweise waren für Kepler von entscheidender Wichtigkeit. Gegeben diese Beobachtungen und **NTG**, ist es jedoch klar, daß die Massen der Planeten nicht beliebig klein sein *können*. Wenn man die Newtonsche Interpretation von Keplers drittem Gesetz (siehe Abschnitt 8.1.3.1) benutzt, ist es leicht, etwa die Masse der Erde über die Mondbahn auszurechnen.[22] Auch wenn solche Kalkulationen nicht ganz genau sind, reichen sie sicherlich hin, um feste untere Schranken für die Massen der Planeten zu setzen. Die Anfangsbedingungen der wirklichen Welt erweisen somit die für den Approximationsprozeß notwendigen Annahmen als kontrafaktisch.

Drittens und am wichtigsten: Es ist zwar wahr, daß eine Art „Korrespondenzprinzip" im Übergang von Kepler zu Newton am Werke war. In gewissem Sinn erklärt **NTG** Keplers Gesetze, diese können irgendwie auf **NTG** reduziert oder in **NTG** eingebettet werden, sie stellen gewissermaßen ein Grenzfall von **NTG** dar. Aber es ist genauso wichtig — und

[22] Diese Behauptung gilt natürlich nur unter Vorbehalt. Erst einmal muß man den Wert der Gravitationskonstante kennen, den man ebenfalls aus den Bewegungen der Himmelskörper, z.B. von Sonne, Erde und Mond, errechnen kann (vgl. Stumpff 1973, §18). Dann ist es eine Konsequenz aus der Newtonschen Korrektur des dritten Keplerschen Gesetzes (siehe Abschnitt 8.1.3.2), daß man tatsächlich nur die Summe der Massen von Erde und Mond direkt aus der Mondbahn erhält; die Erdmasse kann aber bestimmt werden, wenn man zusätzlich die Bahn des Systems Erde-Mond um die Sonne betrachtet. Drittens abstrahieren alle diese Berechnungen natürlich von den „Störungen" durch andere Himmelskörper.

wurde von den Approximationstheoretikern in seiner Bedeutung vielleicht etwas zu wenig hervorgehoben —, daß **NTG KGP** modifiziert, sie korrigiert, ihnen also widerspricht. Um nur ein Beispiel zu nennen: Einer der größten Triumphe der modernen Wissenschaft wäre gar nicht möglich gewesen ohne ebendiese Newtonsche Berichtigung von **KGP**. Im Jahre 1846 sagten Adams und Leverrier unabhängig voneinander die Existenz und die Position des Planeten Neptun aufgrund der Störungen der Uranusbahn voraus. Ich halte es für angezeigt, mehr Gewicht auf die Tatsache zu legen, daß Newtons Gravitationstheorie nicht nur erklärt, warum Keplers Gesetze fast genau richtig, sondern auch, warum sie letzten Endes doch falsch sind.

Dies sind die Gründe für meinen Versuch, eine alternative Analyse des Kepler-Newton-Beispiels zu erbringen, indem ich es als einen Fall von Idealisierung betrachte. Statt verschiedene Zusatzannahmen anzusetzen, die *sowohl* von extern vorgegebenen Werten von ε und T abhängig *als auch* kontrafaktisch sind, werde ich die mit der intertheoretischen Erklärung verbundene Kontrafaktizität hinnehmen und sogar willkommen heißen. Mein Ziel ist es, zu zeigen, daß *eine einzige* idealisierende Annahme genügt, um eine *strikte* (d.h. nicht bloß approximative) Beziehung zwischen **KGP** und **NTG** zu finden.[23] Diese Annahme wird sich jedoch als verschieden von der Annahme herausstellen, welche sozusagen eine Extrapolation der Approximationsperspektive wäre, nämlich der Annahme, daß alle Planeten die Masse null haben. Indem wir die kontrafaktischen Bedingungen, die nötig sind, um **KGP** im Licht von Newton als Idealisierung zu erweisen, den tatsächlichen Gegebenheiten im Sonnensystem gegenüberstellen, werden wir in der Lage sein, Antworten auf die beiden Warum-Fragen nach

[23] Für den speziellen Fall des Kepler-Newton-Beispiels findet sich die Idee, idealisierende Bedingungen zu verwenden, *um sich von Approximationen „zu befreien" oder um sie „loszuwerden"*, auch bei Krajewski (1977, S. 36–38). Er nennt folgende Anfangs- und Randbedingungen für die „Reduktion" des ersten Keplerschen Gesetzes: (1) Entweder besteht „das System" nur aus Sonne und einem Planeten oder die anderen Planeten „haben keine Wirkung"; (2) keine äußeren Kräfte wirken auf das System; (3) die Sonnenmasse ist „unendlich größer" als die Masse des Planeten (dies formalisiert Krajewski durch m/M=0); (4) der Abstand des Planeten von der Sonne überschreitet einen gewissen Minimalwert; (5) die Tangentialkomponente der Anfangsgeschwindigkeit des Planeten überschreitet ein gewisses Minimum und (6) unterschreitet ein gewisses Maximum. Laut Krajewski (S. 38) ist für die „heutige, allgemeinere Form des ersten Keplerschen Gesetzes" — welches die Bewegungen in Zweikörpersystemen auf den gemeinsamen Schwerpunkt bezieht — Bedingung (3) allerdings überflüssig. — Zu den Bedingungen (4)–(6) (die Tangentialkomponente der Anfangsgeschwindigkeit ist übrigens weder für ein Stürzen des Planeten in die Sonne noch für ein Entfliehen desselben von der Sonne allein ausschlaggebend) siehe Fußnote 9, ein Kommentar zu den Bedingungen (1)–(3) ist implizit im nächsten Abschnitt enthalten.

der offensichtlichen Bedeutung und dem schließlichen Scheitern der Keplerschen Gesetze zu geben.

8.1.3 Das Kepler-Newton-Beispiel als ein Fall von Idealisierung

Zunächst wollen wir das durch das Kepler-Newton-Beispiel gestellte Problem noch einmal kurz reformulieren. In irgendeinem — jetzt einmal im Vagen gelassenen — Sinne erklärt **NTG KGP**; andererseits liefert **NTG** auch, wie wir alle wissen, eine Erklärung, warum **KGP** falsch sind. Man könnte dies wieder durch die Wendung ausdrücken, daß der Übergang von Kepler zu Newton ein Musterbeispiel für das Doppelspiel von Kontinuität und Widerspruch in der Wissenschaft ist. Unter Verwendung der Terminologie von Kapitel 7 können wir erwarten, daß **NTG** nicht nur eine konservative oder gute, sondern sogar eine progressive oder überlegene Nachfolgertheorie von **KGP** ist. Wir haben dort einen Erklärungsbegriff bereitgestellt, der den Anspruch stellen kann, die zweigleisige Aufgabe zu lösen, welche uns durch das Kepler-Newton-Beispiel gestellt ist.

Nun sollte es außer Zweifel stehen, daß **NTG** intuitiv eine überlegene Nachfolgertheorie für die Keplerschen Gesetze ist. Und es ist nach der vorangehenden Diskussion genauso offensichtlich, daß **NTG KGP** höchstens als Idealisierung (strikt) erklären kann, während das, was **NTG** realiter (strikt) erklärt, das Scheitern von **KGP** ist.[24] Gehen wir von den Alternativen einer faktischen, potentiellen oder kontrafaktischen Erklärung im Sinne von Definition 7.2.3 aus, dann ist sicher die kontrafaktische oder idealisierende Erklärung die richtige Option. Es ist bequemer, wenn wir uns im folgenden statt auf die Definitionen von Kapitel 7 gleich auf die in den dortigen Theoremen genannten Bedingungen für die Überlegenheit einer Nachfolgertheorie beziehen. Unsere erste Aufgabe ist es, die passende Anwendungsbedingung A für das Kepler-Newton-Beispiel zu finden.[25] Gemäß Theo-

[24] Vgl. Lakatos (1978/82b, S. 95), der den Weg von Kepler zu Newton als eine „Analyse und Synthese, die nicht erklärt, was sie erklären wollte", bezeichnet. Aus dem Kontext gerissen, scheint auch Hegel so etwas zu meinen, wenn er sagt: „Newton, statt die Gesetze Keplers zu beweisen, hat also vielmehr das Gegenteil getan" (Hegel 1970, S. 99). Sieht man aber die Begründung Hegels (angebliche Fehlerhaftigkeit von Newtons Infinitesimalrechnung) und den ganzen §270 der *Enzyklopädie der philosophischen Wissenschaften* an, so wird klar, daß Hegel von der Materie überhaupt nichts verstand (vgl. dazu auch Shea 1981). Feyerabends gelegentliches Lob auf Hegel (z.B. Feyerabend 1981, S. 175) scheint wirklich nur dazu da zu sein, die Verdienste seiner Opponenten, vornehmlich die von Popper (1957), zu schmälern.

[25] Das Finden von passenden Anwendungsbedingungen oder „auxiliary statements" ist

rem 7.4.3 müssen wir dann nachweisen, daß A und **KGP** bezüglich **NTG** theoretisch völlig äquivalent sind. Die „realen" Punkte (d.h. die Punkte betreffs ¬A und ¬**KGP**) sind trivial: Schon Newton war zu dem Ergebnis gekommen, daß die Anwendungsbedingungen für **KGP** vom Sonnensystem nicht erfüllt werden und daß **KGP** (deshalb!) falsch sein müssen.[26] Somit gilt wegen der Identitätsbedingung (T*I) (vgl. Kapitel 4) $\mathbf{NTG}^*_{\neg A} = \mathbf{NTG} = \mathbf{NTG}^*_{\neg KGP}$ (man beachte, daß ich **NTG** hier als Theorie über das Sonnensystem inklusive Randbedingungen auffasse). Die „idealisierenden" Punkte (betreffs A und **KGP**) sind dagegen interessant und werden denn auch wirklich sowohl in Newtons *Principia* als auch in modernen Lehrbüchern abgehandelt. Unsere zweite Aufgabe wird es also sein, zu zeigen, daß $\mathbf{NTG}^*_A = \mathbf{NTG}^*_{KGP}$ oder — was im Gärdenforsschen Revisionsmodell wegen (T*I) auf dasselbe hinausläuft — daß **KGP** ein Teil von $\mathbf{NTG}^*_A$ und daß A ein Teil von $\mathbf{NTG}^*_{KGP}$ ist.

Der in Kapitel 7 entwickelte Ansatz, angewandt auf das Kepler-Newton-Beispiel, unterscheidet sich vom Vorgehen der Approximationstheoretiker in einigen prinzipiellen Hinsichten. Wenn man von vornherein mit Scheibes Formulierung (KGP) der Keplerschen Gesetze beginnt, scheint mir dies die Gefahr zu bergen, daß einem die lehrreichsten Aspekte der Relation zwischen **KGP** und **NTG** entgehen. Es erfordert nicht allzuviel guten Willen, der Behauptung glauben zu schenken, daß die Lösungen von (KGP)(a)-(c) zumindest eine Zeitlang den Lösungen von (NTG) ziemlich nahe kommen, wenn nur m_1 sehr nahe an k und die m_i's für i=2,...,n sehr nahe an null sind. Jedoch braucht es ein beträchtliches Stück mathematischer Arbeit und bedeutet es eine große kognitive Leistung, zu erkennen, daß (KGP)(a)-(c) wirklich äquivalent zur ursprünglichen Formulierung der Keplerschen Gesetze sind. Auch den interessantesten Teil *der Erklärung* von **KGP** (und ihres Scheiterns) durch **NTG** kann man in diesem Schritt vermuten. Das deduktiv-nomologische Erklärungsschema hat zwar vielerlei harsche

die typische Aufgabe in Putnams (1974, S. 231) „Schema II" für wissenschaftliche Probleme. Im Unterschied zu Putnams Schema ist im vorliegenden Fall das Explanandum, weil erstens gesetzesartig und zweitens falsch, allerdings kein „fact".

[26] Man kann natürlich auch umgekehrt von der Ungültigkeit von **KGP** auf das Nichterfülltsein ihrer Anwendungbedingungen schließen. Dies tut etwa Scriven (1963, S. 122f): „I have to fake the premises [about the gravitational laws and the laws of motion etc.] up in order to get Kepler's laws out of them owing to the inconvenient fact that the planets do not obey Kepler's laws by a long shot." Welche Prämissen muß Scriven „fälschen"? Wir werden sehen, daß es nicht Prämissen „über" die Newtonschen Gesetze sind, sondern nur Prämissen über Randbedingungen. Der dem Zitat folgende Satz ist auch noch interessant: „Still, it might be said to be an explanation of Kepler's laws (insofar as they are true)."

Kritik hinnehmen müssen, aber es wäre sicherlich übertrieben zu sagen, in reiner Deduktion (auf der Grundlage mathematischer Theorien) *könne* keine Erklärungsleistung stecken. Wir wollen deshalb unseren Ausgang von den Keplerschen Gesetzen in ihrer üblichen Form nehmen:

(KGP1) Die Planeten bewegen sich in elliptischen Bahnen, in deren einem Brennpunkt die Sonne steht.

(KGP2) Der Radiusvektor von der Sonne zu einem Planeten überstreicht in gleichen Zeiträumen gleiche Flächen.

(KGP3) Das Verhältnis zwischen dem Quadrat der Umlaufzeit und der dritten Potenz der größeren Halbachse der Ellipse (d.h. der dritten Potenz „der mittleren Entfernung" von der Sonne[27]) ist dasselbe für alle Planeten.

Ab sofort soll **KGP** für die Konjunktion von (KGP1), (KGP2) und (KGP3) stehen.[28] Dies heißt jedoch nicht, daß wir (KGP1)–(KGP3) zu einem einzigen Block zusammenschweißen, in welchem die ursprünglichen Formen der Gesetze nicht mehr identifizierbar sind (wie das in (KGP) der Fall ist). Denn wir werden sehen, daß (KGP3) in gewissem Sinne „mehr idealisiert" ist als (KGP1) und (KGP2).[29]

Die Approximationsperspektive legt eine simple Strategie für die Analyse des Kepler-Newton-Beispiels als Idealisierung nahe. Approximative Gültigkeit von **KGP** kann erreicht werden, indem man die Massen der Planeten „gegen null gehen läßt". Dies heißt aber, sagten wir, vom Pfad der Wahrheit abweichen. Also scheint es nicht sehr viel schlimmer zu sein, wenn man einfach die Planetenmassen *gleich* null „setzt". Man könnte

[27] Es hat sich zwar eingebürgert, ist aber gleichwohl etwas künstlich, „die" mittlere Entfernung eines Planeten von der Sonne mit der größeren Halbachse seiner elliptischen Bahn zu identifizieren. Für eine differenziertere Behandlung dieses Punktes siehe Stumpff (1973, §25).

[28] **KGP** ist von nun an also keine Theorie mit empirischem Wissen (Randbedingungen etc.) mehr, so wie ich es von **NTG** vorausgesetzt habe. Dies ist lediglich eine vereinfachte Auffassung von **KGP**, wenn das empirische Wissen in „Keplers Theorie über unser Planetensystem" mit dem empirischen Wissen von **NTG** übereinstimmt. Wenn das nicht der Fall ist, gibt es Probleme, und wir müssen die einfachen Schemata von Kapitel 7 modifizieren. Vergleiche Kapitel 1, Fußnote 80.

[29] Felix Mühlhölzer hat mich darauf aufmerksam gemacht, daß man als den entscheidenden Unterschied zwischen Kepler und Newton vielleicht die Tatsache ansehen kann, daß Kepler die Sonne als ruhend betrachtet, während sie für Newton in Bewegung ist. Keplers *Gesetze* machen allerdings keine Aussage über die absolute Bewegung der Sonne. Wir sehen nun übrigens auch, daß (KGP)(b) ein Zusatz zu den eigentlichen **KGP** ist, denen genauer eine Zusammenfassung von (KGP)(a) und (b) zu $\ddot{r}_i - \ddot{r}_1 = -(k/|r_i - r_1|^3)\cdot(r_i - r_1)$, $i=2,\ldots,n$, entspräche.

versucht sein zu argumentieren, daß die idealisierende Bedingung

A_0 $\quad m_i = 0$ für i=2,...,n

nur eine Extrapolation und konsequente Anerkennung der Kontrafaktizität ist, die ohnehin im Approximationsprozeß mit inbegriffen ist. Dieser Schachzug hätte den Vorteil, die verschiedenen, von ε abhängigen Annahmen der Approximationsperspektive durch eine einzige idealisierende Annahme zu ersetzen, welche (NTG) strikt auf (KGP)(a)–(c) reduziert (vorausgesetzt, daß m_1 identisch mit k gewählt wird). Aber das ist natürlich illegitim! Es macht ganz einfach keinen Sinn, im Kontext der zentralen Axiome Newtons masselose Planeten oder irgendwelche anderen masselosen Teilchen anzunehmen. Die Gravitationstheorie kann die Bahn masseloser Teilchen durchaus nicht erklären, da nach dem charakteristischen Gesetz von **NTG** solche Teilchen überhaupt nie angezogen werden. Noch schlimmer ist, daß man das von **NTG** präsupponierte zweite Newtonsche Gesetz gar nicht so revidieren kann, daß es auf masselose „Körper" anwendbar wäre: Es ist ausgeschlossen, daß irgendeine Kraft welcher Art auch immer auf solche „Körper" wirkt (jedenfalls dann, wenn wir unendliche Beschleunigungen ausschließen). Es wird demnach augenscheinlich, daß die idealisierende Bedingung A_0 mit den fundamentalsten Prinzipien von **NTG** nicht in Einklang gebracht werden kann.[30] Mit den in Abschnitt 7.5 eingeführten Worten ausgedrückt, steht A_0 *in unauflösbarem Konflikt mit* **NTG**: **NTG** muß jede Antwort schuldig bleiben auf die Frage, was denn los wäre, wenn die Planeten die Masse null hätten. Also gibt es überhaupt keine minimale Revision von **NTG** zum Zwecke einer Assimilation von A_0, also ist A_0 auch kein ernsthafter Kandidat für die Anwendungsbedingung, die **NTG** als überlegene Nachfolgertheorie von **KGP** ausweist.

Wir müssen uns bei anderen Quellen als bei den Approximationisten nach neuen Strategien umsehen. Das Material, von dem wir die richtigen Hinweise bekommen, ist nicht schwer beizubringen. Man findet es in Newtons *Principia* (Newton 1963) und in modernen elementaren Lehrbüchern wie dem *Berkeley Physik Kurs* (Kittel u.a. 1979). Wir werden zwei „Grade der Idealisierung" unterscheiden. Die erste, „hochgradigere" Idealisierung bezieht sich auf die Keplerschen Gesetze in ihrer ureigensten Form, die zweite, „realistischere" Idealisierung auf (KGP1), (KGP2) und eine kor-

[30] Dies wurde schon von Popper (1957) betont und ohne Zweifel von den Approximationstheoretikern klar erkannt. Vgl. aber Stegmüller (1986, S. 309): „Die Inkommensurabilität besteht hier darin, daß in der Kepler-Theorie die Planeten Nullmasse haben, während in der Theorie Newtons nur positive Werte der Massenfunktion zugelassen sind."

rigierte Version des dritten Keplerschen Gesetzes. In beiden Fällen werden wir die beiden Richtungen diskutieren, in welche die Idealisierungen „durchgehen" müssen: von Keplers Gesetzen zu ihren Anwendungsbedingungen in **NTG** und umgekehrt von diesen Anwendungsbedingungen zu Keplers Gesetzen. Theorem 7.4.2 schreibt vor, daß beide Wege gangbar sein müssen, damit man **NTG** gemäß den Definitionen von Kapitel 7 als überlegene Nachfolgertheorie von **KGP** bezeichnen kann.

8.1.3.1. Erste Idealisierung: Einkörpersysteme. Der erste Versuch, **KGP** vom Standpunkt von **NTG** aus vermittels einer Idealisierung zu rechtfertigen, behält **KGP** ohne jede Abstriche bei. In der Tat sind die Keplerschen Gesetze sogar der Ausgangspunkt dieser Idealisierung.

Von Keplers Gesetzen zu ihrer Newtonschen Anwendungsbedingung: Einkörpersysteme. Keplers Gesetze, interpretiert als Gesetze über die Bewegung von „Körpern", spielen gleich in den ersten Abschnitten von Newtons *Principia* eine zentrale Rolle. Grob gesagt, beweisen die Propositionen I-III in Buch 1, daß die Planetenbewegung in einer Ebene gemäß dem Flächengesetz (KGP2) (mit dem Sonnenzentrum als Referenzpunkt) äquivalent ist mit der Existenz einer Zentripetalkraft (mit der Sonne als unbeschleunigtem Referenzpunkt). Auf dieser Grundlage zeigt Proposition XI, daß aus dem Ellipsengesetz (KGP1) folgt, daß die Zentralkraft dem Quadrat des Abstands von der Sonne indirekt proportional ist, dies aber nur für (die Bahn eines) jeden Planeten einzeln. Korollar VI von Proposition IV, welches (KGP3) im Spezialfall einer gleichförmigen Kreisbewegung behandelt, weist nach, daß in diesem Fall ein universelles $1/r^2$-Kraftgesetz wirksam ist.[31] Eine moderne Darstellung, wie sich aus (KGP1)-(KGP3) eine $1/r^2$-Zentripetalbeschleunigung ergibt, findet man in Born (1949, S. 129–133), wo insbesondere auch gezeigt wird, wie das auf Ellipsen angewandte (KGP3) die Abhängigkeit des $1/r^2$-Gesetzes von speziellen Eigenschaften (bei Born z.B. die Flächenkonstante und das Semilatus Rectum) der elliptischen Bahn eines Planeten beseitigt.

[31] Ich kann in Newtons *Principia* keine Ableitung des universellen $1/r^2$-Kraftgesetzes aus dem *auf Ellipsen angewandten* dritten Keplerschen Gesetz finden. (Die umgekehrte Implikation wird in Proposition XV formuliert.) Dies hängt möglicherweise mit Newtons Meinung zusammen, daß das $1/r^2$-Kraftgesetz besser über das Ruhen der Apsiden der Planetenbahnen gezeigt wird. Vgl. seinen Kommentar zum Beweis von Proposition II in Buch 3 der *Principia*.
NB: Ich halte mich an die international gebräuchliche Numerierung der Newtonschen Ergebnisse in der amerikanischen Ausgabe (Newton 1934) und spreche demgemäß von „Propositionen" und „Korollaren" statt von „Lehrsätzen" und „Zusätzen" (wie es in Newton 1963 heißt).

Aufgrund dieses Ergebnisses und des zentralen zweiten Newtonschen Gesetzes wird klar, daß die Anwendungsbedingung für **KGP** in **NTG** die Planeten als eine spezielle Art von *Einkörpersystem* behandelt. Dies ist die erste Idealisierung:

A_1 Die Planeten sind je einzelne Körper, welche sich unter dem Einfluß einer zentripetalen $1/r^2$-Gravitationskraft bzgl. eines nichtbeschleunigten Zentrums bewegen.

Man beachte, daß dem jeweils betrachteten Planeten keinerlei Beschränkung seiner Masse auferlegt wird — außer daß sie nicht gleich null sein darf. Die Idealisierung in dieser Anwendungsbedingung ist dennoch sehr einschneidend. Sie vereinzelt die Planeten künstlich und muß dabei nicht nur die Existenz der anderen Planeten, sondern auch die Existenz der Sonne verleugnen, denn interplanetarische Wechselwirkungen würden die Zentripetalkraft stören, und die Sonne *als massiver Körper* würde von dem je betrachteten Planeten angezogen und deshalb beschleunigt werden. Wir dürfen also A_1 nicht durch die Hinzufügung verstärken, daß sich die Sonne tatsächlich im Zentrum der Kraft befindet (was eine weitere Konsequenz aus (KGP1)–(KGP3) wäre). Außerdem wirft A_1 die dringende Frage auf, woher die $1/r^2$-Gravitationskraft denn überhaupt kommt. Ich werde auf die gravierendsten dieser Schwierigkeiten in Abschnitt 8.1.3.2 zurückkommen.

Wir haben gesehen, daß es natürlich nicht — wie manchmal behauptet — die Newtonsche Theorie ist, die aus **KGP** hergeleitet werden kann. Aber innerhalb von **NTG**, wo der Begriff der Kraft verfügbar ist, implizieren **KGP** A_1 in dem Sinne, daß die Planeten, *wenn* sie sich wirklich nach Keplers Rezept bewegen *würden*, Objekte in Einkörpersystem *wären*, wie sie in A_1 beschrieben sind. Wir haben also gerade aufgedeckt, wie **NTG**, um Poppers (1957, S. 29) Worte zu gebrauchen, „abgeändert werden müßte[n] — welche falschen Prämissen angenommen oder welche Bedingungen festgesetzt werden müßten", um **KGP** wahr zu machen.

Von Einkörpersystemen zu Keplers Gesetzen. Gehen wir nun davon aus, daß sich die Planeten unter dem ungestörten Einfluß einer $1/r^2$-Zentripetalkraft bewegen, wie sie in A_1 beschrieben ist. Die auf einen einzelnen Planet wirkende Kraft ist dann gegeben durch

$$\text{(8.1.7)} \quad \mathbf{F} = m\ddot{\mathbf{r}} = -(GMm/|\mathbf{r}|^2)\cdot\hat{\mathbf{r}}.$$

Hier gibt $\mathbf{r}$ den Ort des Planeten bezüglich des (nichtbeschleunigten!) Kraftzentrums an, $\hat{\mathbf{r}} = \mathbf{r}/|\mathbf{r}|$ ist der Einheitsvektor in Richtung von $\mathbf{r}$, und G ist die Gravitationskonstante (die wir nun, um die üblichen Formeln zu bekommen, nicht mehr gleich 1 setzen wollen); m bezeichnet die Plane-

tenmasse, und M repräsentiert die „Sonnenmasse". Aber da (8.1.7) die Einkörpersystem-Bedingung A_1 formalisieren soll, hat man zu gewärtigen, daß M jetzt nicht als wirkliche Masse eines wirklichen Körpers zu verstehen ist, sondern als eine bloße Konstante — nennen wir sie *die Quasimasse* — des Sonnensystems.

Es ist Gegenstand der meisten einführenden Mechaniklehrbücher, daß die Bahnen von Teilchen[32] unter der alleinigen Einwirkung einer solchen $1/r^2$-Zentripetalkraft Kegelschnitte beschreiben. Der von dem Kraftzentrum aus zum Teilchen gezogene Radiusvektor überstreicht, wie man schnell zeigen kann, für Zentralkräfte beliebiger Größe in gleichen Zeitintervallen gleich große Flächen. Schließlich findet man, daß Keplers drittes Gesetz gültig ist, wobei das darin erwähnte Verhältnis $4\pi^2/GM$ ist. Für all dies siehe Kittel u.a. (1979, S. 171–177).

Wenn also Planeten so wären wie in A_1 beschrieben, dann wären sie vollkommen Keplersche Körper. Wir haben in 3.1.1 nachgewiesen, daß A_1 Element von $\mathbf{NTG}^*_{\mathbf{KGP}}$ ist, und eben gezeigt, daß umgekehrt auch **KGP** Element von $\mathbf{NTG}^*_{A1}$ ist. Damit ist nach (T*I) und Theorem 7.4.3 **NTG** wirklich eine überlegene Nachfolgertheorie für **KGP**, die nicht nur **KGP** kontrafaktisch, sondern auch das Scheitern von **KGP** real erklärt: Weil Planeten keine einsamen Teilchen in einem $1/r^2$-Zentripetalfeld um ein nichtbeschleunigtes Zentrum sind, deshalb sind die Keplerschen Gesetze nicht wahr.

8.1.3.2. Zweite Idealisierung: Zweikörpersysteme. Wie jede Idealisierung ist A_1 kontrafaktisch. Das hat Newton klar gesehen, der Abschnitt XI von Buch 1 der *Principia* mit den folgenden Sätzen beginnt:

> Bis jetzt habe ich die Bewegung solcher Körper auseinander gesetzt, welche nach einem unbeweglichen Centrum hingezogen werden, ein Fall, der kaum in der Natur existirt. Es pflegen nämlich Anziehungen auf Körper stattzufinden, jedoch sind die Wirkungen der ziehenden und der angezogenen Körper nach dem dritten Gesetze stets wechselseitig und einander gleich, so dass weder der anziehende noch der angezogene Körper ruhen kann, sondern, wenn ihrer zwei sind, beide (nach Gesetze, Zusatz 4.) sich gleichsam durch wechselseitige Anziehung um den gemeinschaftlichen Schwerpunkt drehen.

[32] Newton nahm große Mühen auf sich, um zu zeigen, daß es keinen Unterschied macht (und also eine relativ harmlose Idealisierung ist), wenn man Planeten als ausdehnungslose Partikeln anstatt als massive Kugeln (was auch wieder eine Idealisierung ist) betrachtet. vgl. Newton (1963, Abschnitt XII von Buch 1) und Kittel u.a. (1979, S. 165–169).

(Newton 1963, S. 166f) Newton deutet hier an, daß möglicherweise das ganze Unterfangen der ersten zehn Abschnitte der *Principia* als im Widerstreit mit dem dritten seiner fundamentalen Axiome, dem Actio-Reactio-Prinzip, stehend angesehen werden muß. In der Tat war der Preis, den wir für die Rettung der Keplerschen Gesetze im vorigen Abschnitt zahlen mußten, sehr hoch. Die Bedingung A_1 bringt uns in große Verlegenheit: Wie sollen wir die Existenz nicht nur der anderen Planeten, sondern auch der Sonne leugnen? Woher soll das Zentralkraftfeld dann kommen? Wie können wir die Quasimasse M in der Newtonschen Formel (8.1.7) für die Gravitationskraft interpretieren? Ich kenne keine vernünftigen Antworten auf diese Fragen. Vor allem aber ist es das nicht loszuwerdende Gefühl, daß A_1 mit Newtons drittem Bewegungsgesetz unvereinbar ist,[33] das uns zu dem Schluß führt, daß A_1 in fundamentalem Konflikt mit **NTG** steht. Zwar ist dieser Konflikt mit dem „Geiste" der Gravitationstheorie offensichtlich ein auflösbarer (vgl. Abschnitt 7.5), denn es bereitet ja keinerlei Schwierigkeit zu sagen, was der Fall wäre, wenn A_1 erfüllt wäre: Keplers Gesetze wären wahr. Dennoch ist es sehr fraglich, ob es wirklich „minimale" Revisionen von **NTG** geben kann, die A_1 mit einschließen, ob also die Keplerschen Gesetze (KGP1)–(KGP3) in **NTG** auch nur per Idealisierung erklärt werden können.

Bedeutet dies nun ein Scheitern der Idealisierungsperspektive? Keineswegs. Ich meine, die obige Analyse zeigt vielmehr ziemlich deutlich, daß Keplers ursprünglicher Block von Gesetzen im Lichte von **NTG** nicht mehr haltbar ist. Denn trotz der approximativen Gültigkeit von **KGP** und trotz ihrer approximativen Erklärung durch **NTG** liefern unsere Überlegungen ein klares Argument dafür, daß **KGP** einer Modifikation bedürfen. Dies ist überraschend, besonders weil wir ja schon erlauben, eine Revision von **NTG** durch idealisierende, kontrafaktische Annahmen durchzuführen. Sehen wir uns deshalb das Argument noch einmal genauer an.

Intuitiv ist Newtons Theorie zweifellos eine überlegene Nachfolgertheorie für Keplers Gesetze. Also sollten wir dies auch auf einer mehr formalen Ebene widerspiegeln können. In Abschnitt 8.1.3.1 versuchten wir nachzuweisen, daß **NTG** eine überlegene Nachfolgertheorie für **KGP** im Sinne von Kapitel 7 ist. Wir fanden heraus, daß A_1 die für das in Theorem 7.4.2 (oder 7.4.3) genannte Kriterium geeignete Anwendungsbedingung ist, d.h. daß A_1 die richtige Anwendungsbedingung von **KGP** in der Newtonschen

[33] Meines Erachtens ist es aber auch nicht mehr als ein bestimmtes Gefühl, denn das dritte Newtonsche Gesetz ist letztendlich doch zu vage, um einen wirklich schlüssigen Nachweis zu gestatten, daß es mit A_1 inkonsistent ist.

Rahmentheorie wäre. Aber A_1 hat — sehr vorsichtig formuliert — einen merklichen Anflug von Unvereinbarkeit mit dem dritten Gesetz Newtons, und wir hielten es für verboten, Annahmen zu **NTG** „hinzuzufügen", die im — wenn auch auflösbaren — Konflikt mit ihren grundlegendsten Axiomen stehen. Somit haben wir allem Anschein nach (KGP1)-(KGP3) als idealisierte Vorgängertheorie von **NTG** oder, anders ausgedrückt, **NTG** als überlegene Nachfolgertheorie von (KGP1)-(KGP3) disqualifiziert. Es können also nicht **KGP** in ihrer ureigenen Form sein, die sich dem Schema der doppelten Erklärung (s. Kapitel 7) durch **NTG** fügen. Glücklicherweise wird sich eine kleine Modifikation von **KGP** als tauglich erweisen. Und diese Modifikation erschließt sich uns auf natürliche Weise, wenn wir Newtons eigenen Gedankengängen folgen. Betrachten wir nun Zweikörpersysteme.

Von Zweikörpersystemen zur Modifikation von Keplers Gesetzen. Die zweite, alternative Idealisierung beginnt damit, daß in der Anwendungsbedingung für **KGP** in **NTG** die Planeten als Teile von *Zweikörpersystemen* behandelt werden:

A_2 Die Planeten sind je einzelne Körper, die sich um die Sonne drehen.

Seien m und M die Newtonschen Massen des Planeten bzw. der Sonne in einem Zweikörpersystem. Es ist wohlbekannt, daß man das Problem der Zweikörperbewegung mathematisch auf die folgende Gleichung und damit auf ein virtuelles Einkörpersystem reduzieren kann (siehe Kittel u.a. 1979, S. 177–179):

(8.1.8) $\mu\ddot{\mathbf{r}} = -(GMm/|\mathbf{r}|^2)\cdot\hat{\mathbf{r}}$.

Hier bezeichnet $\mathbf{r}$ den Ort des Planeten bezüglich der Sonne als Koordinatenursprung, und $\mu = Mm/(M+m)$ wird in der Himmelsmechanik die *reduzierte Masse* genannt. Seltsamerweise bleiben die einschlägigen Darstellungen bei dieser Gleichung stehen, obwohl sie eigentlich noch nicht die Form der Gleichung (8.1.7) eines Einkörpersystems hat, denn auf der linken und rechten Seite sind verschieden „Massen" im Spiel. Durch Umformen mittels der Beziehung $Mm = (M+m)\mu$ kann aber ganz leicht dafür gesorgt werden, daß (8.1.8) zu einer Einkörpergleichung wird, die völlig strukturgleich mit (8.1.7) ist:

(8.1.9) $\mu\ddot{\mathbf{r}} = -(G[M+m]\mu/|\mathbf{r}|^2)\cdot\hat{\mathbf{r}}$.

Erst jetzt kann die Bewegung eines Planeten behandelt werden, *als ob* er sich in einem $1/r^2$-Zentripetalkraftfeld mit der Quasimasse M+m bewegen würde, dessen Mittelpunkt sich im wirklichen Zentrum der Sonne befin-

det. Man darf indes aus der Bewegung des Planeten keinen Schluß auf seine „Als-ob-Masse" ziehen, obgleich diese scheinbar gleich der reduzierten Masse μ sein muß. Da sich nämlich in Einkörpergleichungen die Planetenmasse — ebenso wie μ in (8.1.9) — stets herauskürzt, spielt sie kinematisch keine Rolle. Man halte sich vor Augen, daß wir es trotzdem mit einer echten als-ob-Konstruktion zu tun haben, weil die Sonne, wie wir bereits in Abschnitt 8.1.2 bemerkten, in Zweikörpersystemen kraft **NTG** *nicht* der Ursprung eines Inertialsystems sein kann (sondern sich um den gemeinsamen Schwerpunkt dreht) und weil ihre Masse nach Voraussetzung *nicht* M+m (sondern M) ist.[34] Es ist keine echte Kraft, was durch die linke Seite von (8.1.9) definiert wird.

Gleichung (8.1.9) ist nichtsdestoweniger äußerst hilfreich bei der Untersuchung der Gültigkeit von Keplers Gesetzen. Dank der strukturellen Identität mit (8.1.7) und dank der Tatsache, daß $\mathbf{r}$ den Ort relativ zur Sonne angibt, ist es klar, daß (KGP1) und (KGP2) auch jetzt noch *strenggenommen gültig* sind: In Zweikörpersystemen bewegen sich die Planeten wirklich auf Ellipsen um die Sonne, wobei die Sonne in einem Brennpunkt „steht", und die Radiusvektoren überstreichen wirklich gleiche Flächen in gleichen Zeitintervallen. Im Gegensatz zu diesen Gesetzen, in denen die Gestalt der Bahn relevant ist und ihre Größe nicht, ist das sog. harmonische Gesetz (KGP3) nicht mehr ganz genau gültig. Wenn man im Argument von Kittel u.a. (1979, S. 176) M+m an die Stelle von M und μ an die Stelle von m setzt, dann sieht man, daß das in (KGP3) genannte Verhältnis gleich $4\pi^2/G(M+m)$ ist. Wir müssen (KGP3) also abändern zu

(KGP3*) Das Verhältnis zwischen dem Quadrat der Umlaufzeit und der dritten Potenz der großen Halbachse der Ellipse (der dritten Potenz „der mittleren Entfernung" von der Sonne) ist umgekehrt proportional zur Summe der Massen von Sonne und Planet.

[34] Solche als-ob-Redeweisen sind keine Erfindung von (Wissenschafts-) Philosophen, sondern werden gar nicht selten von Physikern selbst verwendet. So schreibt zum Beispiel Raynor L. Duncombe in der *McGraw-Hill Encyclopedia of Science and Technology*, Artikel „orbital motion", folgendes: „... it is convenient to consider only the relative motion of a planet of mass m about the Sun of mass M *as though* the planet had no mass and moved about a center of mass M+m. The orbit so determined is exactly the same shape as the true orbits of planet and Sun about their common center of mass, but it is enlarged in the ratio (M+m)/M." (Hervorhebung von mir) Diese Stelle ist allerdings kein Glanzstück in der ansonsten ausgezeichneten Enzyklopädie, denn erstens folgt aus **NTG** für masselose Planeten gar nichts und zweitens ist die (positive) Masse des Planeten bei gegebenen Anfangsbedingungen und Quasimasse M+m einerlei.

Häufig wird dieses Gesetz sogar als „Keplers drittes Gesetz" präsentiert.[35] Wir wollen von nun an die Konjunktion von (KGP1), (KGP2) und (KGP3*) **KGP*** nennen. Meine These ist nun, daß diese Variation der Keplerschen Gesetze die richtige „Vorgängertheorie" zur überlegenen Theorie Newtons ist. Man beachte auch, daß unter der idealisierenden Annahme A_2 keinerlei Approximationen nötig sind, um **KGP*** zu erhalten. Andererseits zeigt die idealisierende Theorie **KGP*** ganz genau, warum Scheibes Approximationen (welche auch nur für den Zweikörperfall funktionieren, siehe Abschnitt 8.1.2) notwendig waren: weil man (KGP3) korrigieren — oder in der anderen Richtung: weil man (KGP3) retten — muß. Und gerade hier ist ein rechter Platz für Approximationen im Kepler-Newton-Beispiel: (KGP3) ist approximativ gültig, weil die in (KGP3*) genannten Summen für alle Planeten ungefähr denselben Wert haben.

Von den modifizierten Keplerschen Gesetzen zu ihrer Newtonschen Anwendungsbedingung: simulierte Zweikörpersysteme. Es gibt noch einen Punkt, auf den wir eingehen müssen, wenn **NTG** eine überlegene Nachfolgertheorie für **KGP** sein soll. Wir müssen zeigen, daß A_2 ein Element von $\mathbf{NTG}^*_{\mathbf{KGP}^*}$ ist. Nehmen wir also an, unser Sonnensystem gehorchte den modifizierten Keplerschen Gesetzen **KGP***.[36]

Nebenbei sei bemerkt, daß diese Annahme nur vor dem Hintergrund von **NTG** möglich ist, denn (KGP3*) kann nicht einmal formuliert werden ohne den Newtonschen Massenbegriff.[37] Können wir von **KGP*** zu ihrer Newtonschen Anwendungsbedingung kommen, und wenn ja, wie? Bei der Suche nach einer Antwort auf diese Frage kann man den größten Teil des Bornschen Wegs von **KGP** zu A_1 übernehmen. Nur das letzte Teilstück (Born 1949, S. 132) seines Beweises wird von der Ersetzung von (KGP3) durch (KGP3*) berührt. Und sogar hier kann man leicht Abhilfe schaffen, wenn man M+m anstelle von M verwendet. Das Ergebnis lautet, daß **KGP*** die Gleichung

[35] In der ersten Auflage des *Berkeley Physics Course* ist es (KGP3*), welches aus den Prämissen des Zweikörpermodells bewiesen wird, und die Abweichung von **KGP** wird nicht einmal kommentiert. Erst in der zweiten Auflage, die der deutschen Ausgabe zugrunde liegt, werden die Keplerschen Gesetze im Kontext von Einkörpersystemen diskutiert (Kittel u.a. 1979, S. 176).

[36] Newton wußte, daß diese Annahme kontrafaktisch ist. Man sehe zum Beispiel seine Bemerkungen zu Proposition XIII von Buch 3 der *Principia*.

[37] In Kapitel 7 bin ich davon ausgegangen, daß, intuitiv gesprochen, Vorgängertheorien immer „theoretisch" und Anwendungsbedingungen immer „nichttheoretisch" (bezüglich der Nachfolgertheorie) sind. Es ist interessant, daß das Paar $\langle A_2, \mathbf{KGP}^* \rangle$ diese Vermutung, die Paare $\langle A_0, \mathbf{KGP} \rangle$ und $\langle A_1, \mathbf{KGP} \rangle$ aber gerade ihre Umkehrung zu stützen scheinen.

(8.1.10) $\ddot{\mathbf{r}} = -(G[M+m]/|\mathbf{r}|^2)\cdot\hat{\mathbf{r}}$.

impliziert, was gerade (8.1.9) ist, nachdem man dort μ auf beiden Seiten gekürzt hat.

Was zeigt dieses Ergebnis? Es zeigt zunächst einmal, daß sich gemäß **KGP*** jeder Planet so bewegt, *als ob* er in einem $1/r^2$-Zentralfeld kreiste, dessen Zentrum mit dem Sonnenmittelpunkt zusammenfällt und das durch die Quasimasse M+m charakterisiert wird. Aber dies darf man natürlich nicht wörtlich verstehen, da die Sonne ja wirklich im Zentrum *ist*, aber ihre Masse eindeutig und gleich M ist. Also birgt die „wörtliche" Interpretation von (8.1.10) durch körperlose Zentralkräfte eine noch größere Gefahr des (auflösbaren) Konflikts mit Newtons fundamentalen Axiomen als die Konstruktion aus Abschnitt 8.1.3.1.

Wie wir im letzten Unterabschnitt gesehen haben, genügen Planeten in allen Zweikörpersystemen, in denen sich die Massen von Sonne und Planet zu M+m aufaddieren, der Gleichung (8.1.10). Aber auch wenn wir die je relevanten Massen als gegeben annehmen, dürfen wir nicht kontrafaktisch von (8.1.10) auf A_2 schließen, da es, soweit ich sehe, keinen Weg gibt, aus einer bloßen Differentialgleichung zwingend die Anzahl der beteiligten Körper abzuleiten. Anscheinend müssen wir also A_2 dadurch abschwächen, daß wir kinematisch äquivalente Systeme zulassen:[38]

A_2' Die Planeten bewegen sich wie je einzelne Körper, die sich um die Sonne drehen.

Zugegeben, dies ist kein angenehmes Ergebnis. Es ist unschön, daß man noch eine wie-Phrase zu den bereits idealisierenden, d.h. kontrafaktischen Anwendungsbedingungen hinzufügen muß. Solange wir jedoch nicht in der Lage sind, zu beweisen, daß sich (8.1.10) nur auf Zweikörpersysteme beziehen kann, sehe ich keine Möglichkeit, um dieses Manöver herumzukommen. Immerhin ist es nicht direkt schädlich, denn die (kontrafaktische) Ableitung von **KGP*** innerhalb von **NTG** wird, wie man sich leicht klar macht, nicht beeinträchtigt, wenn A_2' den Platz von A_2 einnimmt. Und es ist ebenso

[38] „Kinematisch äquivalent" soll heißen, daß bei geeigneter Wahl der Koordinatensysteme identische raumzeitliche Zustandsbeschreibungen einander entsprechender Teilchen vorliegen, wohingegen irgendwelche Konstanten, Massen und Kräfte in den Systemen verschieden sein dürfen. Zum Beispiel ist — wie gesagt — jedes Zweikörpersystem mit den Massen M′ und m′ kinematisch äquivalent zu dem Zweikörpersystem mit den Massen M und m, sofern nur M′+m′=M+m (wie man aus (8.1.9) ersieht). — Ein ähnlicher Einwand wäre wohl auch gegen die „Herleitung" von Einkörpersystemen aus **KGP** vorzubringen gewesen. Weil mir der Konflikt mit Newtons drittem Gesetz wichtiger vorkommt, habe ich in Abschnitt 8.1.3.1 jedoch auf die entsprechende Überlegung verzichtet.

einsichtig, daß die rein kinematische Annahme A_2' keinen Schaden an den fundamentalen Axiomen von Newtons Gravitationstheorie verursacht.

Somit haben wir eine Bedingung angegeben, welche die theoretische Signifikanz der Keplerschen Gesetze aus dem Blickwinkel von **NTG** zeigt. In anderen, präziseren Worten gesagt, zeigt diese Bedingung, daß **NTG** eine überlegene Nachfolgertheorie für die verbesserte Version **KGP*** ist in dem Sinn, daß die minimale (kontrafaktische!) Revision von **NTG**, die nötig ist, um A_2' zu akzeptieren, **KGP*** mit einschließt und umgekehrt die minimale (kontrafaktische!) Revision von **NTG**, die nötig ist, um **KGP*** zu akzeptieren, A_2' mit einschließt. Nach der in Kapitel 6 durchgeführten Analyse kann man äquivalent behaupten, daß ein Vertreter von **NTG** folgender doppelter Erklärung zustimmen wird: Wenn Planeten wie einzelne Körper um die Sonne rotierten, dann wäre **KGP*** korrekt; aber weil sie es nicht tun, deshalb scheitert **KGP*** in Wirklichkeit. Damit wird **KGP*** in **NTG** durch A_2' als Idealisierung erklärt, gleichzeitig erklärt A_2' in **NTG** faktisch das Scheitern von **KGP***.

8.1.4 Idealisierung und Approximation im Kepler-Newton-Fall

Dieses Teilkapitel hat, glaube ich, gezeigt, daß Idealisierungen von Approximationen verschieden sind, ein komplementäres Bild liefern und sich anscheinend besser für die Analyse des Kepler-Newton-Beispiels eignen. Sie sind verschieden, weil riesige Planeten **KGP** als Approximationen, aber nicht als Idealisierungen unbrauchbar werden ließen, während in einer Cartesianischen Physik **KGP** als Idealisierungen, aber nicht als Approximationen ihren Wert verlören (vgl. Baigrie 1987).[39] Sie sind komplementär

[39] Ein wichtiges Beispiel, welches als Approximation, aber wohl nicht als Idealisierung zu analysieren ist, ist das *Galilei-Newton-Beispiel*. Welche kontrafaktische Bedingung könnte Galileis Fallgesetz vom Standpunkt von **NTG** aus als Idealisierung ausweisen? Als plausibelster Kandidat scheint die Annahme „Erdradius + Abstand des Körpers vom Erdboden = konstant" in Frage zu kommen. Dies impliziert aber, daß der „fallende Körper" gar nicht fällt, d.h. die Annahme steht in (unauflösbarem?) Konflikt mit dem Galileischen *Fall*gesetz. Auch ein Konflikt von A *mit* T_1 sollte das Zustandekommen einer idealisierenden Erklärung ausschließen, denn es kann dann eigentlich nicht wirklich T_1 sein, was in $T_2{}^*{}_A$ enthalten ist. Abgesehen davon, daß Galileis Gesetz für Newton deduktiv nicht viel „hergibt", kann man also auch über die Idealisierungsanalyse plausibel machen, daß Keplers Gesetze für Newton viel wertvoller waren als das (approximativ gesehen ungefähr gleich gute) Galileische Fallgesetz. Die *Principia* gehen nicht von ungefähr mit **KGP** los — Newtons Geschichte vom fallenden Apfel als Inspiration für die Gravitationstheorie hingegen ist nichts als ein hübsches Märchen.

in dem Sinne, daß die Approximationsperspektive quantitative Relationen zwischen Modellen gewisser kinematischer Differentialgleichungen wie (NTG) und (KGP) betrachtet, wohingegen die Idealisierungsanalyse qualitativ ist und ihr Augenmerk darauf richtet, wie man solche Gleichungen aus Keplers Gesetzen erhält und wie man sie in **NTG** interpretieren kann. Schließlich möchte ich noch einmal zusammenfassen, warum mir die Idealisierungsperspektive für das vorliegende Beispiel lehrreicher als die Approximationsperspektive erscheint.

Worin besteht der Wert von Keplers Gesetzen? Es kann nicht der Punkt sein, daß sie *Approximationen an die Wahrheit* sind. Denn sie waren in der Mitte des siebzehnten Jahrhunderts weder empirisch zweifelsfrei abgesichert noch theoretisch konkurrenzlos. Cassinis Ovale waren ernsthafte Rivalen für die Keplerschen Ellipsen. Wenn man aber eine Ellipsenform annimmt, dann bekommt man aus der „einfachen elliptischen Hypothese" (nach welcher die Planeten sich mit konstanter Winkelgeschwindigkeit bezüglich des leeren Brennpunkts bewegen) praktisch genauso gute Resultate wie aus Keplers Flächengesetz. Schließlich bat Newton 1684 Flamsteed dringend darum, sich der Abweichung der Saturnbahn von den Werten, die sich nach Keplers drittem Gesetz ergeben müßte, anzunehmen. Wissenschaftshistoriker gestehen meistens ohne weiteres zu, daß **KGP** als „empirische Prämissen" für Newton bedeutungslos waren (vgl. Wilson 1969/70 und Baigrie 1987).

Die Pointe von **KGP** wird besser getroffen, wenn man die Tatsache betont, daß sie *Approximationen an* **NTG** sind. Aber wie ich am Ende von Abschnitt 8.1.2 erläutert habe, erfordert der Approximationsprozeß Zusatzannahmen, die *sowohl* von Genauigkeitsgraden ε (und, wenn n>2, von Zeitintervallen T) abhängen *als auch* ein Abweichen von der Wahrheit mit sich bringen, und er läßt die Frage offen, *warum* sich die Keplerschen Gesetze letztendlich doch als inadäquat erweisen.

Idealisierungen im hier verwendeten Sinne scheinen bessere Dienste zu leisten. Man kann keine Approximationen ohne kontrafaktische Annahmen bekommen, aber man kann umgekehrt kontrafaktische Bedingungen formulieren, die uns der Notwendigkeit von Approximationen entheben. In Abschnitt 8.1.3 habe ich idealisierende oder kontrafaktische Annahmen ernst genommen. Wir haben deutliche Hinweise gefunden, daß die ursprünglichen Keplerschen Gesetze vom Standpunkt von **NTG** aus nicht einmal als Idealisierung gerechtfertigt werden können, denn die Existenz von Einkörpersystemen ist in der letzteren Theorie schwerlich vorstellbar. Aus diesem Grunde mußten wir **KGP** etwas modifizieren und landeten so

bei ihrer „theoretischen Konkretisierung“[40] **KGP***. Wir haben gezeigt, daß **KGP*** *völlig korrekt* wären, *wenn* unsere Planeten zusammen mit der Sonne Zweikörpersystemen bilden *würden*, oder genauer: wenn sie sich *wie Planeten in Zweikörpersystemen bewegten*. Die Idealisierungsanalyse stellte heraus, warum **KGP** sogar als Idealisierung scheitern — (KGP3) muß durch (KGP3*) ersetzt werden —, und machte zugleich klar, daß das „falsche“ dritte Gesetz das „richtige“ in einem natürlichen und einfachen Sinne approximiert.

Ich habe versucht, vermittels kontrafaktischer Konditionalsätze aufzuzeigen, wie man **NTG** als überlegene Nachfolgertheorie für **KGP*** erkennen kann, d.h. wie es möglich ist, daß **NTG** sowohl **KGP*** als auch deren Scheitern erklärt. Mein Ergebnis stimmt gut überein mit den Analysen des Wissenschaftshistorikers und Newton-Spezialisten I. Bernard Cohen (1974, S. 319): „Newton's ‚theory' first displays the hypothetical circumstances under which ... Kepler's laws are valid, and then shows the modified form in which all three of these laws occur in Newtonian dynamics.“ Die durch **NTG** geleistete doppelte Erklärung im Sinne von Kapitel 7 stellt bestimmten idealisierenden Randbedingungen kontrastiv die tatsächlich vorliegenden Randbedingungen gegenüber. Oder wie Cohen (1974, S. 315) es ausdrückt: „In this 'real world', as opposed to the hypothesized world presented in the earlier sections of the *Principia*, Kepler's laws will no longer hold exactly.“[41] Der wesentliche Punkt im Kepler-Newton-Beispiel ist, meine ich, die Opposition von *hypothetischer Wahrheit* und *tatsächlicher Falschheit*, nicht die Opposition von approximativer und strikter Wahrheit. Ich möchte nicht den heuristischen oder praktischen Wert der Tatsache bestreiten, daß **KGP** sowohl der empirischen Wahrheit als auch den **NTG**-Konsequenzen für das Planetensystem sehr nahe kommen, aber insofern als unser Anliegen theoretische Erklärungen und abstrakte intertheoretische Relationen sind, wird dies zu einem eher nebensächlichen Phänomen.

[40] Der Terminus stammt von Kuipers (1985, S. 186) und deutet an, daß der Übergang von **KGP** zu **KGP*** sich auf der Basis von theoretischen Hintergrundüberlegungen in **NTG** vollzieht, nicht auf der Basis von empirischer Evidenz. Wir können dies in einem Schema fassen, welches parallel ist zu Kuipers' suggestivem Schema für den Fall des idealen und des Van der Waalsschen Gasgesetzes:

1. Ebene	**KGP**	$\longrightarrow$	$\mathbf{NTG^*}_{A_1}$
			$\downarrow$
2. Ebene	**KGP***	$\longleftarrow$	$\mathbf{NTG^*}_{A_2}$
	kinematische Ebene		*dynamische Ebene.*

[41] Vergleiche auch den letzten Absatz von Cohens Fußnote 30. Nebenbei bemerkt vertritt Cohen sogar die These, daß die Darstellung in den *Principia* mehr oder weniger genau den historischen Lauf der Entdeckungen Newtons wiedergibt (1974, S. 332).

Wir haben in diesem Teilkapitel die axiomatische Struktur von **NTG** nicht explizit dargelegt, und wir haben a fortiori versäumt, die für die Ausführung von **NTG**-Revisionen nötige Relation der theoretischen Wichtigkeit anzugeben. Die Nachlässigkeit scheint allerdings verzeihlich, denn wir können diese Informationen wenigstens in groben Zügen unschwer nachliefern. Wie mehrfach deutlich geworden ist, besteht nach meiner Auffassung **NTG** keineswegs nur aus dem Gravitationsgesetz Newtons. Dieses ist zu ergänzen einerseits durch die noch fundamentaleren drei Grundgesetze der Newtonschen Mechanik, andererseits soll **NTG** als umfassende Theorie über unsere Welt auch empirisches Wissen (z.B. über die Anzahl und Massen der Planeten) beinhalten. Im Konfliktfalle, wie er bei kontrafaktischen („kontratheoretischen") Revisionen auftritt (z.B. um **KGP** hypothetisch zu akzeptieren), waren es stets kontingente Sätze, die aus **NTG** weichen mußten, niemals ein Gesetz. Dies entspricht der Gärdenforsschen Idee, daß Gesetzesartigkeit als Resistenz gegenüber (hypothetischen) Revisionen expliziert werden kann. Wir werden nun im nächsten Teilkapitel ein Beispiel untersuchen, bei welchem eine explizite Betrachtung der Relation der theoretischen Wichtigkeit lohnender ist.

8.2 Das ideale und das van der Waalssche Gasgesetz

8.2.1 Die Inkonsistenz der idealen und der van der Waalsschen Gastheorie

Das ideale Gasgesetz und das van der Waalssche Gasgesetz bilden zusammen ein Paar, das neben dem Kepler-Newton-Fall als vielleicht einfachstes und prägnantestes Musterbeispiel einer bestimmten Nachfolgerrelation zwischen wissenschaftlichen Gesetzen oder Theorien T_1 und T_2 dienen kann. Diese Relation ist dadurch gekennzeichnet, daß sich T_1 und T_2 in gewissem Sinn widersprechen,[42] daß man aber dennoch eine Reduktion oder

[42] Es wird viel zu selten geklärt, in welchem Sinn sich etwa die Keplerschen Gesetze und die Newtonsche Gravitationstheorie oder eben das ideale und das van der Waalssche Gasgesetz widersprechen. Der Widerspruch kann nicht darin bestehen, daß die jeweiligen (gesetzesartigen) Gleichungen (KGP) und (NTG) bzw. unten (IGG) und (VDW) keine gemeinsamen Lösungen haben, denn das haben sie (siehe Scheibe 1973a über Karussellmodelle und Abschnitt 8.2.3. über die Idealkurve). Ein Widerspruch ist nur relativ zu gewissen Zusatzannahmen (über empirische „Randbedingungen") nachweisbar; vgl. Abschnitt 8.2.4.

eine intertheoretische Erklärung zwischen T_1 und T_2 ansetzen möchte, um die offenkundige Kontinuität beim Übergang von T_1 zu T_2 wiedergeben zu können. Wie beim Kepler-Newton-Beispiel werden wir auch im Fall des idealen und des van der Waalsschen Gasgesetzes als Explikate dieser Nachfolgerrelation intertheoretische Approximationen und intertheoretische Idealisierungen betrachten. Der quantitative Begriff der intertheoretischen Approximation soll wieder im zweiten Sinn verstanden werden (vgl. Abschnitt 2.4) und besagen, daß die Gleichung(en) von T_2 beim Übergang gewisser Größen gegen einen Grenzwert Lösungen haben, die bzgl. einer geeigneten Topologie den Lösungen der Gleichung(en) von T_1 beliebig nahe kommen. Demgegenüber soll, wie erinnerlich, der qualitative Begriff der intertheoretischen Idealisierung ausdrücken, daß T_1 „vom Standpunkt von T_2 aus" als *ganz genau* zutreffend erwiesen werden kann, falls man T_2 durch geeignete idealisierende, kontrafaktische Annahmen („Anwendungsbedingungen für T_1") revidiert.

Während es sich bei der Newtonschen Gravitationstheorie von selbst versteht, für wie wichtig ihre Entdeckung einzuschätzen ist, ist die ungeheure historische Bedeutung und der revolutionäre Gehalt der van der Waalsschen Zustandsgleichung heute etwas in Vergessenheit geraten. Wir können dies hier nur andeuten durch die Boltzmannsche Bemerkung in der *Enzykopädie der mathematischen Wissenschaften*, wo er van der Waals als den „Newton für die Theorie der Abweichungen der Gase vom Boyle-Charlesschen Gesetz" (Boltzmann und Nabl 1905, S. 550) bezeichnet, und indem wir daran erinnern, daß die van der Waalssche Zustandsgleichung ihrem Entdecker 1910 den Nobelpreis für Physik einbrachte.[43]

Anders als beim Kepler-Newton-Fall können wir uns im vorliegenden Beispiel weder auf eine stattliche Anzahl von Vorrednern stützen, die die Unverträglichkeit der beiden Gasgesetze herausstreichen, noch haben wir als Vorgabe ausführliche wissenschaftstheoretische Analysen im Rahmen approximativer Ansätze. Dies ist einerseits bedauerlich, andererseits bietet es uns den Anreiz, gleich mit einer systematischeren, wenn auch stark vereinfachten Axiomatisierung der Beispieltheorien zu beginnen, die — in bescheidenem Maße — Betrachtungen über theoretische Wichtigkeit gestattet. Wir werden also zunächst daran gehen, die ideale und die van der Waalssche Gastheorie präzise darzustellen. Diese *Theorien* enthalten in meiner Rekonstruktion als explizite Axiome neben den eigentlichen Zustandsgleichungen noch einige wenige Bedingungen, die man üblicher-

[43] Einen lebhaften Eindruck von der Bedeutung der van der Waalsschen Gasgleichung vermitteln Boltzmann (1898) und Kamerlingh Onnes und Keesom (1912).

weise als Hintergrundwissen oder Präsuppositionen der Gasgesetze auffassen würde. Dann werde ich den Approximations- und den Idealisierungsansatz in ihrer Anwendung auf das vorliegende Beispielspaar prüfen. Da Approximationen und Idealisierungen als Explikate für das Bestehen einer *Reduktion* oder *intertheoretischen Erklärung* gedacht sind, werde ich, in Fortsetzung der klassischen Nagel-Hempelschen Schemata (s. Kapitel 2), mein Augenmerk wieder vor allem darauf richten, ob sich das ideale Gasgesetz auf irgendeine Art und Weise aus dem van der Waalsschen Gasgesetz herleiten läßt. Eine Herleitung kann allerdings nicht auf einfache, direkte Weise vorgenommen werden, weil sich die ideale und die van der Waalssche Gastheorie widersprechen. Nimmt man das Gelingen einer solchen Herleitung zum Maßstab der Möglichkeit einer Reduktion, dann, so wird sich zeigen, bieten Idealisierungen ein ansprechenderes Bild als Approximationen.

Indem wir uns der präzisen Formulierung der idealen und der van der Waalsschen Gastheorie zuwenden, bemerken wir zunächst folgendes: Das ideale Gasgesetz macht nur Aussagen über *Gase im eigentlichen Sinn* (G_e), während das van der Waalssche Gasgesetz auch über Dämpfe, Flüssigkeiten und gasförmig-flüssige Zweiphasensysteme spricht, die wir unter dem Prädikat „*Gase im weiteren Sinn*" (G_w) zusammenfassen wollen. Sowohl das ideale als auch das van der Waalssche Gasgesetz ist eine Gleichung für die Zustandsvariablen Druck (p), (molares) Volumen (V) und (absolute) Temperatur (T) eines beliebigen Systems (S) x eines beliebigen Gases X (im engeren oder weiteren Sinn).[44] Mit a(X) und b(X) als den stoffspezifischen van der Waalsschen Konstanten für den Kohäsionsdruck und das Kovolumen und der positiven reellen Zahl R als „universeller Gaskonstante" können wir die beiden Zustandsgleichungen für ein Mol Gas folgendermaßen anschreiben:

(IGG) $\forall X\, \forall x\, (G_eX \wedge xSX \rightarrow p(x)V(x)=RT(x))$.

(VDW) $\forall X\, \forall x\, (G_wX \wedge xSX \rightarrow (p(x)+(a(X)/V(x)^2))(V(x)-b(X)) = RT(x))$.

In der (IGG) und (VDW) gemeinsamen Objektsprache sind p, V und T

[44] Diese Systeme sollen sich bei jeder Messung im thermischen Gleichgewicht befinden befinden, sind (zwischen den Messungen) im allgemeinen aber *keine* thermisch abgeschlossenen Systeme. Prozesse ohne jeden Wärmeaustausch mit der Umgebung, sog. adiabatische Prozesse, gehorchen (auch bei maximaler Idealisierung) zum Beispiel nicht dem Boyle-Mariotteschen Gesetz pV=konstant (welches für T=konstant eine bekannte Folgerung aus dem idealen Gasgesetz ist), sondern der Adiabatengleichung pV^{c_p/c_v}=konstant. vgl. Abschnitt 8.2.4.

dabei Konstanten für Funktionen, die allen Gas*systemen* reelle Zahlen zuordnen. Im Vokabular des van der Waalsschen Gasgesetzes bezeichnen a und b ebenfalls Funktionen, die aber die Gase (i.w.S.) selbst als Argumente nehmen.

Ich werde im folgenden davon ausgehen, daß das ideale und das van der Waalssche Gasgesetz keine isolierten Sätze sind, sondern daß sie Minitheorien konstituieren, welche in (IGG) und (VDW) stillschweigend gemachte Hintergrundannahmen oder stillschweigend verwendetes Hintergrundwissen explizit mit enthalten. Folgende Annahme soll sowohl Teil der idealen als auch der van der Waalsschen Gastheorie sein:

(8.2.1) Zwei der drei Zustandsgrößen sind frei variierbar.

Aus physikalischen Gründen wäre es plausibel, des weiteren zu fordern, daß Temperatur und Volumen stets größer als null sein sollen; $V(x)\neq 0$ wird ja schon mathematisch von (VDW) präsupponiert. Aus (IGG), nicht aber aus (VDW) folgt dann, daß auch der Druck nur Werte in $\mathbb{R}^+$ annehmen kann. Da wir diese Forderungen aber nicht brauchen werden, will ich sie der Übersichtlichkeit halber unterschlagen. Weil die Variabilität der Zustandsgrößen zumindest nach (VDW) (welches $V(x)<b(X)$ verbietet) doch keine *völlig* freie ist, werden wir uns später (in Abschnitt 8.2.3), wenn es um ein formales Resultat geht, mit einer viel schwächeren Bedingung anstelle von (8.2.1) begnügen. Sie besagt, daß bei vorgegebener Temperatur eines Systems ohne Festlegung seines Molvolumens zumindest zwei verschiedene Drücke möglich sind:

(8.2.2) $\forall X\, \forall x\, (G_e X \wedge xSX \rightarrow \exists y\, (ySX \wedge T(y)=T(x) \wedge p(y)\neq p(x)))$.

Die zweite wichtige Hintergrundbedingung, die zumindest Teil der van der Waalsschen Gastheorie sein soll, ist

(8.2.3) Alle Gase im engeren Sinn sind auch Gase im weiteren Sinn.

Diese Annahme dürfte völlig unproblematisch sein, da (VDW) einen mindestens gleich großen Anwendungsbereich wie (IGG) hat. (Die ebenfalls triviale Aussage, daß es Gase im weiteren Sinn gibt, die keine Gase im engeren Sinn sind, werden wir nicht benötigen.) Die dritte und letzte Hintergrundvoraussetzung, welche wir explizit mit in das van der Waalssche Gesetz aufnehmen wollen, ist

(8.2.4) Für alle Gase im weiteren Sinn sind die van der Waalsschen Konstanten größer als null.

(8.2.4) kann schon deshalb nicht zur idealen Gastheorie gehören, weil a und b in deren Sprache gar nicht vorkommt. Aus den in Abschnitt 8.2.3 folgenden Bemerkungen über die kinetische Interpretation der Zustands-

gleichungen wird jedoch hervorgehen, daß (8.2.4) zum einen ein wichtiger Bestandteil des van der Waalsschen Gesetzes, zum anderen aber, von der Warte van der Waals' aus beurteilt, falsch für den Geltungsbereich des idealen Gasgesetzes sein muß.

Intuitiv würde man meinen, schon die Zustandsgleichungen an sich müßten inkonsistent sein, da sie ja für ein bestimmtes Gas X (im engeren Sinn) unterschiedliche Zusammenhänge zwischen den Größen p, V und T in X-Systemen voraussagen. Bei einer genaueren formalen Prüfung kommt aber zum Vorschein, daß man tatsächlich das in (8.2.1) (oder (8.2.2)), (8.2.3) und (8.2.4) enthaltene Hintergrundwissen braucht, um einen Widerspruch abzuleiten. Wir verschieben diese Prüfung auf später (s. das Ende von Abschnitt 8.2.3) und wollen hier nur eine entsprechende terminologische Entscheidung treffen. Wenn im folgenden von der *van der Waalsschen Gastheorie* die Rede ist, dann ist eine kleine Theorie gemeint, die (VDW) und (8.2.1), (8.2.3) und (8.2.4) als Axiome hat; sie sei durch „**VDW**" bezeichnet. Die *ideale Gastheorie* umfasse (IGG) und (8.2.1) und werde durch „**IGG**" bezeichnet.

8.2.2 Das Beispiel des idealen und des van der Waalsschen Gasgesetzes als ein Fall von Approximation

Eine harmlose Idee von Annäherung oder Approximation wäre es, zu sagen, daß das ideale Gasgesetz für einen gewissen Bereich von Druck-, Volumen- und Temperaturwerten annähernd richtige Voraussagen liefert — wenn das van der Waalsche Gasgesetz als Maßstab der Richtigkeit genommen wird. Dies ist die Approximation im ersten Sinn (vgl. Abschnitt 2.4), wird bei diesem Beispiel aber normalerweise nicht gemeint. Wirklich intendiert ist der interessantere zweite Sinn, nach dem das van der Waalssche Gasgesetz „in das ideale Gasgesetz übergeht", wenn gewisse Größen „gegen einen Grenzwert gehen". Die Frage, *welche* Grenzwertübergänge denn eigentlich vollzogen werden sollen, versucht man am besten mit einen Blick in die Lehrbücher zu beantworten. Die dort zu findenden Angaben hinterlassen aber eine gewisse Verwirrung. Der üblichste Vorschlag ist es, das Molvolumen gegen unendlich[45] oder, was auf dasselbe hinausläuft, die Dichte ϱ gegen null[46] gehen zu lassen. Gelegentlich läßt man den Druck gegen null

[45] Siehe Sommerfeld (1952, S. 53), Frauenfelder und Huber (1968, S. 362) und Elsner (1980, S. 162) (und auch Glymour 1970, S. 343f).

[46] Siehe Sommerfeld (1952, S. 7) und Alonso und Finn (1980, S. 418).

gehen,[47] und man findet auch die Idee, die Temperatur gegen unendlich gehen zu lassen.[48] Außerdem werden häufig (konjunktive oder adjunktive) Kombinationen der Bedingungen an die Zustandsgrößen genannt.[49] Für meine weitere Argumentation ist die Wahl der Grenzwertbedingung nicht wesentlich, weshalb ich der Einfachheit halber bei dem meistgenannten Vorschlag „$V \to \infty$" (bzw. dem äquivalenten „$\varrho \to 0$") bleibe.

Wenn das van der Waalssche Gasgesetz beim Grenzübergang, also speziell bei $V \to \infty$, wirklich in das ideale Gasgesetz übergeht, dann kann man nach der Idee des Approximationsmodells sagen, daß das ideale auf das van der Waalssche Gasgesetz *approximativ reduzierbar* ist oder daß das van der Waalssche das ideale Gasgesetz *approximativ erklärt*.[50]

Die Ausdrucksweise, daß man gewisse Größen „gegen einen Grenzwert gehen läßt" und daß dabei eine Gleichung „in eine andere Gleichung übergeht", ist natürlich eine metaphorische. Bei der Präzisierung des Approximationsgedankens werde ich mich nun nicht auf Modellbetrachtungen verlegen (bei denen meiner Einschätzung nach ganz ähnliche Probleme auftreten würden), sondern mich der alten, eher dem Statement view verpflichteten Idee der logischen Positivisten zuwenden, wonach mit einer Reduktion oder Erklärung eine Art approximativer Ableitbarkeit von (IGG) aus (VDW) gegeben sein müßte. Das dieser Vorstellung ursprünglich zugrundeliegende Hempel-Oppenheim-Schema einer wissenschaftlichen Erklärung kann heute keine allgemeine Gültigkeit mehr beanspruchen, aber das Ableitbarkeitskriterium als notwendige Bedingung für Erklärungen scheint mir trotzdem eine sehr wichtige Intuition wiederzugeben. Wir wollen nun nachprüfen, inwieweit der Approximationsansatz dieses Kriterium erfüllt. Um vom van der Waalsschen Gesetz näher an das ideale Gasgesetz heranzukommen, benötigen wir zunächst einmal einen mathematischen Satz nach Art von

[47] Siehe Sears und Zemansky (1964, S. 451), Schmidt, Stephan und Mayinger (1975, S. 222) und Elsner (1980, S. 26, 120, 124).

[48] Siehe Gehrtsen und Kneser (1969, S. 159) und Reichl (1980, S. 99).

[49] Bedingungen an Volumen und Temperatur nennen Boltzmann (1898, S. 22), Frauenfelder und Huber (1968, S. 363) und Schilling (1972, S. 168), Bedingungen an Volumen und Druck nennen Schaefer (1958, S. 114) und Schmidt, Stephan und Mayinger (1975, S. 227) (und auch Scheibe 1984, S. 87f), Bedingungen an Druck und Temperatur nennen Gehrtsen und Kneser (1969, S. 131). Eine Bedingung an Druck oder Volumen nennt Elsner (1980, S. 125), eine Bedingung an Temperatur oder Dichte nennen Alonso und Finn (1980, S. 421).

[50] Für Krajewski ist das ideale auf das van der Waalssche Gasgesetz reduzierbar, kann aber nicht durch dieses erklärt werden: „We should rather say that in order to explain van der Waals' law we must first consider Boyle-Mariotte's law and then take into account additional factors" (Krajewski 1984, S. 13).

(8.2.5) $\forall X\ \forall \varepsilon\ \exists \delta_{\varepsilon,X}\ \forall x\ (V(x)>\delta_{\varepsilon,X} \rightarrow \Psi[x,X] \approx_\varepsilon \Phi[x])$,

wobei „$\Psi[x,X]$" das Konsequens von (VDW) und „$\Phi[x]$" das Konsequens von (IGG) abkürzt. Man beachte, daß δ nicht nur von ε, sondern auch von a(X) und b(X), also von der Art des Gases X abhängig ist.[51]

Natürlich ist zu sagen, was die Bedingung „$\Psi[x,X] \approx_\varepsilon \Phi[x]$" heißen soll. Offenbar will man damit ausdrücken, daß die Aussage des van der Waalsschen Gasgesetzes nur noch „im Grade ε" von der Aussage des idealen Gasgesetzes entfernt ist. Ein naheliegender Versuch, dies zu explizieren, wäre der folgende:

8.2.6. Definition Sei X ein Gas mit den van der Waals-Konstanten a(X) und b(X) und sei xSX. Dann ist $\Psi[x,X]\approx_\varepsilon\Phi[x]$ genau dann erfüllt, wenn gilt: Erfüllen die Tripel $\langle p_\Phi,V(x),T(x)\rangle$, $\langle p(x),V_\Phi,T(x)\rangle$ und $\langle p(x),V(x),T_\Phi\rangle$ das Schema $\Phi[x]$ und die Tripel $\langle p_\Psi,V(x),T(x)\rangle$, $\langle p(x),V_\Psi,T(x)\rangle$ und $\langle p(x),V(x),T_\Psi\rangle$ das Schema $\Psi[x,X]$, so ist $|p_\Psi-p_\Phi|<\varepsilon$, $|V_\Psi-V_\Phi|<\varepsilon$ und $|T_\Psi-T_\Phi|<\varepsilon$.[52]

Nach Definition 8.2.6 sind (IGG) und (VDW) für ein X-System x also genau dann ε-nahe, wenn sich bei der Berechnung

- von p aus gegebenem V(x) und T(x),
- von V aus gegebenem p(x) und T(x)[53] und
- von T aus gegebenem p(x) und V(x),

die zunächst gemäß $\Psi[x,X]$ und dann gemäß $\Phi[x]$ durchgeführt wird, Abweichungen ergeben, die sowohl für p als auch für V als auch für T kleiner als ε sind.

Wir wollen nun nachprüfen, ob diese Definition mit der Idee von (8.2.5) zusammenpaßt. Hierzu berechnen wir bei vorgegebenen a und b (die Argumente X seien der Einfachheit halber weggelassen) nacheinander $|p_\Psi-p_\Phi|$, $|V_\Psi-V_\Phi|$ und $|T_\Psi-T_\Phi|$:

[51] Diese unangenehme Abhängigkeit kann man beseitigen, indem man davon ausgeht, daß man es nur mit endlich vielen Arten von Gasen zu tun hat, so daß $\delta_\varepsilon=\max_X \delta_{\varepsilon,X}$ gewählt werden kann.

[52] Auch eine Betrachtung der relativen Differenzen, d.h. die Bedingungen $|(p_\Psi-p_\Phi)/p_\Phi|<\varepsilon$, $|(V_\Psi-V_\Phi)/V_\Phi|<\varepsilon$ und $|(T_\Psi-T_\Phi)/T_\Phi|<\varepsilon$, wären möglich, würde aber, wie man nachprüft, die im folgenden beschriebene Situation nicht verbessern.

[53] Man beachte, daß sich V(x) gemäß dem van der Waalsschen Gasgesetz nicht mehr als Funktion der Werte von p(x) und T(x) darstellen läßt, sondern daß für einen gewissen Bereich von (relativ niedrigen) Drücken und Temperaturen nach $\Psi[x,X]$ drei Werte für das molare Volumen zulässig sind.

(8.2.7) $|p_\Psi - p_\Phi| = |RT/(V-b) + a/V^2 - RT/V| = |bRT/(V^2-bV) - a/V^2|;$
$|V_\Psi - V_\Phi| = |RT/(p+a/V_\Psi{}^2) + b - RT/p| = |aRT/p(a-pV_\Psi{}^2) + b|;$
$|T_\Psi - T_\Phi| = |(p+a/V^2)(V-b) - pV| \,/\, R = |a/V - ab/V^2 - bp| \,/\, R.$

Von keinem dieser Beträge kann man ohne weiteres sagen, er gehe für $V \to \infty$ (bzw. $V_\Psi \to \infty$) gegen null, denn überall haben wir noch andere Zustandsgrößen als das Volumen enthalten. Versuchen wir es aber einmal schrittweise. Das übersichtlichste Ergebnis brachte der Temperaturvergleich: Wenn V gegen ∞ und wenn gleichzeitig p gegen 0 geht, dann geht der Unterschied zwischen T_Ψ und T_Φ gegen 0. Danach bringt der Druckvergleich das Ergebnis, daß T/V^2 gegen 0 gehen muß, was z.B. erfüllt ist, wenn T (nach oben und unten) beschränkt ist. In der Abschätzung der Temperaturdifferenz geht der Term $aRT/p(a-pV_\Psi{}^2)$ gegen 0, wenn $(ap-p^2V_\Psi{}^2)/T$ gegen $\pm\infty$ geht. Wir wissen schon, daß p gegen 0 gehen und T beschränkt sein soll, weshalb einerseits der Term ap/T gegen 0 geht. Andererseits geht der Term $p^2V_\Psi{}^2/T$ nur dann gegen ∞, wenn p viel langsamer gegen 0 geht als V_Ψ gegen ∞, oder wenn T noch besonders schnell gegen 0 geht. (Falls der Term $aRT/p(a-pV_\Psi{}^2)$ tatsächlich gegen 0 gehen sollte, so haben wir aber immer noch eine kleine, aber positive Temperaturdifferenz mit Wert b.) Zudem ist als weitere Beschränkung zu berücksichtigen, daß zwischen den Werten von p, V und T für reale Gassysteme eine Abhängigkeit besteht (die in etwa durch das van der Waalssche Gasgesetz wiedergegeben wird), so daß man die Variablen nicht frei gegen irgendwelche Grenzwerte laufen lassen kann. Wir sehen also, daß die Grenzwertbetrachtungen weitaus komplexer sind, als angenommen wurde, und ziehen aus dieser Diskussion den Schluß, daß die Bedingung (8.2.5) noch eine beträchtliche Schuld für den Approximationsansatz offen läßt.

Natürlich ist es aber durchaus möglich, daß eine raffiniertere Definition von $\Psi[x,X] \approx_\varepsilon \Phi[x]$ und eine kunstvolle Abschätzung der Abweichungen des van der Waalsschen vom idealen Gasgesetz gefunden werden kann, die zeigen, daß die Bedingung (8.2.5) — oder eine entsprechende Bedingung mit anderem Antezedens — erfüllt ist. Gestehen wir den Befürwortern einer Approximation also einmal zu, daß (8.2.5) korrekt ist und als rein mathematischer Satz Teil der van der Waalsschen Gastheorie sein darf. Selbst dann bekommt man aus dieser Theorie noch kein Gesetz, das eine Annäherung an das ideale Gasgesetz leistet. Das Ziel ist nämlich die sukzessive Ableitung der ε-abhängigen „Annäherungsgesetze"

(IGG_ε) $\forall X\, \forall x\, (G_e X \wedge xSX \to \Psi[x,X] \wedge \Psi[x,X] \approx_\varepsilon \Phi[x])$,

die dem idealen Gasgesetz immer ähnlicher werden. Um aber auf deduktivem Wege ein Gesetz zu erhalten, das ε-nahe am idealen Gasgesetz liegt, braucht man noch eine zusätzliche Voraussetzung, und zwar

$A_\varepsilon \qquad \forall X\, \forall x\, (G_e X \wedge xSX \rightarrow V(x) > \delta_{\varepsilon,X})$.

Diese Bedingung besagt, daß alle Systeme x eines Gases X im engeren Sinne ein Molvolumen haben, das über $\delta_{\varepsilon,X}$ liegt. Aus den Prämissen (VDW), (8.2.3), (8.2.5) und A_ε ist nun (IGG_ε) endlich deduzierbar. Da wir ε frei, also beliebig klein vorgeben können, kann (IGG_ε) beliebig nahe am idealen Gasgesetz gewählt werden, und es scheint so, als hätten wir das ideale Gasgesetz tatsächlich approximativ aus dem van der Waalsschen abgeleitet.

Intuitiv gesehen, ist dies jedoch ein Trugschluß. Denn mit kleiner werdendem ε wächst δ_ε immer mehr, und es wird äußerst fraglich, ob man dann A_ε noch als sinnvoll bezeichnen kann. Obgleich der Sinn des Gasbegriffs nicht von vornherein strikt festgelegt ist,[54] darf man A_ε nicht als definitorische Bedingung für Gase (im engeren Sinn) auffassen, da es ja von ε abhängig ist, und für irgendein ε wird $\delta_{\varepsilon,X}$ ganz sicher zu hoch sein. Andererseits ist A_ε als empirisch aufgefaßte Bedingung für die meisten ε schlichtweg falsch. Wenn wir also zur Ableitung von (IGG_ε) Instanzen des Schemas A_ε als Prämissen verwenden — und ich sehe nicht, wie wir das vermeiden könnten —, dann spricht (IGG_ε) nicht mehr über alle Gase im engeren Sinn, sondern nur noch über ganz bestimmte Systeme solcher Gase. Je kleiner ε wird, über desto weniger Systeme macht (IGG_ε) noch eine Aussage. Was wir abgeleitet haben, sind also nicht approximative Versionen des idealen Gasgesetzes, das über alle Gassysteme im engeren Sinn spricht, sondern nur approximative Versionen eines Gesetzes mit zusehends verschwindendem Anwendungsbereich.

Wenn intertheoretische Reduktion und Erklärung also stets die Ableitbarkeit der (eventuell modifizierten) Gesetze von T_1 aus T_2 implizieren soll, dann gelingt es im approximativen Ansatz nicht, das ideale auf das van der Waalssche Gasgesetz zu reduzieren oder jenes durch dieses zu erklären. Der Ansatz verfehlt das Ziel einer Reduktion oder Erklärung im klassischen, deduktiv-nomologischen Sinne. Man kann nun die argumentative Strategie verfolgen, daß eine Ableitbarkeit auch in diesem speziellen, sehr einfachen Fall gar nicht notwendig, daß das klassische Schema hier also unangebracht sei. Ich werde aber unten zu zeigen versuchen, daß ein Verzicht auf die

[54] Tatsächlich zwang die van der Waalssche Zustandsgleichung dazu, den Gasbegriff neu zu überdenken. Siehe Abschnitt 8.2.5.

Ableitbarkeitsbedingung nicht nötig ist.

Eine Möglichkeit, die mit kleiner werdendem ε immer drastischeren Bereichseinschränkungen durch A_ε abzufangen, bestünde in der Variation der Parameter a und b. Dieser Versuch wird in den physikalischen Lehrbüchern aber nicht unternommen. Sommerfelds Bemerkung

> Für a=b=0 oder, was dasselbe ist, für hinreichend großes v geht (1) [die van der Waalssche Gleichung], wie es sein muß, in die ideale Gasgleichung über. (Sommerfeld 1952, S. 53)

ist, gelinde gesagt, ungenau und steht in der Literatur allein auf weiter Flur. Der Grund dafür, daß Physiker an den Werten von a und b nicht manipulieren wollen, scheint zu sein, daß das van der Waalssche Gesetz ja das Verhalten realer Gase wiedergeben soll und daß jedes reale Gas X eben feste, dem Gas eigentümliche Van-der-Waals-Konstanten a(X) und b(X) hat. Die Annahme immer kleiner werdender Werte von a und b entfernt sich deshalb immer mehr von der Wahrheit, sie wird — anders gesagt — mehr und mehr kontrafaktisch. Anstelle einer zunehmenden Bereichseinschränkung ist man hier mit einer zunehmenden Kontrafaktizität konfrontiert. Aus der Idealisierungsperspektive ist dies aber durchaus nichts Ungewöhnliches. Will man keine deflationäre Bereichseinschränkung betreiben, dann, so scheint es hier wieder, kommt man bei Approximationen um kontrafaktische Prämissen nicht herum. Die Umkehrung gilt, wie schon in Abschnitt 8.1, jedoch nicht: Wenn man mit kontrafaktischen Annahmen umzugehen vermag, dann kann man auf Approximationsprozesse verzichten. Ich werde nun versuchen, diese These auch am Fall des idealen und des van der Waalsschen Gasgesetzes zu erhärten.

8.2.3 Das Beispiel des idealen und des van der Waalsschen Gasgesetzes als ein Fall von Idealisierung

Es muß nicht immer von vornherein klar sein, ob eine historisch gegebene Theorie eine idealisierte Theorie im Sinne des Dekompositionsgedankens ist. Bis zur Mitte des 19. Jahrhunderts zum Beispiel wurden das Boylesche (1662) und das Gay-Lussacsche (1802) Gasgesetz — die zusammen das ideale Gasgesetz ausmachen[55] — nicht als Idealisierung aufgefaßt. Doch

[55] Vgl. Partington (1949, S. 606f). Sonderbarerweise wurde die heute übliche Formulierung des idealen Gasgesetzes mit der universellen Gaskonstante erst 1873, also dem Promotionsjahr van der Waals', von Horstmann gegeben (laut Partington).

nachdem Rudolf Clausius 1857 seine berühmte Arbeit *Über die Art der Bewegung, welche wir Wärme nennen* veröffentlicht hatte, fiel es zumindest den Anhängern der kinetischen Gastheorie wie Schuppen von den Augen.[56]

Clausius nannte die folgenden — offenbar als hinreichend und notwendig gedachten — Bedingungen für die Gültigkeit des idealen Gasgesetzes:

1. Der Raum, welchen die Moleküle des Gases wirklich ausfüllen, muß gegen den ganzen Raum, welchen das Gas einnimmt, verschwindend klein sein.
2. Die Zeit eines Stoßes ... muß gegen die Zeit, welche zwischen zwei Stößen vergeht, verschindend klein sein.
3. Der Einfluß der Molekularkräfte muß verschwindend klein sein. (Clausius 1970, S. 170)

In der Ableitung des idealen Gasgesetzes aus der kinetischen Gastheorie muß man, damit es eine strenge Ableitung ist, die „verschwindend kleinen" Größen *gleich null setzen*. Es ist aber offensichtlich, daß dies eine Idealisierung ist und daß die Größen in Wirklichkeit *nicht* gleich null sind. Folglich ist zu erwarten, daß das ideale Gasgesetz einer Korrektur bedarf.[57] Van der Waals, dem die Arbeit von Clausius „eine Offenbarung" (van der Waals 1911, S. 4) war, ging nun daran, die ersten Schritte der programmgemäßen, sozusagen notwendigen und dennoch höchst kreativen „De-idealisierung" oder „Konkretisierung" des idealen Gasgesetzes zu vollziehen. Er berücksichtigte zwei Faktoren: die Anziehung zwischen den Teilchen und das Volumen, das sie einnehmen. Für die Zustandsgleichung schlug sich dies in den beiden stoffspezifischen Korrekturkonstanten a und b nieder, welche den sogenannten Binnendruck (Kohäsionsdruck) und das sogenannte Kovolumen (Sperrvolumen) repräsentieren. Die van der Waalsschen Konstanten a und b sind ein auf die phänomenologische Ebene transportiertes Maß für die Größe der intermolekularen Anziehungskräfte bzw. für das Eigenvolumen der Moleküle eines „realen" Gases. Keiner dieser beiden Faktoren wird im idealen Gasgesetz berücksichtigt. Wie Clausius vor Augen führte, sagt das ideale Gasgesetz, was der Fall wäre, wenn sich die Gasmoleküle nicht

[56] Wie Clausius erwähnt, wurden einige seiner Ideen schon ein Jahr früher von August Krönig vorweggenommen. Historisch ungleich wichtiger ist aber der Aufsatz von Clausius.

[57] Eine Korrektur ist nicht *immer* notwendig. Zum Beispiel hat Newton bei seiner Berechnung der Planetenbahnen gezeigt, daß man die Planeten als Massenpunkte betrachten darf, ohne daß dies (gegenüber der Annahme, daß Planeten große Kugeln sind) irgendeine Auswirkung hat. vgl. Fußnote 32.

anzögen (und abstießen) und wenn sie ausdehnungslos, also punktförmig wären. Entsprechend wird ein Verfechter des van der Waalsschen Gasgesetzes wohl die folgenden Sätze akzeptieren: Wenn a und b gleich null wären, dann würde das ideale Gasgesetz gelten; weil sie aber nicht gleich null sind, deshalb gilt das ideale Gasgesetz nicht.[58]

Nach den in Kapitel 7 eingeführten Begriffen kann man nun sagen, daß das van der Waalssche Gasgesetz das ideale Gasgesetz als Idealisierung oder kontrafaktisch erklärt, während das, was es faktisch erklärt, das Scheitern des idealen Gasgesetzes ist. Gelingt uns der Nachweis, daß das van der Waalssche Gesetz diese doppelte Erklärungsleistung tatsächlich vollbringt, dann kann man nach Definition 7.2.2 die van der Waalssche Gastheorie eine überlegene Nachfolgertheorie für die ideale Gastheorie nennen. Wenn wir Theorem 7.4.2 anwenden, erhalten wir folgendes Kriterium: **VDW** ist genau dann eine überlegene Nachfolgertheorie für **IGG**, wenn es (von **VDW** aus gesehen) eine Anwendungsbedingung A für **IGG** gibt, so daß gilt: Die minimale Abänderung von **VDW**, die nötig ist, um A „wahr zu machen", erlaubt die Ableitung von **IGG**, und umgekehrt erlaubt die minimale Abänderung von **VDW**, die nötig ist, um **IGG** wahr zu machen, die Ableitung von A.

Ähnlich wie im Kepler-Newton-Fall soll die idealisierende Anwendungsbedingung A auch hier intuitiv angemessen, von nicht gesetzesartigem Charakter und nicht im (unauflösbaren) Konflikt mit **VDW** sein. Der passende Kandidat für A wurde oben schon erwähnt:

A Für alle Gase im engeren Sinn sind die van der Waalsschen Konstanten gleich null.

Es ist unmittelbar klar, daß A der Konjunktion von (8.2.3) und (8.2.4) und damit der van der Waalschen Gastheorie **VDW** widerspricht. Deshalb ist A relativ zu **VDW** eine idealisierende Anwendungsbedingung. Außerdem widerspricht, wie wir in Kürze nachweisen werden, auch die ideale Gastheorie der van der Waalsschen, womit die spezielle Formulierung des obigen Kriteriums (nur Fall (c) von Theorem 7.4.2) gerechtfertigt ist.

Um dieses recht präzise Kriterium für die Entscheidung, ob die van der

[58] Da Entstehung und Bedeutung des van der Waalsschen Gasgesetzes eigentlich nur über seine kinetische Interpretation zu verstehen sind, kann man es nach dem Vorschlag von Kuipers (1985, S. 186) als eine *theoretische Konkretisierung* des idealen Gasgesetzes bezeichnen (vgl. Fußnote 40). Die vorhandenen, aktuellen und relevanten Daten von Regnault und Andrews zur Abweichung realer Gase vom idealen Gasgesetz dienten van der Waals lediglich als Test für seine aufgrund theoretischer Überlegungen gefundene Gleichung. Ich werde im folgenden die kinetische Grundlage nur implizit für meine Argumentation verwenden, die ganz auf der phänomenologischen Ebene bleibt.

Waalssche Theorie wirklich eine überlegene Nachfolgertheorie für die ideale Gastheorie ist oder nicht, zur Anwendung bringen zu können, müssen wir noch klären, *wie* minimale Abänderungen (Revisionen) an **VDW** vorzunehmen sind. Für die einfache, aus den vier „Axiomen" (VDW), (8.2.1), (8.2.3) und (8.2.4) bestehende Theorie **VDW** benötigen wir nicht das ganze ausgefeilte Modell der theoretischen Wichtigkeit aus Kapitel 3. Es genügt die grundlegende Idee, daß man im Zweifelsfalle immer die am schwächsten verankerten, „theoretisch unwichtigsten" Teile einer Theorie aufgeben soll.

Was ist aber der „theoretisch unwichtigste" Teil von **VDW**? Wir rufen uns in Erinnerung, daß (VDW) das Grundgesetz und sozusagen die „Identitätskarte" der van der Waalsschen Theorie ist; daß (8.2.1) die grundlegende Präsupposition aller erdenklichen Zustandsgleichungen ist; und daß (8.2.3) ein quasi analytischer Satz ist. Im Vergleich zu diesen Axiomen erscheint die Bedingung (8.2.4), so wesentlich für die van der Waalssche Theorie sie auch sein mag, doch von deutlich geringerem Gewicht. Ich gehe im folgenden also davon aus, daß (8.2.4) der theoretisch unwichtigste Teil von **VDW** ist. Unter dieser Voraussetzung können wir nun schlüssig beweisen, daß die van der Waalssche der idealen Gastheorie wirklich überlegen ist.

Von der Anwendungsbedingung A zum idealen Gasgesetz. Das ist die einfachere Richtung. A widerspricht der Konjunktion von (8.2.3) und (8.2.4), während es (VDW) und (8.2.1) offenbar unbehelligt läßt. Da (8.2.3) wichtiger als (8.2.4) ist, muß in der minimalen Abänderung von **VDW**, die zur Annahme von A nötig ist, (8.2.4) aufgegeben werden. Dies heißt aber nicht, daß die ganze in (8.2.4) enthaltene Information verloren gehen muß. Eine naheliegende, mit A kompatible Abschwächung von (8.2.4) ist

(8.2.8) Für alle Gase im weiteren Sinn, die keine Gase im engeren Sinn sind, sind die van der Waalsschen Konstanten größer als null.

Da die van der Waalsche Theorie nur minimal abgeändert werden soll, bleibt (8.2.8) anstelle von (8.2.4) erhalten. Die Revision $\mathbf{VDW}^*_A$ besteht dann aus (VDW), (8.2.1), (8.2.3), (8.2.8) und A. Aus dieser Menge ist aber die ideale Gastheorie, bestehend aus (IGG) und (8.2.1), trivial ableitbar (wobei (8.2.8) nicht benötigt wird).

Vom idealen Gasgesetz zu seiner Anwendungsbedingung A. Wir haben ständig den Verdacht geäußert, daß **IGG** mit **VDW** im Widerspruch steht, und wollen dies nun beweisen. Da wir **IGG** „in **VDW** mit aufnehmen" wollen, verfolgen wir die Strategie, **IGG** durch Streichen des schwächsten Glieds (8.2.4) in **VDW** akzeptabel zu machen. Wenn **IGG** tatsächlich mit **VDW** unverträglich ist, dann *muß* (8.2.4) gestrichen werden. Das eigent-

liche Ziel im Sinne des Nachweises, daß **VDW** eine überlegene Nachfolgertheorie von **IGG** ist, besteht jedoch in der Ableitung von A in $\mathbf{VDW}^*_{\mathbf{IGG}}$.

Gehen wir also von (VDW), (8.2.1) und (8.2.3) aus und sehen, was passiert, wenn wir (IGG) hinzunehmen. Dazu betrachten wir zunächst die Konjunktion der beiden entscheidenden Zustandsgleichungen (IGG) und (VDW). Für ein beliebiges System x eines Gases X im engeren Sinn ergibt sich aus (IGG), daß $p(x)=RT(x)/V(x)$. Wegen (8.2.3) dürfen wir dies in (VDW) einsetzen und erhalten für alle Gassysteme im engeren Sinn die Beziehung

$$(RT(x)/V(x)+(a(X)/V(x)^2))(V(x)-b(X))=RT(x).$$

(Man beachte, daß schon (VDW) die Ungleichung $V(x)>0$ präsupponiert und damit impliziert.) Ausmultiplizieren und Streichen identischer Glieder ergibt

(8.2.9) $\forall X\, \forall x\, (G_eX \wedge xSX \rightarrow (a(X)-RT(x)b(X))V(x)=a(X)b(X))$.

Sei X ein Gas im engeren Sinn. *Angenommen*, A gilt nicht, d.h. $a(X)>0$ oder $b(X)>0$. Wegen $V(x)>0$, was aus (VDW) folgt, ist nach (8.2.9) ausgeschlossen, daß $a(X)>0$ und $b(X)=0$. Also gilt $b(X)>0$. Im Falle $a(X)=0$ folgt wieder wegen $V(x)>0$ sofort $T(x)=0$ — was hieße, daß Systeme von X nur am absoluten Nullpunkt vorkommen —, und wegen (IGG) und $V(x)>0$ folgt auch $p(x)=0$, im Widerspruch zu (8.2.1) bzw. sogar zur schwächeren Bedingung (8.2.2). Sei also auch $a(X)>0$. Dann dürfen wir aus (8.2.9) und (IGG) folgendes erschließen:

(8.2.10) $\forall X\, \forall x\, (G_eX \wedge xSX \rightarrow (V(x) = a(X)b(X)/(a(X)-Rb(X)T(x)) \wedge$
$\wedge\, p(x) = RT(x)/b(X)-(RT(x))^2/a(X)))$.

Dies heißt aber, daß bei vorgegebener Temperatur eines Systems sowohl sein Volumen als auch sein Druck festgelegt ist. (Nebenbei bemerkt, ist die durch (8.2.10) bestimmte Kurve, im pV,p-Diagramm eingetragen, in der Literatur als die *Idealkurve* bekannt. Unter der Voraussetzung, daß R, V(x), a(X) und b(X) größer als null sind, zeigt die Betrachtung von (8.2.10), daß $T(x)<a(X)/Rb(X)$ gelten, d.h. daß die Temperatur von x unter der sog. *Boyle-Temperatur* für X liegen muß.) Die Festlegung von Volumen und Druck durch die alleinige Angabe der Temperatur steht aber im Widerspruch zu (8.2.1), ja sogar zu (8.2.2). Also gilt A. □

Damit haben wir, unter Zuhilfenahme der wichtigsten Elemente der van der Waalsschen Theorie, nämlich (VDW), (8.2.1) und (8.2.3), aus dem idealen Gasgesetz seine Anwendungsbedingung A hergeleitet. A widerspricht aber der Konjunktion von (8.2.3) und (8.2.4), also widerspricht auch **IGG** den vier Axiomen (VDW), (8.2.1), (8.2.3) und (8.2.4), d.h. **VDW**. Da

(8.2.4) am schlechtesten in **VDW** verankert ist, muß (8.2.4) in der minimalen Abänderung von **VDW** zur Aufnahme von (IGG) aufgegeben werden. Wir können (8.2.4) wieder durch (8.2.8) ersetzen, oder wir können die hier vielleicht noch näher liegende Abschwächung

(8.2.11) Für alle Gase im weiteren Sinn sind die van der Waalsschen Konstanten größer oder gleich null.

von (8.2.4) in $\mathbf{VDW}^*_{\mathbf{IGG}}$ beibehalten. Die Revision $\mathbf{VDW}^*_{\mathbf{IGG}}$ besteht dann aus (VDW), (8.2.1), (8.2.3), (8.2.11) und (IGG). Aus dieser Menge folgt, wie wir sahen, A (wobei (8.2.11) für die Ableitung nicht benötigt wird).

Damit haben wir gemäß Theorem 7.4.2 gezeigt, daß die van der Waalssche Theorie in der Tat eine überlegene Nachfolgertheorie für die ideale Gastheorie ist. **VDW** erklärt faktisch das Scheitern von **IGG** und kontrafaktisch **IGG** selbst (oder die „theoretische Signifikanz" von **IGG**). Es bleibt anzumerken, daß die idealisierende Anwendungsbedingung A für die ideale Gastheorie nicht gesetzesartig ist und durch die kinetische Theorie der Materie eine Interpretation findet, die genau das wiedergibt, was man intuitiv erwartet. A steht auch nicht in Konflikt mit einem fundamentalen Axiom von **VDW**, sondern nur mit der relativ unwichtigsten Bedingung (8.2.4). Anstatt auf der phänomenologischen Ebene von Binnendruck und Kovolumen zu reden, kann man im kinetischen Modell die folgende Aussage machen: Wenn die Gasmoleküle Massenpunkte (d.h. nicht räumlich ausgedehnt und nicht wechselwirkend[59]) wären, dann würde das ideale Gasgesetz gelten; weil sie aber keine Massenpunkte sind, deshalb gilt das ideale Gasgesetz nicht.

8.2.4 KGP — NTG und IGG — VDW: Vergleich zweier intertheoretischer Idealisierungen

Wir haben die Ideen der intertheoretischen Approximation und der intertheoretischen Idealisierung an zwei einfachen Fallbeispielen erprobt: am Verhältnis zwischen den Keplerschen Gesetzen und der Newtonschen Gravitationstheorie und am Verhältnis zwischen dem idealen und dem van der Waalsschen Gasgesetz. Obwohl sich herausgestellt hat, daß sich in beiden Fällen die in Kapitel 7 entwickelten Begriffe zur Anwendung bringen las-

[59] Um jede Art von Wechselwirkung auszuschalten, müßten wir genaugenommen hinzufügen, daß die Moleküle als Punkte mit träger, aber nicht schwerer (gravitierender) Masse zu betrachten sind. Ansonsten wären ja (winzig kleine) Gravitationskräfte anzusetzen.

sen, darf man nicht vorschnell schließen, daß die Beispiele völlig parallel laufen. Denn es lassen sich neben einigen Gemeinsamkeiten eine Reihe von Unterschieden festhalten, die wir nun beleuchten wollen.

Sowohl das Paar **KGP — NTG** als auch das Paar **IGG — VDW** stellt intuitiv ein Beispiel dafür dar, daß Nachfolgertheorien häufig in irgendeinem Sinne mit ihren Vorgängern inkonsistent sind. Aber weder (KGP) und (NTG) noch (IGG) und (VDW) sind für sich genommen miteinander inkonsistente Gleichungen. Karussellmodelle (im Sinne von Scheibe 1973a) sind Lösungen von (KGP) und (NTG), die Punkte auf der Idealkurve sind Lösungen von (IGG) und (VDW). Um den intuitiv erwarteten Widerspruch zu erhalten, muß man Theorien betrachten, die als umfassende Theorien über die Welt neben den eigentlichen Gesetzen noch „Hintergrundinformation" enthalten, die in der Regel nicht als physikalisches Gesetzeswissen anzusehen ist, sondern insbesondere sprachliches und mathematisches Wissen und Wissen über kontingente Rand- oder Anfangsbedingungen enthält. Im Kepler-Newton-Beispiel soll hierzu etwa das Wissen gehören, daß es in unserem Sonnensystem mehrere Planeten gibt, die sich nicht karussellförmig um die Sonne bewegen (und die eine positive Masse haben). Im Beispiel des idealen und des van der Waalsschen Gasgesetzes haben wir die relevanten Bedingungen mit (8.2.1)–(8.2.4) explizit angegeben.

Bei den Approximationsprozessen beider Beispiele kann man die Dualität von Bereichseinschränkung und Kontrafaktizität erkennen. Im ersten Fall ist dies davon abhängig, wie man die Keplerschen Gesetze auffaßt. Interpretiert man sie als Aussagen über alle möglichen Planetensysteme, so wird durch $m_i \to 0$ $(i \geq 2)$ der Anwendungsbereich dieses Gesetzes mehr und mehr eingeschränkt; interpretiert man sie — wie ich es getan habe — als Gesetze, die nur in unserem singulären Sonnensystem Gültigkeit beanspruchen, so bedeutet der Grenzübergang $m_i \to 0$ eine zunehmende Entfernung von der Wahrheit. Im zweiten Fall ist das ideale Gasgesetz ein Gesetz, welches eindeutig als Gesetz für alle Gase (im engeren Sinn) zu interpretieren ist. Deshalb entsprechen den Alternativen Bereichseinschränkung und Kontrafaktizität hier zwei verschiedene Typen von Grenzübergang: $V \to \infty$ (und/oder $p \to 0$ und/oder $T \to \infty$ o.ä.) schränkt den Bereich des idealen Gasgesetzes immer mehr ein, $a \to 0$ und $b \to 0$ stellen dagegen für reale Gase zunehmend kontrafaktische Annahmen dar. Ich habe in beiden Fällen der kontrafaktischen Idee den Vorzug gegeben und argumentiert, daß man mit dieser Idee auf den Approximationsprozeß ganz verzichten kann.

Die Idealisierungsperspektive erwies sich dann im Kepler-Newton-Fall als wesentlich interessanter als im Fall des idealen und des van der Waals-

schen Gasgesetzes. Während hier die dem Approximationsprozeß $a \to 0$ und $b \to 0$ analogen Annahmen $a=0$ und $b=0$ *nicht* im Konflikt mit der Modellvorstellung der kinetischen Gastheorie stehen,[60] entsteht dort durch die Annahme $m_i=0$ ein Konflikt mit der Newtonschen Gravitationstheorie, ja sogar schon mit der Newtonschen Mechanik, der unauflösbar erscheint. Vom Standpunkt der van der Waalsschen Gastheorie ist es zwanglos möglich, die durch a und b repräsentierten Faktoren Binnendruck und Kovolumen gleich null zu setzen. Vom Standpunkt der Newtonschen Gravitationstheorie jedoch kann man den Faktor Masse schwerlich gleich null setzen — **NTG** ist ja eine Theorie über genau diesen Faktor! Durch eine kontrafaktische Einschränkung der Menge der beteiligten Objekte, nämlich durch Betrachtung von Ein- und Zweikörpersystemen, gelingt es aber ausgezeichnet, die theoretische Signifikanz der Keplerschen Gesetze aufzuweisen. Die Idealisierung im Kepler-Newton-Fall besteht also in einer in der Physik und allen anderen Wissenschaften wohlbekannten Idee: der praktisch undurchführbaren, nur gedanklich möglichen Isolierung eines kleinen, übersichtlichen Systems von anderen wirkenden Objekten. Ein- und Zweikörpersysteme sind auch und vor allem „gravitationell" abgeschlossene Systeme. Wir können diesen Typ von Idealisierung die *Idealisierung des abgeschlossenen Systems* nennen.

Ganz im Gegensatz dazu müssen die Systeme, die Gegenstand des idealen und des van der Waalsschen Gasgesetzes sind, nicht thermisch isoliert sein. Für Zustandsänderungen thermisch abgeschlossener Systeme, sogenannte adiabatische Zustandsänderungen, gelten die Adiabatengleichungen, die (bei nicht konstanter Temperatur!) pV^γ=konstant bzw. $(p+a/V^2)(V-b)^\gamma$=konstant (vgl. Partington 1949, S. 680) lauten, wobei $\gamma=c_p/c_V$ der Quotient der spezifischen Wärmekapazitäten bei konstantem Druck und konstantem Volumen ist. Das ideale Gasgesetz beschreibt aber auch insbesondere isotherme (Boylesches Gesetz), isochore (Gay-Lussacsches Gesetz) und isobare (Charlessches Gesetz) Prozesse, die allesamt einen perfekten Wärmeaustausch mit der „unendlich groß" angenommenen Umgebung voraussetzen (und deshalb eigentlich „unendlich langsam" ablaufen müßten). Im Fall des idealen und des van der Waalsschen Gasgesetzes liegt also keine Idealisierung des abgeschlossenen Systems vor, sind „Faktoren" nicht mit Objekten gleichzusetzen. Für **KGP** bleiben gegenüber **NTG** die relevanten Eigenschaften der Objekte gleich (Massen, Ort und Geschwindigkeiten), die Anzahl der relevanten Objekte wird aber

[60] Dagegen würden die den bereichseinschränkenden Approximationsprozessen entsprechenden Annahmen $V(x)=\infty$ oder $T(x)=\infty$ die Gesetze (VDW) und (IGG) offensichtlich wertlos, weil unanwendbar, machen; $p(x)=0$ ist nur für (IGG), nicht für (VDW) unsinnig.

manipuliert (nur 1 oder 2 statt 7–10); für **IGG** bleibt gegenüber **VDW** die Anzahl der relevanten Objekte gleich ($6{,}022 \cdot 10^{23}$ pro Mol eines Gases), die Menge der relevanten Eigenschaften dieser Objekte wird aber manipuliert (nur träge Massen, Orte und Geschwindigkeiten, keine räumliche Ausdehnung und gegenseitigen Anziehungskräfte).

Nachzutragen bleibt noch ein wichtiger, vielleicht der augenfälligste Unterschied zwischen den beiden Beispielen. Obgleich man beide Male von einer theoretischen Konkretisierung der alten Theorie durch die neue Theorie sprechen kann, haben die betrachtetetn Beispielspaare eine verschiedene systematische Stellung. Dies sei durch die folgenden Skizzen angedeutet:

		KGP* mit	*kinetische*
KGP	**KGP***	„Störungen“	*Ebene*
↑	↑	↑	
$\mathbf{NTG^*_{A_1}}$	← $\mathbf{NTG^*_{A_2}}$ ←	$\mathbf{NTG^+_{A_0}}$(=**NTG**?)	*dynamische Ebene*

		VDW mit	
IGG	$\overset{*A}{\leftarrow}$ **VDW**	„Maxwell-Konstruktion“	*Makro-Ebene*
↑	↑	↑	
$\mathbf{KGT^*_{A_1}}$	← $\mathbf{KGT^*_{A_2}}$ ←	$\mathbf{KGT^+_{A_0}}$(=**KGT**?)	*Mikro-Ebene*

SKIZZE 8.1

(**KGT** heißt kinetische Gastheorie, A_1 und A_2 sind mehr bzw. weniger idealisierende Anwendungsbedingungen, A_0 steht für die tatsächlich, in der Realität vorliegenden Randbedingungen und kann eventuell als ganz in **NTG** bzw. **KGT** enthalten betrachtet werden.) Die obere Zeile stellt jeweils Gesetze oder Theorien dar, welche eher empirischen, phänomenologischen Charakter haben, die Gesetze oder Theorien in der unteren Zeile sind relativ weit entfernt von der sinnlichen Wahrnehmung und mehr „theoriegeladen“.

Aufgrund der Stellung von **KGP** und **NTG** bzw. von **IGG** und **VDW** in diesen Schemata kann man im ersten Fall eine *vertikale Idealisierung* und im zweiten Fall eine *horizontale Idealisierung* konstatieren. Während man wohl in beiden Fällen von einer *Reduktion* sprechen darf, wird man nur im ersten Fall mit gutem Gewissen sagen können, daß eine echte *Erklärung* vorliegt. Im zweiten Fall wäre statt „**VDW** erklärt (das Scheitern von) **IGG**“ eine genauere Ausdrucksweise wohl „die Interpretation von **VDW**

durch die kinetische Gastheorie erklärt (das Scheitern von) **IGG**." Der Aspekt, daß die Explanans-Theorie irgendwie „tiefer gehen" muß als die Explanandum-Theorie, wird in dem einfachen, in Kapitel 7 entwickelten Modell der intertheoretischen Erklärung nicht erfaßt.

Der letzte Punkt, der eine Gegenüberstellung lohnt, ist die Betrachtung der qualitativen Unterschiede, die der Übergang von **KGP** zu **NTG** bzw. von **IGG** zu **VDW** mit sich bringt. Dieser Punkt ist von so großem Interesse, daß wir ihm einen eigenen Abschnitt widmen wollen.

8.2.5 Quantitative Ähnlichkeiten und qualitative Unterschiede

Das Approximationsmodell liefert Darstellungen der folgenden Art: Die Newtonsche Gravitationstheorie erklärt (approximativ) die Keplerschen Gesetze, das van der Waalssche Gasgesetz (bzw. seine kinetische Deutung) erklärt (approximativ) das ideale Gasgesetz usw. Dadurch wird nahegelegt, daß beim wissenschaftlichen Theorienwandel vor allem auf Ähnlichkeiten zwischen Vorgänger- und Nachfolgertheorie zu achten ist. Nach dem Idealisierungsmodell von Kapitel 7 erklärt die Newtonsche Gravitationstheorie *nicht wirklich* die Keplerschen Gesetze, sondern deren Scheitern, und das van der Waalssche Gasgesetz erklärt, kinetisch interpretiert, *nicht wirklich* das ideale Gasgesetz, sondern dessen Scheitern. Wenn man sich diese Feststellung ganz deutlich vor Augen hält, dann wird man erwarten, daß die Abweichungen zwischen **NTG** und **KGP** bzw. zwischen **VDW** und **IGG** wichtiger sind als ihre Übereinstimmungen. Außerdem kann der Idealisierungsbegriff — im Unterschied zum per definitionem quantitativen Approximationsbegriff — qualitativ modelliert werden, wie dies auch im vorigen und in diesem Kapitel geschehen ist. Wir können daher vermuten, daß *qualitative Unterschiede* als Triebkräfte beim Theorienwandel vielleicht eine größere Rolle spielen als *quantitative Annäherungen*. Unter diesem Gesichtspunkt stellt sich das Beispiel des idealen und des van der Waalsschen Gasgesetzes als interessanter heraus als das Kepler-Newton-Beispiel.

Welche wichtigen qualitativen Unterschiede für das Sonnensystem brachte die Newtonsche Gravitationstheorie gegenüber den Keplerschen Gesetzen mit sich? Zwei Dinge scheinen mir hier besonders nennenswert zu sein. Erstens ist es eine Folgerung aus der Newtonschen Theorie, daß sich die Sonne bewegt, während Keplers Gesetze vom Kopernikanischen Bild einer ruhenden Sonne auszugehen scheinen (vgl. aber Fußnote 29); zweitens wird es mit dem Auftauchen der Gravitationstheorie zu einer offenen,

sehr schwer zu beantwortenden Frage, ob das Sonnensystem stabil ist oder nicht (vgl. aber Fußnote 16), während das nach den Keplerschen Ellipsenbahnen eine Selbstverständlichkeit ist. Die beiden Punkte sind von großem philosophischen und physikalischem Interesse, haben aber den Nachteil, empirisch so schwer nachprüfbar zu sein, daß sie keinen Einfluß auf den historischen Wechsel von **KGP** zu **NTG** haben konnten: Bewegungen von Himmelskörpern können nur als relative Bewegungen beobachtet werden (zudem „wackelt" die Sonne nur sehr wenig), und das Zehnkörpersystem Sonne-mit-Planeten trotzt offenbar bis heute einer endgültigen Berechnung (vgl. Moser 1977–79; zudem gibt es Einflüsse von Monden und Sternen).

Ganz anders ist die Sachlage im Fall des idealen und des van der Waalsschen Gasgesetzes. Es gibt hier eine Fülle von bereits Mitte des 19. Jahrhunderts bekannten Phänomenen und Effekten, die nach dem idealen Gasgesetz gar nicht auftreten dürften, die aber — zumindest qualitativ richtig — vom van der Waalsschen Gasgesetz vorausgesagt (oder „hinterhergesagt") wurden.

Die wichtigste durch das van der Waalssche Gasgesetz eingeführte Neuerung bestand in der (im wesentlichen qualitativ korrekten) Beschreibung der Zustandsänderung von Gasen in der Nähe der Verflüssigung (von „Dämpfen"), während der Verflüssigung (der Koexistenz von gasförmiger und flüssiger Phase) und von Flüssigkeiten. Wie die van der Waalssche Gleichung aussagt, ist eine Verflüssigung unabhängig vom ausgeübten Druck nur unterhalb einer bestimmten, der *kritischen* Temperatur T_k möglich, und diese Temperatur liegt über dem absoluten Nullpunkt, d.h. es lassen sich alle, auch die vormals „permanent" genannten Gase verflüssigen. Weiter bringt nach dem van der Waalsschen Gasgesetz eine bei konstanter Temperatur $T<T_k$ durchgeführte Volumenverkleinerung nicht unbedingt eine Druckerhöhung mit sich — ein Phänomen, das sich während der Verflüssigung von Gasen beobachten läßt. Es gibt sogar, wie nach van der Waals auszurechnen, für ein und dasselbe System zwei Zustände $\langle T,p,V\rangle$ und $\langle T,p',V'\rangle$ mit $p<p'$ und $V<V'$, wobei $\langle T,p,V\rangle$ einen Zustand der flüssigen Phase und $\langle T,p',V'\rangle$ einen Zustand der gasförmigen Phase charakerisiert und sich erstere im Siedeverzug und/oder letztere im Kondensationsverzug befindet. (Solche Zustände sind allerdings instabil.) Die van der Waalssche Zustandsgleichung zeigt, daß Gase nicht beliebig komprimierbar sind (das Molvolumen muß stets größer als b(X) sein) und daß Flüssigkeiten (Gase im weiteren Sinn) bei negativen Drücken existieren können — was beides empirisch nachgewiesen ist. Sie macht verständlich, warum reale Gase unterhalb einer bestimmten Temperatur (der Boyle-Temperatur $T_B=27/8\cdot T_k$)

bei kleinen Drücken mehr, bei größeren Drücken aber weniger kompressibel sind als ein (stets in gleichem Maße kompressibles) ideales Gas und warum über der Boyle-Temperatur die Kompressibilität stets kleiner als die eines idealen Gases sein muß.[61] (Dies ist sehr schön zu sehen in den sogenannten Amagat-Diagrammen.) Schließlich kann mit dem van der Waalsschen Gasgesetz die Temperatur, bei welcher sich der Joule-Kelvin-Effekt (der bei idealen Gasen gar nicht auftreten kann) umkehrt, die sog. Inversionstemperatur, als $T_{JK}=27/4\cdot T_k$ ermittelt werden. Auch in diesen Fällen sind die aus dem van der Waalsschen Gasgesetz gewonnenen Werte nicht unbedingt in quantitativer, wohl aber in qualitativer Hinsicht überzeugend.[62]

Alle diese wichtigen und empirisch bestätigten Phänomene können auf der Grundlage des van der Waalsschen Gesetzes erklärt oder vorhergesagt werden, während das ideale Gasgesetz durchwegs falsche Aussagen dazu macht. Diese qualitativen Unterschiede zum idealen Gasgesetz sind es, die einen entscheidenden, wenn nicht *den* entscheidenden Anteil an der Überlegenheit des van der Waalsschen Gasgesetzes ausmachen. Dagegen erscheint die Tatsache, daß das ideale Gasgesetz in bestimmten, durchaus weiten Bereichen quantitativ angenähert wird, weniger bedeutsam.

Eine Art qualitativen Unterschieds des van der Waalsschen zum idealen Gasgesetz ist philosophisch besonders interessant. Es ist nämlich nicht nur so, daß das van der Waalssche Gasgesetz für Gase und Flüssigkeiten Gültigkeit beanspruchen kann und damit zwei getrennte, wohlunterschiedene Anwendungsbereiche abdeckt. In einem gewissen, physikalisch präzisierbaren Sinn führt das van der Waalssche Gasgesetz sogar „zu einer Aufhebung des Unterschiedes zwischen flüssigem und gasförmigem Zustande" (Schaefer 1958, S. 119). Während man deutliche Beispiele von gasförmigem und deutliche Beispiele von flüssigem Zustand eines Stoffes sehr gut auseinanderhalten kann, ist eine prinzipielle Grenze nicht mehr zu ziehen, denn, wie die van der Waalssche Gleichung zeigt, ist der eine in den anderen Zustand ohne jede Diskontinuität überführbar. Sehr schön ist dies ausgeführt in Boltzmanns (1898, §19) Abschnitt über die Willkürlichkeit der Definitionen von Gas, Dampf und (tropfbarer) Flüssigkeit — Begriffe, die höchstens noch relativ zu einer bestimmten Art der Zustandsänderung (isotherm, iso-

[61] Nur Wasserstoff und die Edelgase sind bei gewöhnlichen Temperaturen weniger kompressibel als ein ideales Gas, weshalb Regnault 1847 Wasserstoff als ein „gas plus que parfait" bezeichnet hatte.

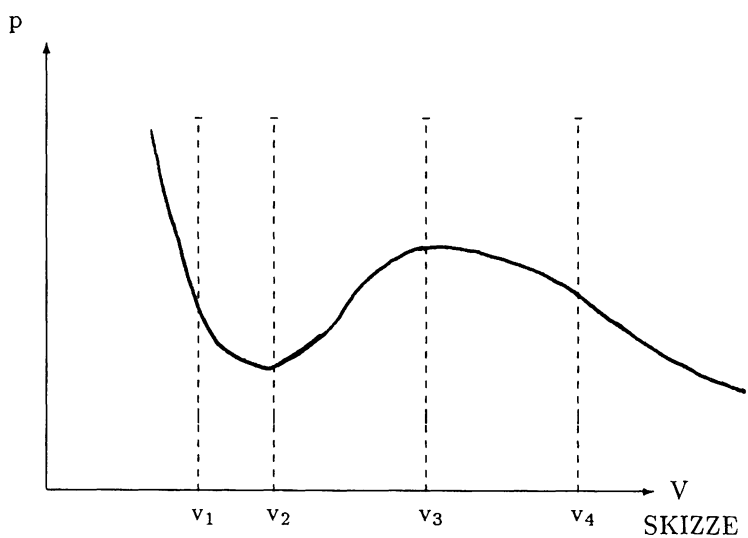
[62] Fast alle einschlägigen Lehrbücher diskutieren fast jeden der in diesem Absatz angesprochenen Punkte. Besonders empfehlenswert sind wohl Boltzmann (1898, S. 1–51), Partington (1949, S. 546–697), Sommerfeld (1952, S. 1-23, 52–68, 92–96) und Elsner (1980, S. 8–38, 154–173, 263–270).

bar, adiabatisch usw.) Sinn machen. Das van der Waalssche Gasgesetz hat Konsequenzen, die prima facie paradox klingen:

> Die Substanz kann sogar aus einem ausgesprochen tropfbar flüssigen in einen ausgesprochen dampfförmigen Zustand nicht durch Verdampfung, sondern durch Kondensation übergeführt werden. ... Ebenso kann man einen ausgesprochenen Dampfzustand durch Verdampfung in einen offenbar tropfbar flüssigen überführen. (Boltzmann 1898, S. 49)

Wir haben hier also den interessanten Tatbestand vorliegen, daß ein neues *Gesetz* alte *Begriffe* — die in vielen Fällen nützlich sind und bedenkenlos angewandt werden dürfen — als theoretisch fragwürdig erweist. Es ist unnötig hinzuzufügen, daß das ideale Gasgesetz zu diesem Thema nichts beitragen kann.

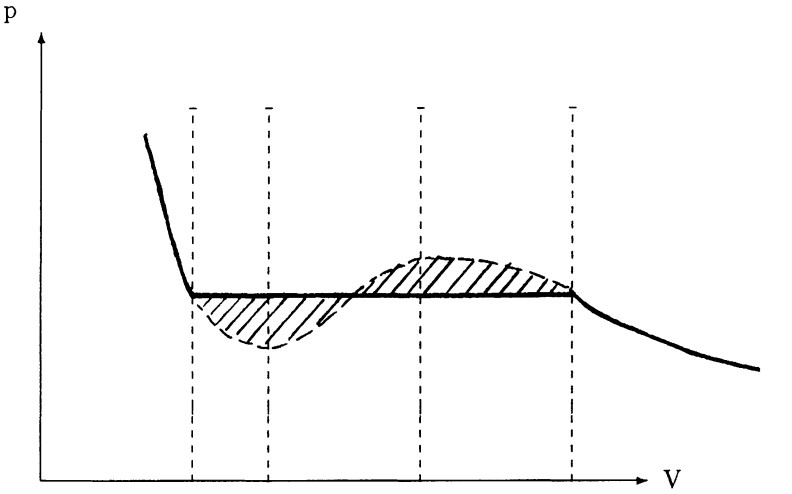
Wir kehren zur Rolle von qualitativen Unterschieden in der Theoriendynamik zurück und betrachten als letzten Punkt die Frage, wie die Entwicklung der Gastheorie nach van der Waals „programmgemäß" weitergehen mußte. Das van der Waalssche Gasgesetz liefert für $T<T_k$ Isothermen der folgenden Form:



SKIZZE 8.2

Die für manche Gase empirisch schon früher gefundenen und von Maxwell

1875 thermodynamisch begründeten Isothermen haben aber die folgende Gestalt:



SKIZZE 8.3

Der van der Waalssche Kurvenverlauf zwischen v_1 und v_2 und zwischen v_3 und v_4 kann durch die instabilen Zustände beim Siede- bzw. Kondensationsverzug prinzipiell nachvollzogen werden. Doch das Stück zwischen v_2 und v_3, wo die Isotherme eine positive Steigung hat, wird gemeinhin als „unphysikalisch" oder „physikalisch sinnlos" zurückgewiesen.[63] Es dürfte nicht übertrieben sein, zu sagen, daß dieser qualitative Mangel der van der Waalsschen Zustandsgleichung viel dringlicher eine Behebung verlangt als alle bloß quantitativen Unzulänglichkeiten. Die Herausforderung an die kinetische Gastheorie war damit schon 1875 ausgesprochen: Unter welchen Bedingungen ist das van der Waalssche Gasgesetz *mit der Maxwellschen Konstruktion* genau gültig? Es dauerte ein knappes Jahrhundert, bis man die mathematischen Schwierigkeiten überwunden hatte und Kac, Uhlenbeck und Hemmer 1963/4, van Kampen 1964 und Lebowitz und Penrose 1966 die Beantwortung dieser alten Frage erfolgreich in Angriff nehmen konnten. Wenn deren Berechnungen auch weit über den Horizont unserer Betrachtungen hinausgehen, so dürfen wir vielleicht doch die Hoffnung äußern, daß die kinetische Gastheorie „inklusive" der Anwendungsbedin-

[63] Van der Waals selbst sah dies anders; vgl. Klein (1974, S. 41f).

gungen, die für die Maxwellsche Modifikation des van der Waalsschen Gasgesetzes notwendig sind, das ursprüngliche van der Waalssche Gesetz als Idealisierung bezüglich ganz bestimmter Faktoren erweist — insofern als die Maxwellsche Modifikation zu einer überlegenen Nachfolgertheorie führt.

Die Betrachtung des idealen und des van der Waalschen Gasgesetzes aus der Idealisierungsperspektive motiviert also ein interessantes wissenschaftshistorisches und -theoretisches Forschungsprogramm. Es wäre zu untersuchen, inwieweit die Vermutung zutrifft, daß auch in anderen, womöglich bedeutsameren Fällen von Theorienwandel qualitativ verschiedene Voraussagen eine wichtigere Rolle spielen als quantitative „Korrespondenzen".

Kapitel 9

Über Idealisierung in der Wissenschaft

9.1 Idealisierungen in der Wissenschaftstheorie

Scriven (1962) hat in seiner vielbeachteten Arbeit ‚The Key Property of Physical Laws — Inaccuracy' die provokante These aufgestellt, die „Natur physikalischer Gesetze" bestehe darin, daß sie inakkurat seien. Beim Versuch, diese Behauptung zu verstehen, muß man das Fremdwort „inakkurat" entschlüsseln. Laut Fremdwörterbuch von *Duden* bedeutet es „ungenau, unsorgfältig", wovon auf Gesetze nur „ungenau" paßt. Aufschlußreicher ist es noch, wenn wir in englischen Wörterbüchern nach der Bedeutung des englischen Wortes „inaccurate" nachforschen. In *Webster's Third New International Dictionary* etwa findet man dort den Eintrag (a) „containing a mistake or error: INCORRECT, ERRONEOUS" und (b) „not functioning with accuracy or precision: FAULTY, DEFECTIVE". Unter „accuracy" wiederum findet man eine Aufspaltung in (a) „...: CORRECTNESS" und (b) „...: EXACTNESS". „Accurate" heißt „CORRECT [vgl. (a)], EXACT [vgl. (b)], PRECISE [vgl. (b)]". Rubrik (a) zielt offenbar auf eine richtig-falsch-Dichotomie, während Rubrik (b) so etwas wie Grade der Genauigkeit zu meinen und auf die Unterscheidung von ganz richtig und nicht ganz richtig hinauszulaufen scheint. Maximale Flexibilität erhalten wir uns, wenn wir drei Fälle unterscheiden: ganz richtig, nicht ganz richtig und überhaupt

nicht richtig.

Nun ist es einfach nicht möglich, daß physikalische Gesetze und auch Gesetze anderer mehr oder weniger erfolgreicher Wissenschaften überhaupt nicht, in keinem erdenklichen Sinne, richtig sind, denn sonst wären diese Wissenschaften doch nicht erfolgreich! Scriven kann also nur gemeint haben, daß wissenschaftliche Gesetze nicht ganz genau richtig sind. Immer noch kann man das verschieden betonen, nämlich so: daß diese Gesetze nicht *ganz genau* (aber doch *ungefähr*) richtig sind, daß sie also nur *strenggenommen* falsch sind (dies entspricht Rubrik (b) oben). Oder aber so: daß diese Gesetze *nicht* ganz genau *richtig sind* (aber unter Umständen *richtig wären*), daß sie vorderhand *in der Tat falsch* sind (dies entspricht Rubrik (a) oben).

Prima facie scheint die erstgenannte Emphase für den Wissenschaftsphilosophen auf jeden Fall die informativere und interessantere zu sein, zumal ja schon bestens bewährte mathematische Methoden zur Behandlung von Approximationen zur Verfügung stehen. Ich möchte hier aber zu bedenken geben, daß die Präferenzen nicht vorschnell verteilt werden sollten. Wenn man sagt, irgend etwas sei ungefähr richtig, dann gibt man sich meist schon mit dieser Feststellung zufrieden. Wissenschaftstheoretiker sind da keine Ausnahme. Es besteht die Gefahr, daß man den Einzelfall aus den Augen verliert und sich darauf beschränkt, den Approximationsbegriff durch Aufbereitung und Erläuterung einiger Fragmente aus Lehrbüchern der Mathematik (bevorzugt der Topologie) zu explizieren. Wenn man aber deutlich macht, daß etwas nicht richtig ist, etwas, das — wie physikalische Gesetze — vielleicht gute Dienste geleistet hat, dann bohrt man nach und fragt, *warum* es nicht richtig ist. Findet man eine Antwort auf diese Frage, war die zweitgenannte Betonung vielleicht doch die bessere.

Nachdem wir im letzten Kapitel anhand von zwei relativ detailliert analysierten Fallbeispielen ein Bild von Idealisierungen — insbesondere in Abgrenzung gegenüber Approximationen — gewonnen haben, wollen wir zum Abschluß dieses Buchs noch einige abstrakt-philosophische Überlegungen zum Begriff der Idealisierung anstellen. Dabei werden wir auf eine Anzahl von Vorschlägen aus der Literatur eingehen, auf die Diskussion weiterer Beispiele wird hingegen verzichtet — hier konsultiere man die zitierten Autoren. Dem Ausflug in spekulativere Gefilde seien noch einige erläuternde Bemerkungen vorausgeschickt.

Wie im Verlaufe der letzten Kapitel deutlich geworden ist, halte ich die These, daß wissenschaftliche Gesetze (mit der unten genannten Einschränkung) praktisch *immer* Idealisierungen sind, für eine zwar gewagte,

aber informative und systematischer Bearbeitung würdige Antwort auf die Frage, warum sie praktisch *nie* ganz richtig sind. Die These besagt, daß wissenschaftliche Gesetze vereinfachen, von sogenannten „Störfaktoren" abstrahieren oder „zu vernachlässigende" Größen außen vor lassen. Oft ist dies dem praktizierenden Wissenschaftler bewußt; dann ist es ein natürliches Forschungsprogramm, die Idealisierungen Schritt für Schritt aufzuheben und die vernachlässigten Faktoren in Rechnung zu stellen, in der Hoffnung, daß am Ende die — dann modifizierten — Gesetze ganz genau richtig sind, oder wenigstens genauer als die in stärkerem Maße idealisierenden Gesetze. Vielleicht das beste Beispiel eines solchen Forschungsprogramms stellt das Bohrsche Atommodell dar.[1]

Manchmal ist es nicht bekannt, daß ein Gesetz idealisiert; dann stellt die Entdeckung neuer Zusammenhänge, in deren Licht die Idealisierung offenbar wird, einen echten, eventuell das Prädikat „revolutionär" verdienenden wissenschaftlichen Fortschritt dar. Ein revolutionäres Beispiel ist wohl der Kepler-Newton-Fall, ein eher nichtrevolutionäres der Übergang vom idealen zum van der Waalsschen Gasgesetz.

Auf eine Einschränkung der Reichweite der Idealisierungsanalyse möchte ich, um eine dauernde Wiederholung im Verlaufe dieses Kapitels zu vermeiden, jetzt schon hinweisen: Ich halte es für ratsam, die folgenden Behauptungen in aller Regel nur auf Gesetze zu beziehen, die nicht *ganz* fundamental sind (wie etwa die Newtonschen Gesetze der Bewegung),[2] und nur auf Theorienwandel, der nicht *ganz* radikal ist (wie etwa der Wechsel von der Phlogistontheorie zur modernen Chemie).

Intertheoretische Idealisierungen wurden bereits mehrfach charakterisiert, dagegen wurde noch nicht gesagt, was es heißt, daß ein Gesetz oder eine Theorie für sich allein genommen eine Idealisierung ist. Auch hier spielen kontrafaktische Konditionalsätze die entscheidende Rolle. Ein Gesetz G soll genau dann ein *idealisierendes* (oder *idealisiertes*) *Gesetz* heißen, wenn gilt: Wenn die-und-die Faktoren nicht vorhanden oder wenn die-und-die Umstände gegeben wären, dann wäre G (ganz genau, strenggenommen) wahr. Wie auch immer man den Idealisierungsbegriff fassen will, ich glaube, daß keine befriedigende Analyse von Idealisierungen ohne einen Rückgriff

[1] Vgl. Lakatos (1970, S. 146–153), Shapere (1974, S. 558–565) und McMullin (1985, S. 259–261, 263f). Zum Programmaspekt von **NTG** vgl. Lakatos (1970, S. 135f) und von **VDW** vgl. Clark (1976, S. 57–60).

[2] Gerade umgekehrt sieht das Nancy Cartwright (1983, S. 3): „Realists are inclined to believe that if theoretical laws are false and inaccurate, then phenomenological laws are more so. I urge just the reverse. When it comes to the test, fundamental laws are far worse off than the phenomenological laws they are supposed to explain."

auf kontrafaktische Konditionale und ohne eine befriedigende Analyse derselben auskommt.[3] Des öfteren übersehen Wissenschaftstheoretiker, die sich auf dem richtigen Weg befinden, Probleme, die in der konditionallogischen Diskussion als Standardprobleme thematisiert und gelöst sind.[4]

Zwischen der Approximations- und der Idealisierungsperspektive besteht offensichtlich ein großer Unterschied, welcher aber in den allermeisten einschlägigen Arbeiten nicht gesehen oder verwischt wird. Eine positive Ausnahme macht schon Scriven, noch etwas expliziter sind Shapere und Suppe, am deutlichsten erkennt Laymon die Trennungslinie. Als Negativbeispiel kann Schwartz (1978) dienen, der die zum Teil treffenden Beobachtungen von Barr (1974) mit reiner Approximationspropaganda zu leugnen versucht. Idealisierungen sind ohnehin kein sehr häufig bearbeitetes Thema in der an der Physik orientierten Wissenschaftstheorie. Dies kann man dadurch erklären, daß die meisten physikalischen Gesetze — im Gegensatz zu sozialwissenschaftlichen Gesetzen — eben ungefähr richtig sind und so die Frage nach dem Warum ihrer Unrichtigkeit in den Hintergrund

[3] Hinsichtlich der Wichtigkeit von kontrafaktischen Konditionalsätzen in der Wissenschaftstheorie befinde ich mich in bester Übereinstimmung mit Goodman (1947/65, S. 3): „Indeed, if we lack the means for interpreting counterfactual conditionals, we can hardly claim to have any adequate philosophy of science." Eine extreme Gegenposition vertritt van Fraassen (1980, S. 118): „... science does not imply the truth of any counterfactual — except in the limiting case of a counterfactual with the same truth-value in all contexts."

[4] So sprechen z.B. Nowak (1972, S. 539) und Yoshida (1977, S. 6) wiederholt von *dem* Teil einer Theorie, der mit einer bestimmten Annahme konsistent ist, obwohl es für Konditionallogiker ein Gemeinplatz ist, daß hier keine Eindeutigkeit vorliegt (tatsächlich ist es, wie in Kapitel 3 gesehen, ein zentrales Problem, wie man zu eindeutigen Kontraktionen kommt). Um einen Eindruck von der Nützlichkeit des Gärdenfors-Modells zu vermitteln, seien zwei Punkte aus der Idealisierungsliteratur herausgegriffen. Nowak (1972, S. 538–541) konstruiert die Revision einer Theorie T bezüglich zweier Annahmen A und B (mit den Symbolen von Kapitel 8.2: es geht um a(X)=0 und b(X)=0), und er schlägt dabei offenbar die folgende Beziehung vor:

$$T^*_{A\wedge B} = (T^-_{\neg(A\wedge B)})^+_{A\wedge B} = (T^-_{\neg A\vee\neg B})^+_{A\wedge B} = (T^-_{\neg A}\cap T^-_{\neg B})^+_{A\wedge B}.$$

Der letzte Schritt ist jedoch fehlerhaft, da — wie (T-7∧8) aus Kapitel 3 zeigt — $T^-_{\neg A}\cap T^-_{\neg B}$ allenfalls mit $T^-_{\neg A\wedge\neg B}$, nicht aber mit $T^-_{\neg A\vee\neg B}$ identisch ist. — Brzeziński e.a. (1976, S. 76) geben im Schema der „dialektischen Korrespondenz" eine Theorie T an, die die Konditionalsätze Wenn W und A, dann Z', Wenn W und A und ¬B, dann Z'' und Wenn W und A und B, dann Z''' enthält. Eine interessante, aber unbeantwortete Frage ist die nach der Vergleichbarkeit dieser Sätze. Bei einer Ramsey-Test-Interpretation erhalten wir $Z'\in T^*_{W\wedge A}$, $Z''\in T^*_{W\wedge A\wedge\neg B}$ und $Z'''\in T^*_{W\wedge A\wedge B}$. Unter Berücksichtigung von $T^*_{W\wedge A}=T^*_{(W\wedge A\wedge B)\vee(W\wedge A\wedge\neg B)}$ und von (T*7∧8) und bei erneuter Ramsey-Test-Interpretation folgt daraus, daß auch Wenn W und A, dann $Z''\vee Z'''$ und entweder Wenn W und A und ¬B, dann Z' oder Wenn W und A und B, dann Z' in T sind.

gedrängt wird.[5] Eine Ausnahme ist die vor allem von der Poznań-Schule vertretene polnische Wissenschaftstheorie (Nowakowa, Nowak, Krajewski, Patryas, Such und andere), die in den 70er Jahren Idealisierungen in das Zentrum ihrer Aufmerksamkeit rückte.[6] Im Suppes-Sneed-Stegmüllerschen Strukturalismus, der wissenschaftliche Gesetze mit Modellmengen identifiziert und den Begriff der Gesetzesartigkeit überhaupt nicht verwendet, gibt es nur appoximative Ansätze, deren inhaltlicher Kern bei der Analyse des paradigmatischen Kepler-Newton-Beispiels in Kapitel 8.1 skizziert wurde. Stegmüller (1986, S. 228) spricht Idealisierung nur als eine von vier Unterarten der Approximation an, zudem eine, die nicht weiter thematisiert wird.[7] Neuerdings ist aber auch hier ein Widerhall der Lewisschen Semantik für kontrafaktische Konditionale zu registrieren (Balzer 1985, S. 101–113; vgl. auch Niiniluoto 1986, S. 281f). Insgesamt wäre es nicht allzu aufwendig, eine Literaturliste zum Thema „Idealisierung in der Wissenschaft" zu erstellen, die Anspruch auf Vollständigkeit erheben kann.

Die Betonung von Idealisierungen, so wie ich sie verstehe, bringt die Diskussion von wissenschaftlichen Gesetzen in einen ganz bestimmten Blickwinkel. Explizit zur Sprache kommen muß die Interpretation kontrafaktischer Konditionalsätze sowie das Problem, welche kontrafaktischen Annahmen in Idealisierungen denn überhaupt als sinnvoll oder zulässig anzusehen sind. Damit ist man schon bei der Frage, wie man Idealisierungen wissenschaftsphilosophisch rechtfertigen kann, die mit der eher wissenschaftspraktischen Frage ihrer empirischen Testbarkeit verwandt ist. Die beiden erstgenannten Probleme treten bei Approximationen natürlich nicht auf, die beiden letztgenannten Fragen beantworten sich dort ziemlich trivial. Trotzdem werde ich dafür plädieren, daß die Vorzüge einer Idealisierungsanalyse

[5] Die relative Unerforschtheit von Idealisierungen ist aber insofern überraschend, als Galilei, der oft der Begründer „der" modernen wissenschaftlichen Methode genannt wird, ausgiebigsten Gebrauch von Idealisierungen gemacht hat. Vgl. dazu Such (1978) und McMullin (1985). Vgl. auch Kitchers (1985, S. 635) beiläufige Bemerkung: „since Galileo explained ballistics to the Gunners of Venice, scientific explanations have been playing ‚let's pretend'. We ignore common perturbations that could have slight effects ... and major potential disruptions with negligible probability ..."

[6] Diese Tradition wird heute noch gepflegt. Siehe die von Brzeziński u.a. (1990) herausgegebenen Sonderbände der *Poznań Studies* über Idealisierung.

[7] Dies ist eigentlich überraschend, denn einer der wichtigsten Punkte des Strukturalismus, nämlich die Idee, eine Theorie anstatt auf „die Welt" als ganzes auf viele kleine intendierte Anwendungen zu applizieren, erscheint für die sehr verbreiteten „Idealisierungen des abgeschlossenen Systems" (s. Abschnitt 8.2.4) geradezu maßgeschneidert. Man beachte auch die von Kuokkanen (1988) aufgewiesenen Parallelen zwischen dem Strukturalismus und der Poznań-Schule.

gewichtiger sind als ihre Nachteile. Das Verhältnis von Idealisierungen und Approximationen ist zu klären, denn selbstverständlich spielen letztere auch dann eine Rolle, wenn man erstere für konzeptuell wichtiger hält. Die Beispiele im letzten Kapitel haben schon gezeigt, daß die beiden Ansätze nicht auf dasselbe hinauslaufen (die ins Spiel kommenden Ähnlichkeitsrelationen sind also nicht äquivalent). Schließlich sollte man sich vergegenwärtigen, ob etwas und — wenn ja — was dafür spricht, Idealisierungen statt Approximationen ins Zentrum der wissenschaftstheoretischen Aufmerksamkeit zu stellen.

Einen großen Teil der eben aufgeführten Fragestellungen werde ich in diesem Kapitel ansprechen. Eine wirklich zufriedenstellende Behandlung würde das vorgegebene Maß an Raum und Zeit jedoch deutlich überschreiten, und so kann ich in den folgenden vier Abschnitten nur sehr vage und kursorisch Richtungen andeuten, in die weiter vorzustoßen wäre. Behauptungen, die kategorisch präsentiert werden, sind in den allermeisten Fällen cum grano salis zu verstehen.

9.2 Zwei Arten von kontrafaktischen Konditionalsätzen und Essentialismus in Naturgesetzen

Eines der meistdiskutierten und offenbar brauchbarsten Abgrenzungskriterien dafür, ob ein wahrer Satz der Form

(9.2.1) Alle F's sind G's.

ein Naturgesetz oder ein nur zufällig wahrer Allsatz (eine evtl. unendliche Konjunktion singulärer Sätze) ist, ist das folgende: Gesetze stützen kontrafaktische Konditionalsätze, bloße Allsätze tun dies nicht. Satz (9.2.1) ist demnach genau dann ein Gesetz, wenn es auch auf (Noch-)Nicht-F's anwendbar ist. Dann und nur dann, wenn man Wenn dieses x ein F wäre, dann wäre es auch ein G, abgekürzt durch

(9.2.2) $Fx \mathrel{\Box\!\!\rightarrow} Gx$,[8]

akzeptiert, betrachtet man (9.2.1) als ein Gesetz. Vielleicht sollte man den Satz (9.2.1) sogar durch $\forall x(Fx \mathrel{\Box\!\!\rightarrow} Gx)$ formalisieren — dies würde uns

[8] Da in diesem Kapitel keine indikativischen wenn-dann- oder weil-Sätze mehr vorkommen werden, schreibe ich statt $A^z \mathrel{\Box\!\!\rightarrow} B$ (s. Kapitel 6, v.a. die Abschnitte 6.5–6.7) im weiteren einfach $A \mathrel{\Box\!\!\rightarrow} B$.

allerdings insofern nicht weiterbringen, als keine allgemein anerkannte Konditionallogik auf prädikatenlogischem Niveau, ja meines Wissens nicht einmal ernstzunehmende Versuche in dieser Richtung, existieren.[9] Als eine sehr allgemein gehaltene Analyse von (9.2.2) wiederum kann folgende Bestimmung dienen: (9.2.2) ist genau dann wahr (oder akzeptabel), wenn gilt:

(9.2.3) Fx, ceteris paribus $\vdash$ Gx .[10]

Passende Verdeutschungen von „ceteris paribus" wären hier etwa „alle anderen Faktoren außer F bleiben gleich" oder „alle anderen Bedingungen außer Fx bleiben gleich".[11] Wenn x sich auf das Wasser in einem vor mir stehenden Topf bezieht, wenn Fx für „x wird auf 100 °C erhitzt" und Gx für „x siedet" steht, dann gehört zu den Ceteris-paribus-Bedingungen zum Beispiel, daß beim Erhitzen die chemische Zusammensetzung des Wassers erhalten und daß der Außendruck gleich bleibt (und zwar gleich dem sogenannten Normalluftdruck von 101325 Pascal), denn der Siedepunkt von Wasser erhöht sich mit zunehmendem Druck (um 0.04 °C pro Hektopascal). Die Schwierigkeit an dieser — ohnehin sehr vorläufigen — Explikation ist, daß nie *alle* Faktoren außer F gleich bleiben können, wenn der Faktor F verändert wird: Der „Faktor" G ist ein Beispiel. Auch der Anteil der gelösten Gase, die elektrische Leitfähigkeit, die Viskosität etc. und natürlich die mittlere kinetische Energie der in x enthaltenen Wassermoleküle werden mit F verändert. Es ist notorisch schwierig und *im allgemeinen* (siehe aber Abschnitt 9.5) höchst kontextabhängig, den Inhalt der Ceteris-paribus-Klausel genau zu bestimmen. Es ist aber wohl erlaubt zu sagen, daß x durch Fx *im wesentlichen* unverändert bleibt (d.h. $\neg$Fx gehört nicht zur „Essenz" von x) und daß es eine *wesentliche* Eigenschaft von x ist (d.h. zur Essenz von x gehört), bei Erwärmung auf 100°C zu sieden. Den *ersten Typ von Essentialismus* in Naturgesetzen nenne ich die in Gesetzen der Form (9.2.1) implizit enthaltene These, daß zwischen F und G ein wesentlicher Zusammenhang bestehen muß.

Die Mängel von (9.2.3) bestehen also in seiner Vagheit und seiner enorm schwierigen Interpretierbarkeit. Von nun an will ich aber davon ausgehen,

[9] Es wären auch mindestens dieselben Schwierigkeiten wie bei Quantifikationen in der üblichen Modallogik zu erwarten. Vgl. Garson (1984) und Cocchiarella (1984).

[10] Auch der Ramsey-Test (s. Kapitel 4) kann als Spezialfall dieser Formulierung aufgefaßt werden.

[11] Es wird häufig erwähnt, ist aber äußerst selten Gegenstand systematischer Betrachtungen, daß Ceteris-paribus-Bedingungen in der Wissenschaft eine große Rolle spielen. Die einzigen mir bekannten Ausnahmen sind Sklar (1964, S. 27–44), Patryas (1975) und Johansson (1975, S. 37–44, und 1980).

daß man Konditionalsätze der Form (9.2.2) auf aussagenlogischer Ebene befriedigend erfassen kann, und zwar durch die formale Analyse über Gärdenforssche Revisionen. Die Formulierung (9.2.3) wird dadurch modelliert, daß die Ceteris-paribus-Bedingungen von T unter der Annahme Fx einfach mit dem Inhalt der Kontraktion $T^{-}_{\neg Fx}$ identifiziert werden. Auf dieser Grundlage können wir uns weiteren und gravierenderen Problemen von Naturgesetzen zuwenden, die in der Literatur bis jetzt weniger häufig angesprochen wurden.

Diese neuen Probleme stellen sich so dar: Gesetze der Form (9.2.1) sind sehr oft entweder leer (d.h. es gibt in Wirklichkeit keine F's) oder falsch (d.h. es gibt zwar F's, die aber keine G's sind; nicht selten gilt sogar, daß kein F ein G ist). Das wird besonders dann augenfällig, wenn in F bzw. G exakte quantitative Beziehungen enthalten sind. Da diese praktisch nie *genau* erfüllt sind, hat man nur die Wahl, sich mit der Leere bzw. Falschheit des Gesetzes abzufinden oder aber das Gesetz auf fiktive, *ideale* F's zu beziehen: auf ideale Gase, auf ideal starre Körper, auf ideal abgeschlossene Systeme, auf ausdehnungslose Massenpunkte usw. Auch am harmlos aussehenden Beispiel des Wasserkochens wird das Dilemma sichtbar. Man kann das „Gesetz", daß Wasser bei 100 °C zu sieden beginnt, auf ideal reines Wasser (gibt es nicht) bei ideal konstantem (Normal-)Luftdruck (gibt es nicht) und ideal kontrollierbarer Energiezufuhr (gibt es nicht) beziehen oder eben zugeben, daß *diese Menge* Wasser x nicht wirklich bei genau 100°C zu sieden anfängt, etwa weil der hydrostatische Druck im Wasser oder die in x gelösten Stoffe die Siedetemperatur erhöhen (zum letzteren Punkt vgl. Schaefer 1958, S. 252–260, 311–315). Die statistische Mechanik als gegenüber der phänomenologischen Thermodynamik „überlegene" Theorie kann sogar dahin gehend interpretiert werden, daß es (praktisch und theoretisch) unmöglich ist, daß Wasser bei genau 100°C kocht. Die kinetische Energie der einzelnen Wassermoleküle variiert beträchtlich (ein einzelnes Molekül hat keine Temperatur), und Energie- bzw. Temperaturfluktuationen innerhalb des Wassertopfes sind (prinzipiell) nicht auszuschließen. (Dahinter steckt wieder die allgemeine Moral, daß neue, bessere Theorien sehr oft, vielleicht sogar prinzipiell die Gesetze ihrer Vorgängertheorien — mit manchmal großen, manchmal kleinen Abweichungen — als ungültig erweisen.) Eine besondere Schwierigkeit in diesem Beispiel besteht außerdem darin, daß man die Idealisierung nicht zu weit treiben darf: Sehr reines Wasser in einem sehr glatten Gefäß beginnt *nicht* bei 100 °C zu sieden, weil ohne sog. Blasenkeime ein Siedeverzug eintritt, wie er vom van der Waalsschen Gasgesetz (aber nicht von der Maxwell-Konstruktion,

vgl. Abschnitt 8.2.5) vorausgesagt wird. Wasser kann bis 270 °C überhitzt werden (s. Theimer 1978, S. 24–27).

Es ist also nicht nur so, daß bestimmte Faktoren (wie z.B. der Außendruck) *konstant* gehalten werden müssen, sondern man muß zusätzlich oder stattdessen postulieren, daß viele Faktoren, die man de facto (z.B. gelöste Fremdstoffe) oder prinzipiell (z.B. Energieschwankungen) nicht ausschalten kann, wegfallen oder zumindest völlig wirkungslos bleiben, um ein einfaches und präzis gültiges Gesetz aufstellen zu können. Statt „ceteris paribus" ist hier das Schlagwort „ceteris absentibus"[12] oder „ceteris neglectis":

(9.2.4) $Fx, \text{ceteris absentibus} \vdash Gx$.

Dieses künstliche, vereinfachende, abstrahierende, idealisierende Außerachtlassen von „Stör-" oder „Randfaktoren" kann nicht mehr, jedenfalls nicht mehr ohne Verdrehung des üblichen Verständnisses, mit (9.2.2) abgekürzt werden. Das kontrafaktische Konditional läßt wahre — bzw. nach der aktuellen Theorie T wahre — Hintergrundannahmen unerwähnt. Falsche (bzw. nach T falsche), *idealisierende* Hintergrundannahmen hingegen, die für die genaue Gültigkeit vieler Gesetze unabdingbar sind, bleiben in der Formulierung von Gesetzen zwar normalerweise unerwähnt, müssen in der Analyse jedoch als kontrafaktische Antezedenzien explizit gemacht werden:

(9.2.5) $I \mathrel{\Box\!\!\rightarrow} (Fx \mathrel{\Box\!\!\rightarrow} Gx)$.

Durch den Satz I seien hierbei die idealisierenden Hintergrundannahmen bezeichnet, die besagen, daß bestimmte, in $Fx \rightarrow Gx$ nicht erwähnte Faktoren unwirksam sind. Wenn man die Ceteris-absentibus-Bedingungen I zusammen mit den Ceteris-paribus-Bedingungen von $Fx \mathrel{\Box\!\!\rightarrow} Gx$ in einer Konjunktion A zusammenfaßt, kann man das zweite Konditional $\mathrel{\Box\!\!\rightarrow}$ in (9.2.5) eventuell[13] durch das materiale Konditional $\rightarrow$ ersetzen:

(9.2.6) $A \mathrel{\Box\!\!\rightarrow} (Fx \rightarrow Gx)$.

In vielen Fällen werden die idealisierenden Bedingungen I so beschaffen sein, daß sie von *keinen einzigen Objekt* des in Frage kommenden Gegenstandsbereichs erfüllt werden, d.h. daß gilt: $\forall x \neg Ix$. Andererseits scheinen wir bei Idealisierungen von der allquantifizierten Annahme $\forall x Ix$ ausgehen zu können, so daß x in I nicht frei vorkommt, und wir dürfen (9.2.5) in etwas natürlicherer Sprache wiedergeben:

(9.2.7) $I \mathrel{\Box\!\!\rightarrow}$ Alle F's sind G's .

[12] Zu meiner großen Überraschung habe ich diesen etwas absonderlich klingenden Ausdruck auch in Josephs (1985, S. 581) Rezension von Cartwright (1983) entdeckt.

[13] Vgl. aber Kapitel 4, Fußnote 5.

Die Trennung zwischen den Faktoren, bzgl. deren idealisiert wird, und den Faktoren, welche explizit in die Formulierung von Gesetzen (großer Abstraktheit, Allgemeinheit oder Einfachheit) eingehen, ist eine Trennung von unwesentlichen und wesentlichen Faktoren. Unwesentliches wird in der quasi elliptischen Formulierung (9.2.1) vernachlässigt und kommt erst in der Rekonstruktion, in I, zum Vorschein. In Naturgesetzen geht es also in zweifacher Hinsicht um wesentliche Eigenschaften von „Objekten wie x": Nicht nur die Verknüpfung zwischen F und G wird als eine wesentliche gekennzeichnet, sondern auch der Anspruch erhoben, daß in dem Gesetz, besonders in G, die wesentlichen Eigenschaften und Faktoren in Beziehung gesetzt werden. „Störungen" dieses Gesetzes durch „Nebenfaktoren" haben als unwesentlich zu gelten. Diese in Gesetzen der Form (9.2.1) implizit enthaltene These ist *der zweite Typ von Essentialismus* in Naturgesetzen. Während der Pfeil $\Box\!\!\rightarrow$ in (9.2.2) das typische Kennzeichen des Essentialismus im ersten Sinn ist, versinnbildlicht der Pfeil $\Box\!\!\rightarrow$ in (9.2.7) den Essentialismus im zweiten Sinn.[14]

In vielen Fällen wird man (9.2.7) noch weiter verallgemeinern und die I's einzelner Gesetze zum I einer ganzen Theorie T zusammenfassen können. Dann darf man schreiben:

(9.2.8) $\mathrm{I} \Box\!\!\rightarrow \mathrm{T}$.

Es ist klar, daß das Problem des zweiten Typs von Essentialismus in Naturgesetzen größer ist als das des ersten. Die ganze Problematik kontrafaktischer Konditionalsätze und ihrer Analyse durch Ceteris-paribus-Bedingungen ist beim zweiten Typ genauso akut wie beim ersten: Welche Wahrheiten kann man als bestehen bleibend ansehen, wenn man idealisierenderweise, sei es bewußt oder unbewußt, seine Gesetze und Theorien auf Voraussetzungen aufbaut, die falsch sind? Darüber hinausgehend haben wir jetzt aber das mindestens ebenso schwierige Problem, unsere Idealisierungen explizit machen zu müssen. Im ersten Fall von Essentialismus brauchen wir von den wahren (als wahr akzeptierten) Sätzen „nur" so lange welche zu streichen, bis eine Konsistenz mit Fx vorliegt. Im zweiten Fall aber wird die Frage vordringlich, welche idealisierenden, d.h. kontrafaktischen Annahmen I man überhaupt machen darf, damit ein gelungenes, erfolgreiches, sinnvolles oder in irgendeinem Sinne brauchbares wissenschaftliches Gesetz resultiert. Wie kann man ein solches I systematisch rechtfertigen? (Zu dieser Frage vgl. Abschnitt 9.4.) Wie kam die Wissenschaft, historisch

[14] Ein schönes Beispiel der Vermengung dieser beiden Arten von Essentialismus findet man bei Suppe (1976, S. 262f), der allerdings sehr gut die Unterscheidung zwischen Idealisierungen und Approximationen durchführt.

gesehen, zu guten Idealisierungen, und wie wurden sie gerechtfertigt?[15]

9.3 Idealisierungen und Theoriendynamik

Eine von den im letzten Abschnitt angestellten Überlegungen unabhängige Art, sich Idealisierungen zu nähern, ist die Betrachtung wissenschaftlichen Wandels. Wir haben uns in den ersten beiden Kapiteln mit der altehrwürdigen These der westlichen Wissenschaftstheorie beschäftigt, daß alte Theorien — eventuell nur im Grenzfall — auf die neueren, besseren Nachfolgertheorien reduzierbar seien. Das Pendant zu dieser These in der idealisierungsorientierten polnischen Wissenschaftstheorie ist angelegt in einer Idee von Izabella Nowakowa (1974; 1975), die den Verlauf der Wissenschaft als durch die Relation der „dialektischen Korrespondenz" bestimmt ansieht. Stark vereinfacht kann ihr Schema wie folgt dargestellt werden. Zunächst haben wir eine alte Theorie, die der Einfachheit halber aus nur einem Gesetz bestehen soll:[16]

T_1: $\forall x\,(Fx \rightarrow G_1x)$.

Dabei soll G_1 für eine quantitative Gleichung stehen, die alle für den betrachteten Objektbereich „wesentlichen" oder „primären" Größen in einen theoretischen Zusammenhang bringt. Später im Laufe der wissenschaftlichen Entwicklung entdeckt man, daß T_1 nur bei Vernachlässigung von gewissen sekundären Faktoren, d.h. nur unter bestimmten idealisierenden Annahmen gilt. Man geht von T_1 zur „Abstraktion" (L. Nowak) von T_1 über:

---

[15] Auf einer wichtigen Intuition basiert sicher auch die Idee von Lyon (1976/77, S. 117), der zwei Arten von kontrafaktischen Konditionalsätzen unterscheidet, die aus gesetzesartigen Aussagen „erschlossen werden können". Die erste Art stimmt mit unserer ersten Art (9.2.2) überein, die zweite Art ist jedoch völlig anders als (9.2.7):

Whatever else were to happen in the universe $\Box\!\!\rightarrow$ Alle F's sind G's .

Diese Formulierung soll den „nomischen Charakter" von Gesetzen ausdrücken, welche „in-this-world-unfalsifiable-come-what-may" seien. Lyon nennt solche Gesetze „ultimate laws" (S. 120) und behauptet, daß sie idealisierte Gesetze gemäß (9.2.7) ebenso erklärten wie das „contingent clutter", welches idealisierte Gesetze letztlich falsch macht. Ich finde die Idee Lyons ansprechend, aber seine Analyse ist so, wie sie dasteht, untauglich: Man nehme nur $Fa \wedge \neg Ga$ als Instanziierung von „whatever else were to happen in the universe". — Eine Lyon genau entgegengesetzte Sichtweise von letzten Gesetzen („fundamental laws") vertritt Cartwright (1983).

[16] Eine Theorie T sei i.f. wieder eine Satzmenge. Um T selbst satzförmig darstellen zu können, werde ich T gelegentlich mit der Konjunktion ihrer Axiome identifizieren (was zumindest dann legitim erscheint, wenn T endlich axiomatisierbar ist).

$\mathbf{A}(T_1)$: $\forall x\ (Ix \rightarrow (Fx \rightarrow G_1x))$.

In der Regel kann Ix kann hier so verstanden werden, daß bestimmte Parameter (Werte der Faktoren) für x gleich Null gesetzt werden. Nach Nowakowa darf man jedoch nicht bei $\mathbf{A}(T_1)$ stehen bleiben. Eine Abstraktion von T_1 ist nur dann zulässig, wenn gleichzeitig eine „Konkretisierung" $T_2 = \mathbf{K}(\mathbf{A}(T_1))$ von $\mathbf{A}(T_1)$ bereitgestellt werden kann:

T_2: $\forall x\ (\neg Ix \rightarrow (Fx \rightarrow G_2x))$.

Dabei soll in G_2x auch die Abhängigkeit der primären Faktoren von den in $\neg Ix$ erwähnten sekundären Faktoren (zumindest teilweise) mit verarbeitet sein. Einerseits ist G_2x im Normalfall unverträglich mit G_1x; andererseits soll der Zusammenhang zwischen T_1 und T_2 so sein, daß sich die Gleichung G_2x, wenn man darin die in Ix beschriebenen idealen Werte von x (in der Regel 0) einsetzt, auf G_1x reduziert (natürlich vorausgesetzt, daß die idealen Werte überhaupt in den entsprechenden, gegebenenfalls erweiterten Definitionsbereichen von G_2 liegen). Von diesem Einerseits-Andererseits rührt der Name „dialektische Korrespondenz" her. Während I. Nowakowa diese intertheoretische Relation zwischen T_1 und $\mathbf{A}(T_1)$ ansiedelt, plädieren andere Autoren (Krajewski, Such) für die naheliegendere Sprechweise, nach der T_1 und T_2 dialektisch korrespondieren. Jedenfalls ist es T_2, welches als bessere oder überlegene Nachfolgertheorie für T_1 in Frage kommt, da $\mathbf{A}(T_1)$ zu wenig informativ ist.

Als erste Verbesserung des Schemas von Nowakowa möchte ich vorschlagen, in $\mathbf{A}(T_1)$ das kontrafaktische Konditional $\Box\!\!\rightarrow$ an die Stelle des ersten materialen Konditionals $\rightarrow$ einzusetzen. Denn Ix ist ja tatsächlich *nicht* erfüllt. Besser also

$\mathbf{A}(T_1)$: $\forall x\ (Ix \Box\!\!\rightarrow (Fx \rightarrow G_1x))$.

Der schon im letzten Abschnitt benutzte Trick soll uns helfen, vom prädikaten- auf das aussagenlogische Niveau herunterzukommen. In den meisten Beispielfällen ist Ix für kein einziges x des fraglichen Anwendungsbereiches erfüllt, sei es aus empirisch-kontingenten, sei es aus (T_2-)theoretischen Gründen. Deshalb verallgemeinern wir hier und ziehen das (per Allquantifikation zu einer geschlossenen Formel gemachte) I — eine allgemeine Beschreibung des fiktiven, idealen Anwendungsbereiches — vor den Quantor:

$\mathbf{A}(T_1)$: $I \Box\!\!\rightarrow \forall x(Fx \rightarrow G_1x)$

oder einfach

$\mathbf{A}(T_1)$: $I \Box\!\!\rightarrow T_1$.

Die Form dieser Version der „Abstraktion" von T_1 ist uns bereits aus dem letzten Abschnitt bekannt (s. (9.2.8)). Wenn wir das $\neg I$ bei T_2 analog herausziehen:

T_2: $\quad \neg I \rightarrow \forall x(Fx \rightarrow G_2x)$,

dann brauchen wir natürlich kein kontrafaktisches Konditional an der ersten Stelle (unter der Voraussetzung von (T*E) würde es aber auch nichts schaden). Ja, hier kann das Antezedens sogar ganz weggelassen werden, da — vom Standpunkt der neuen Theorie T_2 aus betrachtet[17] — das Nichtvorliegen der idealen Bedingungen klar, d.h. ein Teil von T_2 ist. Damit sind wir bei meinem zweiten Vorschlag zur Verbesserung des Nowakowaschen Modells:

T_2: $\quad \forall x\ (Fx \rightarrow G_2x)$.

Die Abstraktion $\mathbf{A}(T_1)$ von T_1 ist intuitiv und wohl auch im Sinne von Nowakowa ein Teil von T_2 („Korrespondenz" zwischen T_2 und T_1). Dies bedeutet, im einfachen Statement view ausgedrückt, $\mathbf{A}(T_1) \in T_2$, was nun nach der Analyse von kontrafaktischen Konditionalsätzen durch den (starken) Ramsey-Test mit Randbedingungen (vgl. Kapitel 6, Bedingung (6.7.7)) nichts anderes heißt als

(9.3.1) $\quad \neg I, \neg T_1 \in T_2 \wedge T_1 \in T_2{}^*{}_I$.

$T_2{}^*{}_I$ ist natürlich wieder die minimale Änderung von T_2, die für einen T_2-Theoretiker nötig ist, damit er I akzeptieren kann. Mit (9.3.1) sind wir bei einer Bedingung gelandet, die an verschiedenen Stellen dieses Buchs[18] als vielversprechendes Explikat für eine zentrale intertheoretische Relation zwischen Nachfolgertheorien in Erscheinung trat.

Zwei frühere Versuche, Izabella Nowakowas Schema zu verbessern, sollen noch erwähnt werden. Krajewskis (1974, 1976, 1977) „sophisticated" oder „renewed implicative version" des Korrespondenzprinzips ist dazu gedacht, die nach seiner Ansicht bei Nowakowa verloren gegangene Implikationsbeziehung zwischen Vorgänger- und Nachfolgertheorie wiederherzustellen. Krajewski sieht das Problem, daß I oft mit der neuen Theorie T_2 (und nicht „nur" mit empirischen Gegebenheiten) unverträglich ist.[19]

[17] „Theorie" wird hier, wie stets in diesem Buch, in dem sehr umfassenden Sinn verstanden, daß das Wissen um relevante empirisch-kontingente Anfangsbedingungen mit zur Theorie gehören soll.

[18] Vgl. die Bedingung (1.10.2), das Ende von Abschnitt 2.4.1, die Bedingung (7.4.4) und ihre Anwendung in Kapitel 8.

[19] Krajewski (1977, S. 10) schreibt: „The situation is paradoxical. We draw conclusions from false assumptions although we know from elementary logic that 'the falsity entails everything'. Hence, we perform a kind of reasoning forbidden by logic." Gutwillig

Er schlägt vor, eine mit I konsistente, „abstraktisierte“[20] Version T_2^* von T_2 zu verwenden, mit dem Effekt, daß man statt $I \boxdot\!\!\rightarrow T_1$ das gewöhnliche $I \rightarrow T_1$ verwenden kann, oder anders ausgedrückt:

(9.3.2) $T_2^*, I \vdash T_1^*$.

T_1^* soll hier für eine Interpretation von T_1 im Lichte von T_2 stehen, in der — so Krajewski — die idealisierenden Annahmen von T_1 aufgedeckt werden. Meines Erachtens hat Krajewski hier genau das richtige Problem getroffen. Aber erstens ist seine Formalisierung arg verfehlt (was schon von Nowak 1974 und Yoshida 1978 bemängelt wird), und zweitens ist seine „Lösung“ gar keine richtige Lösung, wenn man nicht weiß, *wie* man allgemein von T_2 zu T_2^* (und von T_1 zu T_1^*) kommen soll. Schließlich haben wir oben schon quasi von selbst eine Ableitungsbeziehung erhalten, denn $T_1 \in T_2^*{}_I$ ist für eine *deduktiv abgeschlossene* Revision $T_2^*{}_I$ äquivalent mit $T_2^*{}_I \vdash T_1$. Was Krajewski mit T_2^* vermutlich meint, aber unexpliziert läßt, nämlich die Kontraktion $T_2{}^-{}_{\neg I}$, wird im Gärdenforsschen Revisionsmodell transparent und über die Relation der theoretischen Wichtigkeit mehr oder weniger konstruktiv darstellbar.

Der zweite und solidere Verbesserungsvorschlag stammt von Niiniluoto (1986). Zunächst einmal läßt er — wie wir oben — das Antezedens $\neg Ix$ in T_2 weg. Der wesentlichere Unterschied zu Nowakowa ist aber, daß Niiniluoto auch ein „intensional if-then-connective“ (S. 266; wir wählen i.f. dafür provisorisch das Zeichen „$\hookrightarrow$“) bei der Formulierung von Gesetzen ansetzt, allerdings an ganz anderer Stelle als wir. Er beschreibt den Prozeß der Konkretisierung, ebenfalls weitgehend vereinfacht dargestellt, so (vgl. Niiniluoto 1986, S. 277f):

T_1: $\forall x\,(Fx \hookrightarrow G_1x)$,
$\mathbf{A}(T_1)$: $\forall x\,(Fx \wedge Ix \hookrightarrow G_1x)$,
T_2: $\forall x\,(Fx \hookrightarrow G_2x)$.[21]

interpretiert, heißt dies wohl, daß die Expansion der Theorie T_2 durch die falschen, idealisierenden Annahmen I zur inkonsistenten Theorie $T_\perp$ führt.

[20] Das ist hier leider ein Homonym und hat mit dem obigen — Nowaks — Gebrauch von „Abstraktion“ nichts zu tun, sondern ist formal sogar eher das Gegenteil: Bei Nowak erhält man eine Abstraktion durch das Voranstellen eines (idealisierenden) Antezedens, während man bei Krajewski durch das Streichen eines (realistischen) Antezedens zu einer Abstraktion kommt.

[21] Wie bereits angedeutet, fände ich es besser, von der prädikatenlogischen Ebene Abstand zu nehmen und wenigstens statt $\mathbf{A}(T_1)$ bescheidener $I \hookrightarrow T_1$ zu schreiben. Von Niiniluotos Vorschlag aus könnte man dies eventuell dadurch rechtfertigen, daß man von $\mathbf{A}(T_1)$ auf $\forall x(Ix \hookrightarrow (Fx \hookrightarrow G_1x))$ schließt (vgl. aber Kapitel 4, Fußnote 5), und von dort wiederum auf $(\forall x Ix) \hookrightarrow (\forall x(Fx \hookrightarrow G_1x))$, was mit $I \hookrightarrow T_1$ identisch ist. Wie in Abschnitt

In G_2 sollen die in I genannten und in G_1 ignorierten Faktoren Berücksichtigung finden. Falls das *Korrespondenzprinzip* erfüllt ist, dann bedeutet dies bei Niiniluoto (1986, S. 278) per Definition, daß die neue Beziehung G_2x bei Annäherung an den durch die idealisierenden Bedingungen Ix charakterisierten Zustand zur alten Beziehung G_1x degeneriert.[22] In der Terminologie Niiniluotos heißen Gesetze, die wie T_1 und T_2 keine kontrafaktischen Annahmen explizit machen, „factual laws", alle anderen „idealizational laws"; von den letzteren unterscheidet er sogenannte „idealized laws", die wie hier T_1 (und vermutlich auch $\mathbf{A}(_1)$ und T_2) zumindest einen relevanten Faktor unerwähnt lassen.

Der Gebrauch des „intensionalen Wenn-dann-Konnektivs" $\hookrightarrow$ bei Niiniluoto entspricht offenbar dem, was wir oben als die erste Art von Essentialismus lokalisiert haben. Dieser ist, im Gegensatz zur zweiten Art, nicht spezifisch für idealisierte Gesetze, und das typische Problem, welche I's noch als geglückte Idealisierungen anzusehen seien, wird nicht berührt. Deshalb ist Niiniluotos Ansatz auch nicht völlig ausreichend.

Wenn wir also das verbesserte Modell der dialektischen Korrespondenz als für wissenschaftlichen Fortschritt typisch akzeptieren (die polnischen Wissenschaftstheoretiker liefern eine ganze Anzahl von in der Regel allerdings recht oberflächlich behandelten Belegbeispielen) und *wenn* ein solcher Fortschritt nie zum Stillstand kommt, dann sind alle wissenschaftlichen Gesetze und Theorien — im Lichte ihrer Nachfolgertheorien besehen — idealisiert.[23] Dies bestätigt auch Duhem (1906/78, S. 227–236), der

---

9.2 bemerkt, haben in der letzten Formulierung die beiden $\hookrightarrow$ ganz verschiedene Funktion. Zum Beispiel sollte jetzt nicht mehr $\forall x \neg Fx$ gelten, wohingegen Niiniluoto explizit annimmt, daß $\forall x \neg Ix$ gilt.

[22] Dies läßt sich natürlich besser ausdrücken, wenn man — wie Niiniluoto u.a. — Gesetze stets als quantitative Gesetze angibt. In Anwendung auf

T_1: $\forall x\,(Fx \hookrightarrow f(x)=g_1(x))$,
$\mathbf{A}(T_1)$: $\forall x\,(Fx \wedge h(x)=0 \hookrightarrow f(x)=g_1(x))$ und
T_2: $\forall x\,(Fx \hookrightarrow f(x)=g_2(x,h(x)))$

(wobei f, g_1, h, g_2 reellwertige Funktionen seien, und für alle wirklichen Objekte x $h(x)>0$ gelte) liest sich das Korrespondenzprinzip zum Beispiel so:

$$\forall x \lim_{y \to 0} g_2(x,y) = g_1(x).$$

In diesem Fall macht es Sinn zu sagen, „T_2 impliziert T_1 (oder $\mathbf{A}(_1)$) als Grenzfall", auch wenn 0 gar nicht im Definitionsbereich für das zweite Argument von g_2 liegt. (Man beachte übrigens, daß x in I nicht unbedingt vorzukommen braucht, weshalb es manchmal irreführend ist, Ix oder $h(x)=0$ für I zu schreiben.)

[23] Dies darf man aber nicht mit der noch fortschrittsoptimistischeren Behauptung verwechseln, die Wissenschaft finde immer neue grundlegendere, einfachere, einheitlichere Theorien, welche ihre Vorgängertheorien erklären. Das Nowakowasche Modell schreitet

davon spricht, daß jedes physikalische Gesetz „provisorisch" sei.[24] So gesehen wären die in Abschnitt 9.2 betrachteten Idealisierungen sozusagen die Projektion eines dynamischen Phänomens („Korrespondenz" oder „Reduktion") auf die statische Ebene einer isoliert betrachteten Theorie.

9.4 Wie rechtfertigt man Idealisierungen?

Wenn nun Naturgesetze Idealisierungen, also entweder leer oder falsch sind, warum hält man dann an ihnen fest? Worin besteht ihre Rechtfertigung und ihr Wert? Wodurch unterscheiden sie sich von anderen, völlig verrückten falschen Aussagen und Generalisierungen? Ich denke, es gibt drei deutlich verschiedene Antworten auf diese Fragen.[25]

Diese Antworten haben eine gewisse Ähnlichkeit mit Noretta Koertges (1973, Abschnitte IIIA, IIIB und IVC) zwei Antworten auf die Frage, warum manche (Kern-)Aussagen wissenschaftlicher Theorien als „preferred statements" behandelt werden, und ihrer dritten Antwort auf die Frage, wie man solche bevorzugten Aussagen kritisieren kann.[26]

---

von einfacheren zu komplexeren Theorien fort, und auch in meiner kleinen Abwandlung wird das G_2 im allgemeinen viel komplizierter sein als das G_1. Für Krajewski (1984) macht das einen terminologischen Unterschied: Seiner Meinung nach kann eine kompliziertere Nachfolgertheorie ihre Vorgängertheorie nicht *erklären*, auch wenn letztere im Sinne einer Korrespondenzrelation auf erstere *reduzierbar* ist.

[24] Duhem nennt zwei Arten von Gründen für den ephemeren Charakter von physikalischen Gesetzen: erstens, weil sie „angenähert", und zweitens, weil sie „symbolisch" seien. Der erste Punkt entspricht der Approximations-, der zweite Punkt recht genau der hier vertretenen Idealisierungsperspektive.

[25] Die Beantwortung der philosophischen Frage, wie man Idealisierungen rechtfertigen kann, läßt einige Aufschlüsse für die Antwort auf die praxisnähere Frage zu, wie man Idealisierungen testen kann. Auf das letztere Problem gehe ich hier nicht ein, sondern verweise auf Suppe (1974b), Patryas (1977) und Laymon (1985).

[26] Vgl. hiermit auch Scriven (1961, S. 101): „The breakthrough comes with the first step — with a wild approximation whose virtues are wholly negative, or simply vaguely related in a limited area. This, I believe, is why we happily refer to propositions as laws which are known to be inaccurate. It is not mere historical habit, but a recognition of these propositions as *framework claims*, valid for certain limiting cases and known-to-be-false assumptions which are nevertheless *related to* actual conditions." (Erste Hervorhebung von mir.) Vgl. damit auch Toulmins (1961) gleichzeitig vorgebrachten „ideals of natural order". — In neuerer Zeit schreibt Ellis (1979, S. 100) ähnlich : „For me, scientific laws are generally *framework principles*, providing idealizations of behaviour, against which actual behaviour may be measured or compared." (Hervorhebung von mir.) Mir scheint, mit solchen „Rahmenprinzipien" ist dasselbe gemeint ist wie mit Koertges „bevorzugten Aussagen". Vor diesem Hintergrund sieht die folgende *These* plausibel aus: Alle Idealisierungen sind — mehr oder weniger — bevorzugte Aussagen.

Die *erste Antwort* liegt nahe: Idealisierte Gesetze sind eben nicht schlichtweg, sondern nur *genaugenommen* falsch, d.h. annähernd wahr, jedenfalls für einen gewissen Anwendungsbereich. Diese Antwort, die auf den Approximationsbegriff (im ersten Sinn) abhebt, ist die bei weitem verbreitetste. Natürlich ist sie eine gute Antwort und hat eine große praktische Relevanz für den Entdeckungszusammenhang: Wie kommt man sonst überhaupt auf dieses oder jenes idealisierte Gesetz? Dennoch kommt sie mir systematisch nicht völlig befriedigend vor.[27] Es ist in den meisten Fällen ja gar nicht schwer, zu einer endlichen Menge von Beobachtungswerten oder Meßergebnissen eine Kurve zu finden, die in einem gewissen Bereich einigermaßen nahe der vorgegebenen Punkte verläuft.[28] Will man eine bestimmte, größere Genauigkeit erreichen, schränkt man den Gültigkeits- oder Anwendungsbereich der Kurve einfach geeignet ein. Oft gibt es auch Fälle, die man eigentlich gerne als intendierte Anwendungen mit erfaßt hätte, in denen der Graph des Gesetzes jedoch leider sehr weit von den empirisch ermittelten Werten entfernt ist. Denn diejenigen Faktoren, welche vom Standpunkt des fraglichen Gesetzes aus als Rand- oder Störfaktoren einzuordnen sind, können unter bestimmten Umständen ganz enorme Auswirkungen haben. Unmittelbare „Wahrheitsnähe" oder approximative Gültigkeit eines Gesetzes ist vermutlich nicht der springende Punkt bei Idealisierungen. Laymon (1977) weist darauf hin, daß — ein typisches Beispiel — die ersten *erfolgversprechenden* Berechnungen Einsteins zur Brownschen Bewegung immerhin um den Faktor 6 bis 7 von den gemessenen Werten Svedbergs abwichen.[29]

Das Wort „erfolgversprechend" führt zur *zweiten Antwort*. Grob gesagt,

Möglicherweise ist diese These sogar als a priori wahr einzusehen; ich möchte aber nicht näher darauf eingehen. Die Umkehrung der These ist wohl nicht gültig, wie man — *pace* Cartwright — am Newtonschen Gravitationsgesetz sehen kann.

[27] Das ist sogar, wie in Kapitel 8 deutlich geworden sein dürfte, etwas untertrieben. Ich habe genau denselben Verdacht wie Frederick Suppe (1976, S. 247): „While the approximate truth approach has plausibility, my suspicion is that it is not the most promising way to deal with the inaccuracy of laws; for I suspect that the inaccuracy of physical laws is a manifestation of relatively deep structural and epistemological properties of laws and theories which will be obscured or missed if one attempts to analyze laws as approximately true generalizations (or other sorts of propositions)."

[28] Manchmal handelt es sich um einen „mittleren", „normalen" Bereich, manchmal handelt um einen „extremen" Rand- oder Grenzbereich. Diese erste Antwort wird um so interessanter, je besser man natürliche, wohldefinierte Bereiche abgrenzen kann, in denen das idealisierte Gesetz möglichst genau zutrifft, während es in dazu komplementären Bereichen möglichst klar inadäquat ist.

[29] Interessanterweise wurden in diesem Fall ziemlich bald nicht die theoretischen Berechnungen, sondern die experimentellen Ergebnisse korrigiert.

ist ein idealisiertes Gesetz genau dann *erfolgversprechend*, wenn das korrekte Miteinbeziehen aller im Gesetz vernachlässigten, d.h. kontrafaktisch als wirkungslos angenommenen Faktoren zu einer Modifikation des Gesetzes führt oder führen würde[30] , die im ganzen Anwendungsbereich mit den empirischen Werten ausgezeichnet im Einklang steht. Es ist zwar von großem wissenschaftlichen, aber nur von geringem philosophischem Interesse, ob man eine Idealisierung tatsächlich explizit zu einer endgültigen „Konkretisierung"[31] verfeinern kann oder ob man sich mit groben Abschätzungen der Wirkung von Störfaktoren begnügen muß, d.h. mit einer Abschätzung, daß, sofern nur genügend Daten, Gesetze und Mathematik für die Berechnung der komplizierten Zusammenhänge zur Verfügung *stünden*, eine solche endgültige Konkretisierung herauszubekommen *wäre*.[32] Wichtig ist, daß man „auf lange Sicht" oder „prinzipiell" bei einer sehr guten — oder jetzt eigentlich besser: bei einer exakten — Gültigkeit des (modifizierten) Gesetzes landet bzw. landen würde.

Die zweite Antwort besteht also in einem solchen Argument: Wenn man auf das in Frage stehende Gesetz Schritt für Schritt alle ignorierten Faktoren „daraufsetzt"[33] (bzw. daraufsetzen würde), so lange, bis man alles berücksichtigt und eine „realistische" Beschreibung des Objektbereichs hat (hätte), dann muß (müßte) das solchermaßen modifizierte Gesetz stimmen. Im Laufe dieses Prozesses wird aber das ursprüngliche Gesetz komplizierter und komplizierter, in der Regel so kompliziert, daß man praktisch nichts Interessantes mehr effektiv ausrechnen kann. Außerdem gewinnt man, in einem bestimmten Sinne (vgl. Krajewski 1984 und 1977, S. 30 und 39), nichts an Erklärungskraft, sondern verliert eher daran, weil man an Einfachheit verliert. Deshalb ist es besser und das vornehmliche Ziel aller Wissenschaftler, anstelle einer „zurechtgeflickten" Konkretisierung eine völlig neue Theorie zu finden, die mindestens genauso einfach wie das alte idealisierte Gesetz ist, die aber — über Brückenprinzipien — bereits eine

[30] Man beachte die neue Art von Kontrafaktizität im zweiten Adjunktionsglied: Es muß durchaus nicht der Fall sein, daß alle relevanten Faktoren effektiv, d.h. ausrechenbar, verarbeitet werden können. Vgl. auch Scriven (1962, S. 215) zum „promissory character of explanations".

[31] Dies ist der in der Poznań-Schule gebräuchliche Terminus. Krajewski benutzt den Terminus „Faktualisierung".

[32] Für Shapere (1974, S. 561f, Fußnote 65) ist das ein Unterscheidungsmerkmal von „Vereinfachungen" und „Idealisierungen".

[33] L. Nowak schreibt sein Schema so, daß sich die Konkretisierung eines Gesetzes immer aus dem ursprünglichen Gesetz und einer geignet verarbeiteten Korrekturgröße zusammensetzt.

Konkretisierung des alten Gesetzes abzuleiten erlaubt. Damit sind wir bei der *dritten Antwort.*

Kandidaten für solche einfacheren, überlegenen Theorien findet man glücklicherweise immer wieder im Verlauf der Wissenschaften: für Keplers Gesetze der Planetenbewegung ist dies die Newtonsche Gravitationstheorie, für das ideale Gasgesetz die kinetische Gastheorie, für Newtons Dynamik die spezielle Relativitätstheorie, für die klassische Thermodynamik die statistische Mechanik, für die klassische Genetik die Molekulargenetik. Man kann dann sagen, die neue, grundlegendere Theorie erklärt die alte idealisierte Theorie und korrigiert sie gleichzeitig — per „Konkretisierung". Im bestmöglichen und trotzdem nicht ganz utopischen Fall (s. Kapitel 8) gibt die neue Theorie in einem präzisierbaren Sinn (s. Kapitel 7) Auskunft auf die Fragen Warum ist das alte Gesetz falsch? (oder, was auf dasselbe hinausläuft, Inwiefern ist es „nur" eine Idealisierung? und Unter welchen Bedingungen wäre es tatsächlich gültig? (Eine verwandte, aber keineswegs gleichwertige Frage wäre Unter welchen Bedingungen ist es annähernd gültig?[34]) Auf die erste Frage antwortet eine neue, überlegene Theorie Weil das alte Gesetz idealisierenderweise so getan hat, als seien die-und-die Faktoren ohne Belang, weil sie aber faktisch doch relevant sind, auf die zweite Frage Wenn die genannten Faktoren tatsächlich ohne Belang wären, dann würde das alte Gesetz stimmen. (Oft, aber keineswegs automatisch ist auch dies richtig: Wenn diese Faktoren klein genug sind, dann stimmt das alte Gesetz ungefähr.[35])

Neue Theorien sind allerdings auch nicht perfekt. Sie müssen in der Praxis immer, im theoretischen Aufbau so gut wie immer ebenfalls idealisieren. Die statistische Thermodynamik ist sicher „realistischer" als die

[34] Eine gewisse Gleichwertigkeit ergäbe sich höchstens dann, wenn man vermutlich falsche, metaphysisch klingende Postulate wie „Die Welt ist stetig." oder „Kleine Ursachen haben kleine Wirkungen." als gültig annehmen könnte.

[35] Eine weitere verwandte Frage ist: „Wie muß man welche Bedingungen variieren, damit das idealisierte Gesetz besser stimmt?" Laymon (z.B. 1982, S. 115) meint, daß die typische Rechtfertigung eines idealisierten Gesetzes so aussieht: Wenn man von immer realistischeren Anfangsbedingungen ausgeht, so liefert das Gesetz immer genauere Voraussagen. Dies ist dem mathematischen Grenzwertbegriff sehr ähnlich, allerdings heißt $\lim_{x \to 0} f(x) = 0$ nicht „für kleineres x ist f(x) stets näher an 0", sondern nur „letztendlich ist für kleine x f(x) beliebig nahe an 0." Man beachte, daß auch der Grenzwertprozeß die in der letzten Fußnote genannten metaphysisch klingenden Postulate präsupponiert. Meiner Ansicht nach ist das *eine*, zugegebenermaßen eine sehr wichtige Methode, man mit kontrafaktischen Annahmen (speziell mit einer Grenzwertannahme wie x=0) umzugehen. Das kontrafaktische Modell ist aber allgemeiner, da es nicht von vornherein auf diese Methode festgelegt ist.

klassische, aber ihre Modelle sind, soweit sie praktisch handhabbar sind, immer noch weit von dem entfernt, was man als tatsächlich gegeben ansieht: Ihre Gasmoleküle sind ideal rund, sie rotieren nicht, sie stoßen ideal elastisch zusammen, das Potential der intermolekularen Kräfte ist ideal einfach, sie stehen nicht in Wechselwirkung mit ihrer Umgebung (Gefäßwand etc.), die Geschwindigkeitsverteilung ist nach Größe und Richtung idealisiert etc.

Demgemäß kann man vermutlich die intertheoretische Rechtfertigung (siehe die dritte Antwort) und die intertheoretische Kritik (vgl. Koertge 1973, Abschnitt IVC) immer weiter fortsetzen. Immer wieder glückt — so hofft man — den Wissenschaftlern die Entdeckung solcher tiefliegenden, relativ einfachen neuen Theorien (was oft eine „ontologische Reduktion", speziell eine „Mikroreduktion" mit sich bringt). Aber selbst wenn das ewig so weitergeht, es gibt keine Aussicht, daß man irgendwann mit irgendwelchen sehr allgemeinen Universaltheorien in der Praxis irgendetwas Interessantes *genau* ausrechnen kann, da diese Aufgabe allzu kompliziert erscheint. Deswegen wird man, abgesehen vielleicht von den grundlegendsten Grundlagenproblemen, immer idealisierte Gesetze brauchen (und dabei von ihrer approximativen Wahrheit profitieren).

9.5 Was ist eine wissenschaftliche Theorie?

Noch einmal: Wissenschaftliche Gesetze und Theorien sind falsch, wenn auch nur genaugenommen falsch. Sie sind Idealisierungen, sie sind nützliche Fiktionen. Irgendetwas sollte also an einer Theorie — an der Theorie selbst! — dran sein, was uns zu entscheiden erlaubt, ob sie eine dumme, abwegige oder eine kluge, nützliche Idealisierung darstellt. In der Literatur über Idealisierungen kann man mehrere Vorschläge finden, diesem Problem (welches sich nur für die Idealisierungsperspektive stellt, aber auch nur von dieser aus gesehen werden kann) durch einen raffinierteren Theorienbegriff gerecht zu werden. Ich will drei solche Ansätze vorstellen, indem ich die jeweiligen Definitionen einer Theorie zitiere:

> ... any principle may be construed as a set of statements — usually laid out in roughly the order of their degree of approximateness — in which more and more of the scope conditions are suppressed. ... A theory, then, is a set of such explanatory principles (statement sets) together with the „families" of explanations and anomaly contexts in which they occur. (Hum-

> phreys 1968, S. 154)
>
> *Theory* may be identified with a sequence of statements T^k, T^{k-1}, ..., T^1, T^0, when T^k is a law of science and T^{i-1} is a concretization of T^i; subscript[sic!] i shows the amount of idealizing conditions. (Nowak 1975, S. 39)
>
> A theory then may be considered as a function that maps increasingly realistic but still idealized descriptions (i.e., models of the same phenomena) into members of the prediction space. (Laymon 1985, S. 156)

Obwohl diese Ideen, bloß als kurze Zitate präsentiert, nicht besonders gut verständlich sind, wird schon klar, daß sie sich nicht allzu viel gemeinsam haben. Humphreys legt seine Theoriendefinition im Anschluß an eine Diskussion von Idealisierungen vor, wobei sich dieser Terminus bei ihm auf „the use of scope or boundary conditions which are never realized in nature" (Humphreys 1968, S. 121) bezieht.[36]

Nowak bleibt ebenfalls im Rahmen der Idealisierungsidee: Wenn man die idealisierenden Annahmen des Gesetzes T^k nach und nach beseitigt, indem man die verschiedenen vernachlässigten Faktoren in Rechnung stellt, dann landet man bei dem minimal idealisierten — und evtl. „faktischen" — Gesetz T^0; ist T^0 (annährend) wahr, so stellt T^k eine gute Idealisierung dar. Dagegen beschreibt Laymon im wesentlichen eine Art Grenzwertprozeß: Wenn man die tatsächlich realisierbaren Anwendungsbedingungen der Theorie nach und nach so abändert, daß sie den idealisierenden Annahmen immer näher kommen, bekommt man immer wieder neue Voraussagen; sind diese Voraussagen entsprechend näher an der empirisch ermittelten Wahrheit, dann stellt die Theorie eine gute Idealisierung dar.[37]

[36] Vollständig verständlich wird Humphreys' Definition nicht, denn bei seiner dem Zitat vorangehenden Beispielsdiskussion (die Entdeckung des Neptun) spielen „closure assumptions" eine entscheidende Rolle, und der Zusammenhang solcher Annahmen mit „scope conditions" (Toulmins Begriff) ist nicht ganz klar. (Zu „closure assumptions" vgl. auch Scrivens (1963, S. 112–116) „completeness claims" und Johanssons (1980) „closure clauses"). — Übrigens sei darauf hingewiesen, daß der in Humphreys' Definition auftauchende Approximationsbegriff nicht im Standardsinn zu verstehen ist. Man sehe zum Beispiel: „An approximative inference in a physical theory is, by definition, one in which observation statements are deduced without full knowledge of the scope conditions affecting the system we are dealing with." (Humphreys 1968, S. 124) Während die meisten Autoren Approximationen und Idealisierungen nur mangelhaft auseinanderhalten, indem sie letztere unreflektiert zu ersteren rechnen, scheint Humphreys sozusagen umgekehrt Approximationen von vornherein als Idealisierungen zu verstehen.

[37] Laymons Kriterium ist auch nicht zwingend. Denn wenn zwei falsche Annahmen sich zufällig so ausgleichen, daß sie zu einem sehr guten Ergebnis führen, dann führt die

Wenn sich die drei Theorienbegriffe auch im einzelnen deutlich unterscheiden, so haben sie doch das Entscheidende gemeinsam. Eine Theorie wird stets dynamisch gesehen, als etwas, was mit mehr oder weniger idealisierten (d.h. weniger oder mehr realistischen) Bedingungen zu tun hat. Ein Unterschied zwischen den drei Vorschlägen liegt im Unterschied zwischen Theorie und Anwendung. Bei Humphreys und Nowak wird dieser dynamische Aspekt explizit in den Theorienbegriff mit eingebaut. Laymon dagegen verlagert ihn in die Anwendungen und betrachtet, strukturalistisch interpretiert, eine Sequenz von idealisierten intendierten Anwendungen I_k, I_{k-1}, ..., I_1, I_0, wobei I_k die Klasse derjenigen potentiellen Modelle ist, für die die Theorie am meisten ausrechnen kann (weil es am wenigsten Faktoren gibt), während I_0 die potentiellen Modelle umfaßt, die die wirkliche Welt am getreulichsten wiedergeben (nur dies wären wohl die real existierenden „intendierten Anwendungen" des Strukturalisten).[38]

Aus diesen Überlegungen geht hervor, daß eine bloße Satzmenge — ich bleibe hier, wie meist in diesem Buch, im Statement view — zu wenig Information beinhaltet, um gute von schlechten Idealisierungen zu unterscheiden. Denn sei T eine idealisierte Theorie. Wir benötigen Informationen für zwei Richtungen. In der einen, der „Vorwärts-Richtung" ist der folgende kontrafaktische Konditionalsatz interessant: Wenn wir alle in T unterschlagenen Randfaktoren mit einbeziehen würden, dann wäre T korrekt. Die Richtigkeit dieses Satzes ist jedoch sehr schwer zu beurteilen: Eine *explizite* Verrechnung *aller* sogenannten Rand-, Rest- oder Störfaktoren ist praktisch immer ausgeschlossen, man wird sich im allgemeinen mit Argumenten begnügen müssen, ob damit überhaupt die „Wahrheitsnähe" der Voraussagen vergrößert würde oder nicht.[39] In der anderen, der „Rückwärts-Richtung" kann man aber viel mehr sagen. Man kann ja davon ausgehen, daß T Nachfolgertheorie einer älteren, schlechteren Theorie T′ ist und daß T — dies wäre die spekulative These — „in geringerem Maße idealisiert" ist als T′,[40] oder konkreter: daß in T (mindestens) ein Faktor I, der in T′

Ersetzung einer Annahme durch eine realistische Anwendungsbedingung zu schlechteren Ergebnissen. Ein bekanntes Beispiel für solch einen glücklichen Ausgleich zweier Fehler stellen die Messungen von Eratosthenes und von Poseidonos dar, die zweifellos gute Theorien zur Bestimmung des Erdumfangs hatten.

[38] Ein Plädoyer dafür, das Kontrafaktische an Theorien nicht auf die Gesetze selbst, sondern auf ihre Anwendung zu beziehen, findet man bei Suppe (1974a, S. 42–45, 213–215, 223–226).

[39] Dies ist ein wichtiger Punkt Laymons, vgl. z.B. die „modal auxiliaries" in Laymon (1980).

[40] Dies heißt in etwa, daß T im Sinne von Mühlhölzer (1988) objektiver ist als T′.

als Randfaktor ignoriert wurde, explizit mit einbezogen wird. Ein typischer kontrafaktischer Konditionalsatz von T lautet dann etwa so: Wenn $\forall x(I(x)=0)$ wahr wäre, dann würde T' stimmen.

Meines Erachtens sollten solche typischen Konditionalsätze als *Teile der Nachfolgertheorie* aufgefaßt werden. Das heißt, sie müßten schon *allein durch T* als wahr oder falsch (akzeptabel oder inakzeptabel) zu erweisen und nicht erst — wie etwa van Frassen (1980, 1981) meint — durch den Kontext verständlich und auflösbar sein. Damit haben wir aber genau den Punkt markiert, in dem der alte Statement view ohne Zusatz nicht ausreicht. Denn einer Satzmenge allein sieht man nicht an, wie man sie zu verändern hätte, um hypothetische Annahmen wie $\forall x(I(x)=0)$ zu machen. Die Satzmenge braucht eine zusätzliche Struktur, damit sie zu einer Theorie wird, die den Umgang mit kontrafaktischen Annahmen und Konditionalsätzen regeln kann. Mein Vorschlag ist, daß die in Kapitel 3 untersuchte Relation $\leq$ der theoretischen Wichtigkeit ein taugliches Modell für diesen Zweck abgibt. Wie das funktioniert, wurde in Kapitel 8 angedeutet, muß aber an weiteren Beispielen im Detail überprüft werden. Wenn eine größere Anzahl von Rekonstruktionsversuchen glückt, dann ist die Idee gestützt, daß man eine Theorie als ein Paar $\langle T,\leq\rangle$ modellieren kann und soll, wobei T eine Satzmenge und $\leq$ die zugehörige Relation der theoretischen Wichtigkeit (in T) ist. Jedenfalls aber meine ich — wie Humphreys, Nowak und Laymon —, daß man wissenschaftliche Theorien als Satzmengen (oder, im Non-Statement view, als Modellklassen) *zusammen mit einer Struktur* auf der Menge aller Sätze (bzw. aller potentiellen Modellen) auffassen sollte, um die Behandlung von Idealisierungen und kontrafaktischen Konditionalsätzen in solchen Theorien möglich zu machen. Von speziellem Interesse ist, daß diese Strukturen nicht gleichwertig mit den Strukturen eines Approximationsprozesses sein müssen (s. Kapitel 8). Die Relation $\leq$ der theoretischen Wichtigkeit eröffnet weiter die Möglichkeit, Lakatos' (1970, S. 155, 164) ausgezeichnete Beobachtung verständlich zu machen, daß sich mathematisch äquivalente Theorien in ihrer „heuristischen Kraft" unterscheiden können. Schon verschiedene Axiomatisierungen derselben Theorie können zu verschiedenen Revisionen führen, wenn wir einem Axiom qua Axiom einen bevorzugten Rang in $\leq$ zuschreiben. Aber diese Gedanken sind, wie gesagt, bisher noch reine Spekulation und warten darauf, mit Substanz gefüllt zu werden.

Literaturverzeichnis

ACHINSTEIN, PETER (1964), 'On the Meaning of Scientific Terms', *Journal of Philosophy* **61**, 497–509.

ADAMS, ERNEST (1959), 'The Foundations of Rigid Body Mechanics and the Derivation of Its Laws from Those of Particle Mechanics', *The Axiomatic Method*, hrsg.v. LEON HENKIN, PATRICK SUPPES und ALFRED TARSKI, North-Holland, Amsterdam, 250–265.

ADAMS, ERNEST W. (1970), 'Subjunctive and Indicative Conditionals', *Foundations of Language* **6**, 89–94.

ALCHOURRÓN, CARLOS E., und DAVID MAKINSON (1982), 'On the Logic of Theory Change: Contraction Functions and Their Associated Revision Functions', *Theoria* **48**, 14–37.

ALCHOURRÓN, CARLOS E., PETER GÄRDENFORS und DAVID MAKINSON (1985), 'On the Logic of Theory Change: Partial Meet Contraction and Revision Functions', *Journal of Symbolic Logic* **50**, 510–530.

ALSONSO, MARCELO, und EDWARD J. FINN (1980), *Fundamental University Physics*, Volume 1: *Mechanics and Thermodynamics*, Addison-Wesley, Reading, Mass.

BAIGRIE, BRIAN S. (1987), 'Kepler's Laws of Planetary Motion, Before and After Newton's *Principia*: an Essay on the Transformation of Scientific Problems', *Studies in History and Philosophy of Science* **18**, 177–208.

BALZER, WOLFGANG (1982), *Empirische Theorien — Modelle, Strukturen, Beispiele*, Vieweg, Braunschweig und Wiesbaden.

BALZER, WOLFGANG (1985), *Theorie und Messung*, Springer, Berlin u.a.

BALZER, WOLFGANG, C. ULISES MOULINES und JOSEPH D. SNEED (1987), *An Architectonic for Science*, Reidel, Dordrecht u.a.

BALZER, WOLFGANG, und JOSEPH D. SNEED (1977), 'Generalized Net Structures of Empirical Theories I', *Studia Logica* **36**, 195–211.

BALZER, WOLFGANG, und JOSEPH D. SNEED (1978), 'Generalized Net Structures of Empirical Theories II', *Studia Logica* **37**, 167–194.

BALZER, WOLFGANG, und JOSEPH D. SNEED (1983), 'Verallgemeinerte Netz-Strukturen empirischer Theorien', *Zur Logik empirischer Theorien*, hrsg.v. WOLFGANG BALZER und MICHAEL HEIDELBERGER, De Gruyter, Berlin, New York, 117–168. (Deutsche Übersetzung von BALZER und SNEED (1978) und (1979))

BARR, WILLIAM F. (1974), 'A Pragmatic Analysis of Idealizations in Physics', *Philosophy of Science* **41**, 48–64.

BLAU, ULRICH (1982/83), *Einführung in die logische Sprachanalyse*, Vor-

lesungsskriptum, Seminar für Philsophie, Logik und Wissenschaftstheorie, Universität München.

BOLTZMANN, LUDWIG (1898), *Vorlesungen über Gastheorie, II. Theil*, in LUDWIG BOLTZMANN, *Gesamtausgabe*, hrsg.v. ROMAN U. SEXL, Band 1, Akademische Druck- und Verlagsanstalt und Vieweg, Graz und Braunschweig und Wiesbaden 1981.

BOLTZMANN, LUDWIG, und J. NABL (1905), 'Kinetische Theorie der Materie', *Enzyklopädie der Mathematischen Wissenschaften*, Band V.1, Teubner, Leipzig 1903–1921, 493–557.

BOOLOS, GEORGE (1979), *The Unprovability of Inconsistency — An Essay in Modal Logic*, Cambridge University Press, Cambridge u.a.

BORN, MAX (1949), *Natural Philosophy of Cause and Chance*, Clarendon Press, Oxford.

BRUSH, STEPHEN G. (1977), 'Statistical Mechanics and the Philosophy of Science: Some Historical Notes,' *PSA 1976*, Vol. 2, hrsg.v. FREDERICK SUPPE und PETER D. ASQUITH, Philosophy of Science Association, East Lansing, Michigan, 551–584.

BRZEZIŃSKI, JERZY, JOLANTA BURBELKA, ANDRZEJ KLAWITER, KRZYSZTOF LASTOWSKI, SLAWOMIR MAGALA und LESZEK NOWAK (1976), 'Law and Theory: A Contribution to the Idealizational Interpretation of the Marxist Methodology', *Poznań Studies in the Philosophy of the Sciences and the Humanities* **2**, No.3, 61–80.

BRZEZIŃSKI, JERZY, FRANCESCO CONIGLIONE, THEO A.F. KUIPERS und LESZEK NOWAK (Hrsg.) (1990), *Idealization*, 2 Bände, *Poznań Studies in the Philosophy of the Sciences and the Humanities* **16-17**, Rodopi, Amsterdam und Atlanta.

CARTWRIGHT, NANCY (1980), 'The Truth Doesn't Explain Much', *American Philosophical Quarterly* **17**, 159–163.

CARTWRIGHT, NANCY (1983), *How the Laws of Physics Lie*, Clarendon Press und Oxford University Press, Oxford und New York.

CHELLAS, BRIAN F. (1980), *Modal Logic — an Introduction*, Cambridge University Press, Cambridge u.a.

CLARK, PETER (1976), 'Atomism Versus Thermodynamics', *Method and Appraisal in the Physical Sciences*, hrsg.v. COLIN HOWSON, Cambridge University Press, Cambridge u.a., 41–105.

CLAUSIUS, RUDOLF (1970), 'Über die Art der Bewegung, welche wir Wärme nennen', *Kinetische Theorie I — Die Natur der Gase und der Wärme*, von STEPHEN G. BRUSH, Vieweg, Braunschweig, 164–193.

COCCHIARELLA, NINO B. (1984), 'Philosophical Perspectives on Quantification in Tense and Modal Logic', *Handbook of Philosophical Logic*, hrsg.v. DOV

GABBAY und F. GUENTHNER, Bd. 2, Dordrecht u.a., 309–353.

COHEN, I. BERNARD (1974b), 'Newton's Theory vs. Kepler's Theory and Galileo's Theory: An Example of a Difference Between a Philosophical and a Historical Analysis of Science', *The Interaction Between Science and Philosophy*, hrsg.v. YEHUDA ELKANA, Humanities Press, Atlantic Highlands, New Jersey, 299–338.

DAVIS, MARTIN (1980), 'The Mathematics of Non-monotonic Reasoning', *Artificial Intelligence* **13**, 73–80.

DAY, MICHAEL A. (1985), 'Adams on Theoretical Reduction', *Erkenntnis* **23**, 161–184.

DOYLE, JON (1979), 'A Truth Maintenance System', *Artificial Intelligence* **12**, 231–272.

DUDMAN, VICTOR H. (1984), 'Conditional Interpretations of if-Sentences', *Australian Journal of Linguistics* **4**, 143–204.

DUHEM, PIERRE (1906/78), *La Théorie Physique: Son Object, Sa Structure*, Chevalier et Rivière, Paris; zitiert nach der deutschen Ausgabe *Ziel und Struktur der physikalischen Theorien*, Meiner, Hamburg 1978.

EBBINGHAUS, HEINZ-DIETER, JÖRG FLUM und WOLFGANG THOMAS (1978), *Einführung in die mathematische Logik*, Wissenschaftliche Buchgesellschaft, Darmstadt.

EBERLE, ROLF A. (1971), 'Replacing One Theory by Another under Preservation of a Given Feature', *Philosophy of Science* **38**, 486–501.

EHLERS, JÜRGEN (1986), 'On Limit Relations between, and Approximative Explanations of, Physical Theories', *Proceedings of the 7th International Congress of Logic, Methodology and Philosophy of Science*, hrsg.v. RUTH BARCAN MARCUS, PAUL WEINGARTNER und GEORG J.W. DORN, North-Holland, Amsterdam, 386–403.

ELLIS, BRIAN (1979), *Rational Belief Systems*, Basil Blackwell, Oxford.

ELSNER, NORBERT (1980), *Grundlagen der Technischen Thermodynamik*, 2. Auflage, Vieweg, Braunschweig und Wiesbaden.

ETHERINGTON, DAVID W. (1987), 'Formalizing Nonmonotonic Reasoning Systems', *Artificial Intelligence* **31**, 41–85.

FEYERABEND, PAUL K. (1958), 'An Attempt at a Realistic Interpretation of Experience', *Proceedings of the Aristotelian Society*, N.S., Vol. **58**, 143–170.

FEYERABEND, PAUL K. (1962), 'Explanation, Reduction, and Empiricism', *Minnesota Studies in the Philosophy of Science*, Vol. III, *Scientific Explanation, Space and Time*, hrsg.v. HERBERT FEIGL und GROVER MAXWELL, University of Minnesota Press, Minneapolis, 28–97.

FEYERABEND, PAUL K. (1963), 'How to Be a Good Empiricist — A Plea for

Tolerance in Matters Epistemological', *Philosophy of Science — The Delaware Seminar*, Vol. 2(1962–63), hrsg.v. BERNARD BAUMRIN, Interscience Publishers, New York, London, Sydney, 3–39.

FEYERABEND, PAUL K. (1965a), 'Problems of Empiricism', *Beyond the Edge of Certainty*, hrsg.v. ROBERT G. COLODNY, Prentice-Hall, Englewood Cliffs, New Jersey, 145–260.

FEYERABEND, PAUL K. (1965b), 'Reply to Criticism — Comments on Smart, Sellars and Putnam', *In Honor of Philipp Frank*, (=*BSPS* **2**), hrsg.v. ROBERT S. COHEN und MARX W. WARTOFSKY, Humanities Press, New York, 223–261.

FEYERABEND, PAUL K. (1965c), 'On the „Meaning" of Scientific Terms', *Journal of Philosophy* **62**, 266–274.

FEYERABEND, PAUL K. (1970), 'Consolations for the Specialist', *Criticism and the Growth of Knowledge*, hrsg.v. IMRE LAKATOS und ALAN MUSGRAVE, Cambridge University Press, Cambridge u.a., 197–230.

FEYERABEND, PAUL K. (1977), 'Changing Patterns of Reconstruction', *British Journal for the Philsophy of Science* **28**, 351–369.

FEYERABEND, PAUL K. (1978), *Der wissenschaftstheoretische Realismus und die Autorität der Wissenschaften*, Ausgewählte Schriften, Band 1, Vieweg, Braunschweig und Wiesbaden.

FEYERABEND, PAUL K. (1981), *Probleme des Empirismus, Schriften zur Theorie der Erklärung, der Quantentheorie und der Wissenschaftsgeschichte*, Ausgewählte Schriften, Band 2, Vieweg, Braunschweig und Wiesbaden.

FEYERABEND, PAUL K. (1983), *Wider den Methodenzwang — Skizze einer anarchistischen Erkenntnistheorie*, revidierte und erweiterte Fassung, Suhrkamp, Frankfurt am Main.

FINE, ARTHUR I. (1967), 'Consistency, Derivability, and Scientific Change', *Journal of Philosophy* **64**, 231–240.

FINE, ARTHUR I. (1975), 'How to Compare Theories: Reference and Change', *Noûs* **9**, 17–32.

FRAUENFELDER, PAUL, und PAUL HUBER (1968), *Einführung in die Physik*, 1. Band, *Mechanik, Hydrodynamik, Thermodynamik*, 3. Auflage, Ernst Reinhardt, Basel.

FUHRMANN, ANDRÉ (1989), 'Reflective Modalities and Theory Change', *Synthese* **81**, 115–134.

GÄRDENFORS, PETER (1979), 'Conditionals and Changes of Belief', *The Logic and Epistemology of Scientific Change*, hrsg.v. ILKKA NIINILUOTO und RAIMO TOUMELA, *Acta Philosophica Fennica* **30** (1978,2–4), North-Holland, Amsterdam, 381–404.

GÄRDENFORS, PETER (1980), 'A Pragmatic Approach to Explanations', *Phi-*

losophy of Science **47**, 404–423.

GÄRDENFORS, PETER (1981), 'An Epistemic Approach to Conditionals', *American Philosophical Quarterly* **18**, 203–211.

GÄRDENFORS, PETER (1982), 'Rules for Rational Changes of Belief', *(320311): Philosophical Essays Dedicated to Lennart Åqvist on His Fiftieth Birthday*, hrsg.v. T. PAULI, University of Uppsala (=University of Uppsala Philosophical Studies no. **34**), Uppsala, 88–101.

GÄRDENFORS, PETER (1984), 'Epistemic Importance and Minimal Changes of Belief', *Australasian Journal of Philosophy* **62**, 136–157.

GÄRDENFORS, PETER (1985), 'Epistemic Importance and the Logic of Theory Change', *Foundations of Logic and Linguistics — Problems and Their Solutions*, hrsg.v. GEORG J.W. DORN und PAUL WEINGARTNER, Plenum Press, New York, London, 345–367.

GÄRDENFORS, PETER (1986), 'Belief Revisions and the Ramsey Test for Conditionals', *Philosophical Review* **95**, 81–93.

GÄRDENFORS, PETER (1987), 'Variations on the Ramsey Test: More Triviality Results', *Studia Logica* **46**, 321–332.

GÄRDENFORS, PETER (1988), *Knowledge in Flux: Modeling the Dynamics of Epistemic States*, Bradford Books, Cambridge/Mass., London.

GÄRDENFORS, PETER, und DAVID MAKINSON (1988), 'Revisions of Knowledge Systems Using Epistemic Entrenchment', *Theoretical Aspects of Reasoning about Knowledge*, ed. MOSHE Y. VARDI, Morgan Kaufmann, Los Altos, California, 83–95.

GARSON, JAMES W. (1984), 'Quantification in Modal Logic', *Handbook of Philosophical Logic*, hrsg.v. DOV GABBAY und F. GUENTHNER, Bd. 2, Dordrecht u.a., 249–307.

GEHRTSEN, CHRISTIAN, und HANS O. KNESER (1969), *Physik*, 10. Auflage, Springer, Berlin u.a.

GELFOND, MICHAEL (1987), 'On the Notion of „Theoremhood" in Autoepistemic Logic', *Abstracts of the 8th International Congress of Logic, Mehtodology and Philosophy of Science*, Band 1, Nauka, Moskau, 231–233.

GIEDYMIN, JERZY (1973), 'Logical Comparability and Conceptual Disparity Between Newtonian and Relativistic Mechanics', *British Journal for the Philosophy of Science* **24**, 270–276.

GINSBERG, MATTHEW L. (1986), 'Counterfactuals', *Artificial Intelligence* **30**, 35–79.

GLYMOUR, CLARK (1970), 'On Some Patterns of Reduction', *Philosophy of Science* **37**, 340–353.

GOODMAN, NELSON (1947), 'The Problem of Counterfactual Conditionals',

Fact, Fiction, and Forecast, von NELSON GOODMAN, Athlone Press, London 54, 13–34; zweite Auflage Bobbs-Merrill, Indianapolis, New York, Kansas City 1965.

GROVE, ADAM (1988), 'Two Modellings for Theory Change', *Journal of Philosophical Logic* **17**, 157–170.

HALPERN, JOSEPH Y., und YORAM MOSES (1984), 'Towards a Theory of Knowledge and Ignorance: Preliminary Report', *Non-Monotonic Reasoning Workshop*, AAAI, New Paltz, N.Y., 125–143.

HARPER, WILLIAM L. (1977), 'Rational Conceptual Change', *PSA 1976*, Vol. 2, hrsg.v. FREDERICK SUPPE und PETER D. ASQUITH, Philosophy of Science Association, East Lansing, Michigan, 462–494.

HEGEL, GEORG W.F. (1970), *Enzyklopädie der philosophischen Wissenschaften im Grundriß (1830), Zweiter Teil: Die Naturphilosophie, mit den mündlichen Zusätzen*, Werke in 20 Bänden, Band 9, Suhrkamp, Frankfurt am Main.

HEMPEL, CARL G. (1965), *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, Free Press und Collier-MacMillan, New York und London.

HEMPEL, CARL G. (1966), *Philosophy of Natural Science*, Prentice Hall, Englewood Cliffs, New Jersey.

HEMPEL, CARL G. (1969), 'Reduction: Ontological and Linguistic Facets', *Philosophy, Science, and Method — Essays in Honor of Ernest Nagel*, hrsg.v. SIDNEY MORGENBESSER, PATRICK SUPPES und MORTON WHITE, St. Martin's Press, New York, 179–199.

HEMPEL, CARL G., und PAUL OPPENHEIM (1948), 'Studies in the Logic of Explanation', *Philosophy of Science* **15**, 135–175. (Wiederabdruck in HEMPEL 1965, 245–290.)

HOERING, WALTER (1984), 'Anomalies of Reduction', *Reduction in Science*, hrsg.v. WOLFGANG BALZER, DAVID PEARCE und HEINZ-JÜRGEN SCHMIDT, Reidel, Dordrecht u.a., 33–50.

HUGHES, GEORGE E., und MAX J. CRESSWELL (1968), *Introduction to Modal Logic*, Methuen, London.

HUMPHREYS, WILLARD C. (1968), *Anomalies and Scientific Theories*, Freeman & Cooper, San Francisco.

JEFFREY, RICHARD C. (1965), *The Logic of Decision*, Chicago University Press, Chicago 1965.

JOHANSSON, INGVAR (1975), *A Critique of Karl Popper's Methodology*, Scandinavian University Books, Stockholm.

JOHANSSON, INGVAR (1980), 'Ceteris Paribus Clauses, Closure Clauses and Falsifiability', *Zeitschrift für allgemeine Wissenschaftstheorie* **11**, 16–22.

JOSEPH, GEOFFREY (1985), Rezension von CARTWRIGHT (1983), *Philosophical Review* **94**, 580–583.

KÄSBAUER, MAX (1976), 'Definitionen der wissenschaftlichen Erklärung', *Erkenntnis* **10**, 255–273.

KAMERLINGH ONNES, H., und W.H. KEESOM (1912), 'Die Zustandsgleichung', *Enzyklopädie der Mathematischen Wissenschaften*, Band V.1, Teubner, Leipzig 1903–1921, 615–945.

KAMLAH, ANDREAS (1985), 'On Reduction of Theories', *Erkenntnis* **22**, 119–142.

KEMENY, JOHN G., und PAUL OPPENHEIM (1956), 'On Reduction', *Philosophical Studies* **7**, 6–19.

KITCHER, PHILIPP (1985), 'Two Approaches to Explanation', *Journal of Philosophy* **82**, 632–639.

KITTEL, CHARLES, WALTER D. KNIGHT, MALVIN A. RUDERMAN, A. CARL HELMHOLTZ und BURTON J. MOYER (1979), *Berkeley Physik Kurs*, Band 1, *Mechanik*, 3. Auflage, Vieweg, Braunschweig und Wiesbaden.

KLEIN, MARTIN J. (1974), 'The Historical Origins of the van der Waals Equation', *Physica* **73**, 28–47.

KOERTGE, NORETTA (1973), 'Theory Change in Science', *Conceptual Change*, hrsg.v. GLENN PEARCE und PATRICK MAYNARD, Reidel, Dordrecht u.a., 167–198.

KONOLIGE, KURT (1988), 'On the Relation Between Default and Autoepistemic Logic', *Artificial Intelligence* **35**, 343–382.

KOURA, ANTTI (1988), 'An Approach to Why-Questions', *Synthese* **74**, 191-206.

KRAJEWSKI, WLADYSLAW (1974), 'The Role of the Correspondence Principle in the Development of Physics', *Dialectics and Humanism* **1**, No.3, 61–63.

KRAJEWSKI, WLADYSLAW (1976), 'Correspondence Principle and the Idealization', *Formal Methods in the Methodology of Empirical Sciences*, hrsg.v. MARIAN PRZELECKI, KLEMENS SZANIAWSKI und RYSZARD WÓJCICKI, Reidel und Ossolineum, Dordrecht, Boston und Wroclaw, 380–386.

KRAJEWSKI, WLADYSLAW (1977), *Correspondence Principle and Growth of Science*, Reidel, Dordrecht u.a.

KRAJEWSKI, WLADYSLAW (1984), 'May We Identify Reduction and Explanation of Theories?', *Reduction in Science — Structure, Examples, Philosophical Problems*, hrsg.v. WOLFGANG BALZER, DAVID PEARCE und HEINZ-JÜRGEN SCHMIDT, Reidel, Dordrecht u.a., 11–15.

KRATZER, ANGELIKA (1981a), 'Partition and Revision: The Semantics of Counterfactuals', *Journal of Philosophical Logic* **10**, 201–216.

KRATZER, ANGELIKA (1981b), 'The Notional Category of Modality', *Words, Worlds, and Contexts*, hrsg.v. HANS-JÜRGEN EIKMEYER und HANNES RIESER, De Gruyter, Berlin, New York, 38–74.

KUHN, THOMAS S. (1970), *The Structure of Scientific Revolutions*, 2. Auflage, Chicago University Press, Chicago.

KUHN, THOMAS S. (1986), 'Possible Worlds in History of Science', Nobel Symposium, August 1986, *unveröff. Manuskript*, 46 S.

KUIPERS, THEO A.F. (1985), 'The Paradigm of Concretization: The Law of van der Waals', *Poznań Studies in the Philosophy of the Sciences and the Humanities* **8**, Rodopi, Amsterdam, 185–199.

KUOKKANEN, MARTTI (1988), 'The Poznań School Methodology of Idealization and Concretization from the Point of View of a Revised Structuralist Theory Conception', *Erkenntnis* **28**, 97–115.

LAKATOS, IMRE (1970), 'Falsification and the Methodology of Scientific Research Programmes', *Criticism and the Growth of Knowledge*, hrsg.v. IMRE LAKATOS und ALAN MUSGRAVE, Cambridge University Press, Cambridge u.a., 91–196.

LAKATOS, IMRE (1977/82a), *The Methodology of Scientific Research Programmes: Philosophical Papers*, Band 1, hrsg.v. JOHN WORRALL und GREGORY CURRIE, Cambridge University Press, Cambridge u.a.; zitiert nach der deutschen Ausgabe *Die Methodologie wissenschaftlicher Forschungsprogramme*, Vieweg, Braunschweig und Wiesbaden 1982.

LAKATOS, IMRE (1977/82b), *Mathematics, Science and Epistemology: Philosophical Papers*, Band 2, hrsg.v. JOHN WORRALL und GREGORY CURRIE, Cambridge University Press, Cambridge u.a.; zitiert nach der deutschen Ausgabe *Mathematik, empirische Wissenschaft und Erkenntnistheorie*, Vieweg, Braunschweig und Wiesbaden 1982.

LAYMON, RONALD (1977), 'Feyerabend, Brownian Motion, and the Hiddenness of Refuting Facts', *Philosophy of Science* **44**, 225–247.

LAYMON, RONALD (1980), 'Idealization, Explanation, and Confirmation', *PSA 1980*, hrsg.v. PETER D. ASQUITH und RONALD N. GIERE, Vol. 1, Philosophy of Science Association, East Lansing, Mich., 336–350.

LAYMON, RONALD (1982), 'Scientific Realism and the Hierarchical Counterfactual Path from Data to Theory', *PSA 1982*, hrsg.v. PETER D. ASQUITH und THOMAS NICKLES, Vol. 2, Philosophy of Science Association, East Lansing, Mich., 107–121.

LAYMON, RONALD (1985), 'Idealizations and Testing of Theories by Experimentation', *Observation, Experiment and Hypothesis in Modern Physical Science*, hrsg.v. PETER ACHINSTEIN und OWEN HANNAWAY, MIT Press, Cambridge, Mass. und London, 147–173.

LEVESQUE, HECTOR J. (1990), 'All I know: A Study in Autoepistemic Logic', *Artificial Intelligence* **42**, 263–309.

LEVI, ISSAC (1977), 'Subjunctives, Dispositions and Chances', *Synthese* **34**, 423–455.

LEVI, ISAAC (1979), 'Serious Possibility', *Essays in Honour of Jaakko Hintikka*, hrsg.v. ESA SAARINEN, RISTO HILPINEN, ILKKA NIINILUOTO und MERRILL PROVENCE HINTIKKA, Reidel, Dordrecht, 219–236.

LEVI, ISAAC (1988), 'Iteration of Conditionals and the Ramsey Test', *Synthese* **76**, 49–81.

LEWIS, DAVID (1973a), *Counterfactuals*, Basil Blackwell, Oxford.

LEWIS, DAVID (1973b), 'Causation', *Journal of Philosophy* **70**, 550–567.

LEWIS, DAVID (1979a), 'Counterfactual Dependence and Time's Arrow', *Noûs* **13**, 455–476.

LEWIS, DAVID (1979b), 'Scorekeeping in a Language Game', *Journal of Philosophical Logic* **8**, 339–359.

LEWIS, DAVID (1986), 'Postscripts to „Counterfactual Dependence and Time's Arrow"', *Philosphical Papers*, von DAVID LEWIS, Band II, Oxford University Press, New York und Oxford, 52–66.

LUDWIG, GÜNTHER (1978), *Die Grundstrukturen einer physikalischen Theorie*, Springer, Berlin, Heidelberg, New York.

ŁUKASZEWICZ, WITOLD (1986), 'Formalization of Knowledge and Ignorance: Introduction to Non-monotonic Reasoning', *CC-AI* **3**, 3–31.

LYON, ARDON (1976–77), 'On Immutable Laws of Nature', *Proceedings of the Aristotelian Society* **77**, 107–126.

MACKIE, JOHN L. (1962), 'Counterfactuals and Causal Laws', *Analytical Philosophy*, hrsg.v. R.J. BUTLER, Blackwell, Oxford, 66–80.

MAKINSON, DAVID (1985), 'How to Give It Up: A Survey of Some Formal Aspects of the Logic of Theory Change', *Synthese* **62**, 347–363. (Errata dazu in *Synthese* **68** (1986), 185–186.)

MAKINSON, DAVID (1987), 'On the Status of the Postulate of Recovery in the Logic of Theory Change', *Journal of Philosophical Logic* **16**, 383–394.

MAKINSON, DAVID (1989a), 'General Theory of Cumulative Inference', *Non-Monotonic Reasoning*, hrsg.v. MICHAEL REINFRANK, JOHAN DE KLEER, MATTHEW GINSBERG und ERIK SANDEWALL, Springer, Berlin u.a., 1–18.

MAKINSON, DAVID (1989b), 'The Gärdenfors Impossibility Theorem in Non-Monotonic Contexts', *Studia Logica*, erscheint.

MAREK, WIKTOR, und MIROSLAW TRUSZYŃSKI (1989), 'Relating Autoepistemic and Default Logic', *Principles of Knowledge Representation and Rea-*

soning, hrsg.v. R.J. BRACHMAN, H.J. LEVESQUE und R. REITER, Morgan Kaufmann, San Mateo, 276–288.

MAYR, DIETER (1976), 'Investigations of the Concept of Reduction I', *Erkenntnis* **10**, 275–294.

MAYR, DIETER (1981a), 'Investigations of the Concept of Reduction II', *Erkenntnis* **16**, 109–129.

MAYR, DIETER (1981b), 'Approximative Reduction by Completion of Empirical Uniformities', *Structure and Approximation in Physical Theories*, hrsg.v. A. HARTKÄMPER und H.-J. SCHMIDT, Plemum Press, New York, London, 55–70.

McCALL, STORRS (1983), 'If, Since and Because: A Study in Conditional Connection', *Logique et Analyse* **26**, 309–321.

McCALL, STORRS (1984), 'Counterfactuals Based on Real Possible Worlds', *Noûs* **18**, 463–477.

McCALL, STORRS (1987), *Tree-Semantics for Conditionals Based on Connection*, McGill University, Montreal, Second Draft, June 1987.

McCARTHY, JOHN (1980), 'Circumscription — a Form of Non-monotonic Reasoning', *Artificial Intelligence* **13**, 27–39.

McDERMOTT, DREW (1982), 'Nonmonotonic Logic II: Nonmonotonic Modal Theories', *Journal of the Association for Computing Machinery* **29**, 33–57.

McDERMOTT, DREW, & JON DOYLE (1980), 'Non-Monotonic Logic I', *Artificial Intelligence* **13**, 41–72.

McGRAW-HILL ENCYCLOPEDIA OF SCIENCE AND TECHNOLOGY (1987), 20 Bände, 6. Auflage, McGraw-Hill, New York u.a.

McMULLIN, ERNAN (1985), 'Galilean Idealization', *Studies in History and Philosophy of Science* **16**, 247–273.

MINSKY, M. (1974), 'A Framework for Representing Knowledge', *AIM-306*, MIT Artificial Intelligence Laboratory, Cambridge/MA. (Teilweise auch in *The Psychology of Computer Vision*, hrsg.v. P. WINSTON, McGraw-Hill, New York 1975; in *Mind Design: Philosophy, Psychology and Artificial Intelligence*, hrsg.v. J. HAUGELAND, Bradford Books, Montgomery, Vermont; und in *Readings in Knowledge Representation*, hrsg.v. R.J. BRACHMANN und H.J. LEVESQUE, Morgan Kaufmann, Los Altos 1985.)

MOORE, ROBERT C. (1984), 'Possible-World Semantics for Autoepistemic Logic', *Non-Monotonic Reasoning Workshop*, New Paltz/NY, 344–354. (Auch als als *Technical Note* 337, August 1984, Menlo Park/CA.)

MOORE, ROBERT C. (1985), 'Semantical Considerations on Nonmonotonic Logic', *Artificial Intelligence* **25**, 75–94.

MORMANN, THOMAS (1984), 'Strukturalistische Reduktionsbeziehungen als

Morphismen von Constraintkategorien', *unveröffentlichtes Manuskript*, Universität Bielefeld.

MOSER, JÜRGEN (1977-79), 'Is the Solar System Stable?', *Mathematical Intelligencer* **1**, 65–71.

MOULINES, C. ULISES (1976), 'Approximative Application of Empirical Theories: A General Explication', *Erkenntnis* **10**, 201–227.

MOULINES, C. ULISES (1980), 'Intertheoretic Approximation: The Kepler-Newton Case', *Synthese* **45**, 387–412.

MOULINES, C. ULISES (1981), 'A General Scheme for Intertheoretic Approximation', *Structure and Approximation in Physical Theories*, hrsg.v. A. HARTKÄMPER und HEINZ-JÜRGEN SCHMIDT, Plenum Press, New York und London, 123–146.

MÜHLHÖLZER, FELIX (1988), 'On Objectivity', *Erkenntnis* **28**, 185–230.

NAGEL, ERNEST (1949), 'The Meaning of Reduction in the Natural Sciences', *Science and Civilization*, hrsg.v. ROBERT C. STAUFFER, University of Wisconsin Press, Madison, 99–145. (Zitiert nach dem Wiederabdruck in *Philosophy of Science*, hrsg.v. ARTHUR DANTO und SIDNEY MORGENBESSER, Meridian Books, New York 1960, 288–312.)

NAGEL, ERNEST (1951), 'Mechanistic Explanation and Organismic Biology', *Philosophy and Phenomenological Research* **11**, 327–338.

NAGEL, ERNEST (1961), *The Stucture of Science — Problems in the Logic of Scientific Explanation*, Harcourt, Brace & World, New York, Chicago und Burlingame.

NAGEL, ERNEST (1970), 'Issues in the Logic of Reductive Explanations', *Contemporary Philosophic Thought*, hrsg.v. MILTON K. MUNITZ, Band 2: *Mind, Science and History*, Albany, New York, 117–137.

NERNST, WALTHER (1922), 'Zum Gültigkeitsbereich der Naturgesetze', *Naturwissenschaften* **10**, 489–495.

NEWTON, SIR ISAAC (1963), *Mathematische Prinzipien der Naturlehre*, hrsg.v. JACOB P. WOLFERS, Wissenschaftliche Buchgesellschaft, Darmstadt; am. Ausgabe *Sir Isaac Newton's Mathematical Principles of Natural Philosophy and His System of the World*, 1729 übersetzt von ANDREW MOTTE, Übersetzung revidiert von FLORIAN CAJORI, University of California Press, Berkeley 1934.

NICKLES, THOMAS (1973), 'Two Concepts of Intertheoretic Reduction', *Journal of Philosophy* **70**, 181–201.

NICKLES, THOMAS (1974), 'Heuristics and Justification in Scientific Research: Comments on Shapere', *The Structure of Scientific Theories*, hrsg.v. FREDERICK SUPPE, University of Illinois Press, Urbana, Chicago und London, 571–589.

NIINILUOTO, ILKKA (1980), 'The Growth of Theories: Comments on the Struc-

turalist Approach', *Theory Change, Ancient Axiomatics, and Galileo's Method*, hrsg.v. JAAKKO HINTIKKA, DANIEL GRUENDER und EVANDRO AGAZZI, Reidel, Dordrecht, 3–47.

NIINILUOTO, ILKKA (1986), 'Theories, Approximations, and Idealizations', *Logic, Methodology and Philosophy of Science VII*, hrsg.v. RUTH BARCAN MARCUS, GEORG J.W. DORN und PAUL WEINGARTNER, North-Holland, Amsterdam u.a., 255–289.

NOWAK, LESZEK (1972), 'Idealizational Laws and Explanation', *Logique et Analyse* **15** (H.59/60), 527–545.

NOWAK, LESZEK (1974), 'Of Some Modifications of the Concept of Dialectical Correspondence, *Dialectics and Humanism* **1**, No.3, 73–78.

NOWAK, LESZEK (1975), 'Idealization: A Reconstruction of Marx's Ideas', *Poznań Studies in the Philosophy of the Sciences and the Humanities* **1**, No.1, 65–70.

NOWAK, LESZEK (1980), *The Structure of Idealization: Towards a Systematic Interpretation of the Marxian Idea of Science*, Reidel, Dordrecht u.a.

NOWAKOWA, IZABELLA (1974), 'The Concept of Dialectical Correspondence, *Dialectics and Humanism* **1**, No.3, 51–55.

NOWAKOWA, IZABELLA (1975), 'Idealization and the Problem of Correspondence', *Poznań Studies in the Philosophy of the Sciences and the Humanities* **1**, No.1, 65–70.

PARTINGTON, J.R. (1949), *An Advanced Treatise on Physical Chemistry*, Volume 1: *Fundamental Principles, the Properties of Gases*, Longmans, Green and Co, London u.a.

PATRYAS, WOJCIECH (1975), 'An Analysis of the „Caeteris Paribus" Clause', *Poznań Studies in the Philosophy of the Sciences and the Humanities* **1**, No.1, 59–64.

PATRYAS, WOJCIECH (1977), 'The Sense of Empirical Testing', *Poznań Studies in the Philosophy of the Sciences and the Humanities* **3**, 180–198.

PEARCE, DAVID (1981), 'Is There Any Justification for a Nonstatement View of Theories?', *Synthese* **46**, 1–39.

PEARCE, DAVID (1982), 'Logical Properties of the Structuralist Concept of Reduction', *Erkenntnis* **18**, 307–333.

PEARCE, DAVID (1985), 'Remarks on Physicalism and Reductionism', *Action, Logic and Social Theory (= Acta Philosophica Fennica* **38**), hrsg.v. G. HOLMSTROEM und A. JONES, Eripainos, Helsinki, 247–271.

PEARCE, DAVID (1987), *Roads to Commensurability*, Reidel, Dordrecht u.a.

PEARCE, DAVID, und VEIKKO RANTALA (1983a), 'New Foundations for Metascience', *Synthese* **56**, 1–26.

PEARCE, DAVID, und VEIKKO RANTALA (1983b), 'Correspondence as an Intertheory Relation', *Studia Logica* **42**, 363–371.

PEARCE, DAVID, und VEIKKO RANTALA (1984a), 'A Logical Study of the Correspondence Relation', *Journal of Philosophical Logic* **13**, *47–84.*

PEARCE, DAVID, und VEIKKO RANTALA (1984b), 'Limiting Case Correspondence between Physical Theories', Reduction in Science, hrsg.v. WOLFGANG BALZER, DAVID PEARCE und HEINZ-JÜRGEN SCHMIDT, Reidel, Dordrecht u.a., 153–185.

POPPER, KARL R. (1934), *Logik der Forschung*, Julius Springer, Wien.

POPPER, KARL R. (1957), 'Über die Zielsetzung der Erfahrungswissenschaft', *Ratio* **1**, 21–31.

POPPER, KARL R. (1972), *Objective Knowledge — An Evolutionary Approach*, Clarendon Press, Oxford.

PUTNAM, HILARY (1965), 'How Not to Talk about Meaning — Comments on J.J.C. Smart', *In Honor of Philipp Frank*, (=*BSPS* **2**), hrsg.v. ROBERT S. COHEN und MARX W. WARTOFSKY, Humanities Press, New York, 205–222.

PUTNAM, HILARY (1974), 'The „Corroboration" of Theories', *The Philosophy of Karl Popper*, hrsg.v. PAUL A. SCHILPP, Book I, Open Court, La Salle, Ill., 221–240.

RAMSEY, FRANK P. (1931), 'General Propositions and Causality', *The Foundations of Mathematics and Other Essays*, von FRANK P. RAMSEY, Kegan Paul u.a., London, 237–255.

RANTALA, VEIKKO (1987), 'Explaining Superseded Laws', *unveröffentlichtes Manuskript*, Helsinki, 11 S.

RANTALA, VEIKKO (1988), 'Counterfactual Reduction', *Criticism and the Growth of Knowledge — After 20 Years*, hrsg.v. KOSTAS GAVROGLU, YORGOS GOUDAROULIS u.a., Reidel, Dordrecht u.a.

REICHL, L.E. (1980), *A Modern Course in Statistical Physics*, University of Texas Press, Austin.

REITER, RAYMOND (1980), 'A Logic for Default Reasoning', *Artificial Intelligence* **13**, 81–132.

ROTT, HANS (1984), *Epistemische Interpretationen von* wenn-dann- *und* weil-*Sätzen nach Kratzer, Gärdenfors und Spohn*, Magisterarbeit, München.

ROTT, HANS (1990), 'A Nonmonotonic Conditional Logic for Belief Revision', *On the Logic of Theory Change*, hrsg.v. André Fuhrmann und Michael Morreau, Springer, Berlin u.a., erscheint.

RYLE, GILBERT (1963), '„If," „So," and „Because"', *Philosophical Analysis*, hrsg.v. MAX BLACK, 2. Auflage, Prentice-Hall, Englewood Cliffs, N.J., 302–318.

SCHAEFER, CLEMENS (1958), *Einführung in die theoretische Physik*, 2. Band, *Theorie der Wärme, Molekular-kinetische Theorie der Materie*, 3. Auflage, de Gruyter, Berlin.

SCHAFFNER, KENNETH F. (1967), 'Approaches to Reduction', *Philosophy of Science* **34**, 137–147.

SCHEIBE, ERHARD (1973a), 'Die Erklärung der Keplerschen Gesetze durch Newtons Gravitationsgesetz', *Einheit und Vielheit — Festschrift für Carl Friedrich von Weizsäcker*, hrsg.v. ERHARD SCHEIBE und GEORG SÜSSMANN, Vandenhoeck & Rupprecht, Göttingen, 98–118.

SCHEIBE, ERHARD (1973b), 'The Approximative Explanation and the Development of Physics', *Logic, Methodology and Philosophy of Science IV*, hrsg.v. PATRICK SUPPES, LEON HENKIN, ATHANASE JOJA und GR.C. MOISIL, North-Holland, Amsterdam, London, 931–942.

SCHEIBE, ERHARD (1975), 'Vergleichbarkeit, Widerspruch und Erklärung', *Physik und Philosophie*, hrsg.v. RUDOLF HALLER und JOHANN GÖTSCHL, Vieweg, Braunschweig, 57–71.

SCHEIBE, ERHARD (1976), 'Gibt es Erklärungen von Theorien?', *Allgemeine Zeitschrift für Philosophie***1**, *26–45.*

SCHEIBE, ERHARD (1982), 'Zum Theorienvergleich in der Physik', Physik, Philosophie und Politik — Festschrift für Carl F. von Weizsäcker zum 70. Geburtstag, hrsg.v. KLAUS M. MEYER-ABICH, Hanser, München und Wien, 291–309.

SCHEIBE, ERHARD (1983), 'Two Types of Successor Relations between Theories', *Zeitschrift für allgemeine Wissenschaftstheorie* **14**, 68-80.

SCHEIBE, ERHARD (1984), 'Explanation of Theories and the Problem of Progress in Physics', *Reduction in Science*, hrsg.v. WOLFGANG BALZER, DAVID A. PEARCE und HEINZ-JÜRGEN SCHMIDT, Reidel, Dordrecht u.a., 71–94.

SCHEIBE, ERHARD (1988), 'The Physicist's Conception of Progress', *Studies in History and Philosophy of Science* **19**, 141–159.

SCHILLING, HEINZ (1972), *Statistische Physik in Beispielen*, VEB Fachbuchverlag, Leipzig.

SCHMIDT, ERNST, KARL STEPHAN und FRANZ MAYINGER (1975), *Technische Thermodynamik*, Band 1, *Einstoffsysteme*, 11. Auflage, Springer, Berlin u.a.

SCHROEDER-HEISTER, PETER, und FRANK SCHAEFER (1989), 'Reduction, Representation and Commensurability of Theories', *Philosophy of Science* **56**, 130–157.

SCHWARTZ, ROBERT J. (1978), 'Idealizations and Approximations in Physics', *Philosophy of Science* **45**, 595–603.

SCRIVEN, MICHAEL (1961), 'The Key Property of Physical Laws — Inaccuracy', *Current Issues in the Philosophy of Science*, hrsg.v. HERBERT FEIGL und GROVER MAXWELL, Holt, Rinehart, and Winston, New York, 91–101.

SCRIVEN, MICHAEL (1962), 'Explanations, Predictions, and Laws', *Minnesota Studies in the Philosophy of Science*, Vol. III, *Scientific Explanation, Space, and Time*, hrsg.v. HERBERT FEIGL und GROVER MAXWELL, University of Minnesota Press, Minneapolis, 170–230.

SCRIVEN, MICHAEL (1963), 'The Limits of Physical Explanation', *Philosophy of Science*, Vol. 2 (1962–63), hrsg.v. BERNARD BAUMRIN, Interscience Publishers, New York, London und Sydney, 107–135.

SEARS, FRANCIS W., und MARK W. ZEMANSKY (1964), *University Physics*, 3. Auflage, Addison-Wesley, Reading, Mass.

SEGERBERG, KRISTER (1982), *Classical Propositional Operators: An Exercise in the Foundations of Logic*, Oxford: Clarendon Press.

SHAPERE, DUDLEY (1966), 'Meaning and Scientific Change', *Mind and Cosmos*, hrsg.v. ROBERT G. COLODNY, University of Pittsburgh Press, Pittsburgh, 41–85.

SHAPERE, DUDLEY (1974), 'Scientific Theories and Their Domains', *The Structure of Scientific Theories*, hrsg.v. FREDERICK SUPPE, University of Illinois Press, Urbana u.a., 518–565.

SHEA, WILLIAM (1981), 'The Young Hegel's Quest for a Philosophy of Science, or Pitting Kepler Against Newton', *Scientific Philosophy Today*, hrsg.v. JOSEPH AGASSI und ROBERT S. COHEN, Reidel, Dordrecht u.a., 381–397.

SINTONEN, MATTI (1984), 'On the Logic of Why-Questions', *PSA 1984*, Volume 1, 168–179.

SKLAR, LAWRENCE (1964), *Inter-theoretical Reduction in Natural Science*, Dissertation, Princeton University.

SKLAR, LAWRENCE (1967), 'Types of Inter-Theoretic Reduction', *British Journal for the Philosophy of Science* **18**, 109–124.

SMART, J.J.C. (1965), 'Conflicting Views about Explanation', *In Honor of Philipp Frank*, (=*BSPS* **2**), hrsg.v. ROBERT S. COHEN und MARX W. WARTOFSKY, Humanities Press, New York, 157–169.

SMORYNSKI, CRAIG (1984), 'Modal Logic and Self-Reference', *Handbook of Philosophical Logic*, hrsg.v. DOV M. GABBAY und FRANZ GUENTHNER, Vol. II, *Extensions of Classical Logic*, Reidel, Dordrecht u.a, 441-495.

SNEED, JOSEPH D. (1971), *The Logical Structure of Mathematical Physics*, Reidel, Dordrecht.

SNEED, JOSEPH D. (1976), 'Philosophical Problems in the Empirical Science: A Formal Approach', *Erkenntnis* **10**, 115–146.

SOMMERFELD, ARNOLD (1952), *Vorlesungen über theoretische Physik*, Band V, *Thermodynamik und Statistik*, Dieterich'sche Verlagsbuchhandlung, Wiesbaden.

SPOHN, WOLFGANG (1983), 'Deterministic and Probabilistic Reasons and Causes', *Erkenntnis* **19**, 371–396.

SPOHN, WOLFGANG (1988a), 'Ordinal Conditional Functions', *Causation in Decision, Belief Change, and Statistics*, hrsg.v. WILLIAM L. HARPER und BRIAN SKYRMS, Bd. II, Kluwer, Dordrecht u.a., 105–134.

SPOHN, WOLFGANG (1988b), 'A General Non-Probabilistic Theory of Inductive Reasoning', *Proceedings of the 4th Workshop on Uncertainty in Artificial Intelligence*, University of Minnesota, Minneapolis, 315–322.

STALNAKER, ROBERT (1968), 'A Theory of Conditionals', *Studies in Logical Theory*, hrsg.v. NICHOLAS RESCHER, APQ Monograph Series **2**, Blackwell, Oxford, 98–112.

STALNAKER, ROBERT (1980), 'A Note on Non-monotonic Modal Logic', *Unpublished Ms.*, Department of Philosophy, Cornell University, Ithaca/NY.

STALNAKER, ROBERT C. (1981), 'A Defense of Conditional Excluded Middle', *Ifs: Conditionals, Belief, Decision, Chance and Time*, hrsg.v. WILLIAM L. HARPER, ROBERT C. STALNAKER und GLENN PEARCE, Reidel, Dordrecht, 87–104.

STALNAKER, ROBERT C. (1984), *Inquiry*, Bradford Books, Cambridge, London.

STEGMÜLLER, WOLFGANG (1979), *The Structuralist View of Theories — A Possible Analogue of the Bourbaki Programme in Physical Science*, Springer, Berlin, Heidelberg und New York.

STEGMÜLLER, WOLFGANG (1983), *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie*, Band I, *Erklärung — Begründung — Kausalität*, 2. Auflage, Springer, Berlin.

STEGMÜLLER, WOLFGANG (1985), *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie*, Band II, *Theorie und Erfahrung*, 2. Teilband, *Theorienstrukturen und Theoriendynamik*, 2. Auflage, Springer, Berlin u.a.

STEGMÜLLER, WOLFGANG (1986), *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie*, Band II, *Theorie und Erfahrung*, 3. Teilband, *Die Entwicklung des neuen Strukturalismus seit 1973*, Springer, Berlin u.a.

STRAWSON, PETER F. (1952), *Introduction to Logical Theory*, Methuen und Wiley, London und New York.

STUMPFF, KARL (1973), *Himmelsmechanik*, Vol. I, 2. Auflage, VEB Deutscher Verlag der Wissenschaften, Berlin.

SUCH, JAN (1978), 'Idealization and Concretization in Natural Sciences', *Poznań Studies in the Philosophy of the Sciences and the Humanities* **4**, 49–73.

SUPPE, FREDERICK (1974a), 'The Search for Philosophic Understanding of Scientific Theories', *The Structure of Scientific Theories*, hrsg.v. FREDERICK SUPPE, University of Illinois Press, Urbana u.a., 1–241.

SUPPE, FREDERICK (1974b), 'Theories and Phenomena', *Developments in the Methodology of Social Science*, hrsg.v. WERNER LEINFELLNER und ECKEHART KÖHLER, Reidel, Dordrecht u.a., 45–91.

SUPPE, FREDERICK (1976), 'Theoretical Laws', *Formal Methods in the Methodology of Empirical Sciences*, hrsg.v. MARIAN PRZELECKI, KLEMENS SZANIAWSKI und RYSZARD WÓJCICKI, Reidel und Ossolineum, Dordrecht, Boston und Wroclaw, 247–267.

SUPPE, FREDERICK (1977), 'Afterword — 1977', *The Stucture of Scientific Theories*, hrsg.v. FREDERICK SUPPE, 2. Auflage, University of Illinois Press, Urbana, Chicago und London, 615–730.

SUPPES, PATRICK (1957), *Introduction to Logic*, Van Nostrand, Princeton.

TARSKI, ALFRED (1930), 'Über einige fundamentale Begriffe der Metamathematik', *Comptes Rendus des séances de la Société des Sciences et des Lettres de Varsovie* **23**, Cl. III, 22–29.

TEMPLE, DENNIS (1988), 'Discussion: The Contrast Theory of Why-Questions', *Philosophy of Science***55**, *141–151.*

THEIMER, WALTER (1978), Handbuch naturwissenschaftlicher Grundbegriffe, dtv, München 1978.

THOMSON, A.J., und A.V. MARTINET (1980), *A Practical English Grammar*, 3. Auflage, Oxford University Press, Oxford.

TOEPLITZ, OTTO (1949), *Die Entwicklung der Infinitesimalrechnung*, 1. Band, hrsg.v. GOTTFRIED KÖHLER, Springer, Berlin, Göttingen, Heidelberg.

TOULMIN, STEPHEN (1961), *Foresight and Understanding*, Hutchinson, London.

TUOMELA, RAIMO (1978), 'On the Structuralist Approach to the Dynamics of Theories', *Synthese* **39**, 211–231.

TOURETZKY, DAVID S. (1986), *The Mathematics of Inheritance Systems*, Pitman und Morgan Kaufmann, London und Los Altos/CA.

TURNER, RAYMOND (1984), 'Towards a Semantic Theory of Non-Monotonic Inference', Kapitel 5 in *Logics for Artificial Intelligence*, von RAYMOND TURNER, Ellis Horwood, Chicester u.a., 59–76.

VAN DER WAALS, JOHANNES D. (1911), *Die Zustandsgleichung — Rede, gehalten am 12. Dez. 1910 in Stockholm bei Empfang des Nobelpreises für Physik*, Akademische Verlagsgesellschaft, Leipzig.

VAN FRAASSEN, BAS C. (1980), *The Scientific Image*, Clarendon Press, Oxford.

VAN FRAASSEN, BAS C. (1981), 'Essences and Laws of Nature', *Reduction, Time and Causality*, hrsg.v. RICHARD HEALEY, Cambridge University Press, Cambridge u.a., 189–200.

WILSON, CURTIS A. (1969/70), 'From Kepler's Laws, So-called, to Universal Gravitation — Empirical Factors', *Archive for History of Exact Sciences* **6**, 89–170.

WÓJCICKI, RYSZARD (1988), *Theory of Logical Calculi*, Kluwer, Dordrecht u.a.

YOSHIDA, RONALD M. (1977), *Reduction in the Physical Sciences*, Dalhousie University Press, Halifax, Nova Scotia.

YOSHIDA, RONALD M. (1978), Rezension von KRAJEWSKI (1977), *British Journal for the Philosophy of Science* **29**, 285–289.