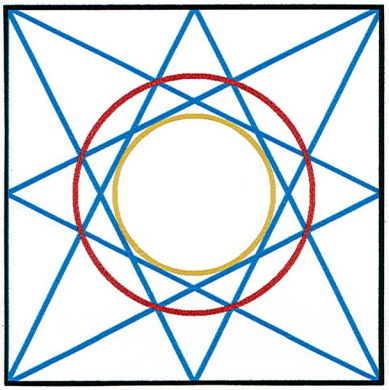
Wilhelm K. Essler: Induktive Logik



Wilhelm K. Essler

Induktive Logik

Grundlagen und Voraussetzungen

EX LIBRIS
W. K. ESSLER

Verlag Karl Alber Freiburg/München

Als Habilitationsschrift auf Empfehlung der Philosophischen Fakultät
der Ludwig-Maximilians-Universität München
gedruckt mit Unterstützung der Deutschen Forschungsgemeinschaft

Alle Rechte vorbehalten — Printed in Germany
© Verlag Karl Alber GmbH Freiburg/München 1970
Herstellung: Union Druckerei GmbH Stuttgart
ISBN 3 495 47196 0

© Wilhelm K. Essler & Verlag Karl Alber
Frankfurt am Main, 2-te unveränderte
Auflage 2010

Therefore, we must class Carnap's contribution to the problem of induction among the greatest achievements of modern Philosophy.

John G. Kemeny in [9], S. 737

Inhalt

Vorwort 9

I. Die geschichtliche Entwicklung des Induktionsbegriffs 13

1. Der Induktionsbegriff der antiken Philosophie 13
2. Enumerative und eliminative Induktion 26
3. Über die Begründung des induktiven Schließens 36
4. Die mathematischen Wahrscheinlichkeitsbegriffe 57

II. Analyse des logischen Wahrscheinlichkeitsbegriffs 65

1. Subjektive, objektive und logische Wahrscheinlichkeitsbegriffe 65
2. Die Form des logischen Wahrscheinlichkeitsbegriffs 76
3. Qualitation, komparative und quantitative Wahrscheinlichkeit 81
4. Über Möglichkeit und Grenzen der induktiven Logik 83

III. Logische Grundlagen der Wahrscheinlichkeit 89

1. Reguläre Bestätigungsmethoden 89
2. Reguläre Wahrscheinlichkeitsfunktionen 100
3. Strenge Kohärenz 118
4. Die logische Stärke der Wahrscheinlichkeitsaxiome 135

IV. Das Kontinuum der induktiven Methoden 139

1. Adäquate induktive Methoden 139
2. Die Normalform der adäquaten induktiven Methoden 148
3. Weitere Adäquatheitsbedingungen für induktive Methoden . 157
4. Erkenntnistheoretische Konsequenzen aus dieser Normalform 173

V. Das System der Adäquatheitsbedingungen 191

1. Die Widerspruchsfreiheit und Vollständigkeit der Adäquatheitsbedingungen 191
2. Der Gehalt der Adäquatheitsbedingungen 201
3. Die Rechtfertigung der Adäquatheitsbedingungen 205
4. Die Wahrscheinlichkeit genereller Hypothesen 217

VI. Induktive Logik und Wahrscheinlichkeitslehre 225

1. Logische und objektive Wahrscheinlichkeit 225
2. Über statistische Verteilungen 245
3. Die mathematische Wahrscheinlichkeitslehre 263
4. Über den mathematischen und den logischen Ansatz 274

VII. Über die Voraussetzungen der Induktion 291

1. Das Problem der zerrütteten Prädikate 291
2. Auf dem Wege zur Lösung dieses Problems 302
3. Das Problem der Dispositionsprädikate 316
4. Über die Begründung von Theorien 337

Anhang: Der Erfolg mit adäquaten induktiven Methoden 553

Literaturverzeichnis 369

Vorwort

Man sagt nicht ganz zu Unrecht, daß ein Wissenszweig um so mehr genötigt ist, sich mit seiner eigenen Geschichte zu befassen, je weniger er systematisch entwickelt worden ist. So wird man etwa in mathematischen Lehrbüchern kaum Andeutungen der historischen Entwicklung dieser Disziplin finden, während der Gegenstand etwa bestimmter soziologischer und auch philosophischer Richtungen weitgehend ihre eigene Geschichte zu sein scheint.

Trotzdem beginne ich dieses Buch mit einem historischen Kapitel, und zwar aus einer zweifachen Überlegung heraus. Einmal möchte ich damit zeigen, um wieviel klarer man das Induktionsproblem mit den begrifflichen Hilfsmitteln der modernen Wissenschaftstheorie als mit denen der traditionellen Philosophie behandeln kann, sei es in der *Formulierung* der Probleme, sei es in ihrer *Lösung*. Vor allem aber soll durch diese Untersuchung nachgewiesen werden, daß der Begriff des induktiven Arguments, des logischen Wahrscheinlichkeitsschlusses, so wie er in diesem Buch analysiert wird, ausdrücklich oder versteckt von allen, die sich mit dem Induktionsproblem befaßt haben, benützt und vorausgesetzt worden ist. Dieses Ergebnis bildet die Grundlage für die im II. Kapitel vorgenommene Explikation des Begriffs des induktiven Schlusses als partieller bzw. gradueller logischer Implikation.

Im I. Kapitel habe ich jene Beiträge zum Induktionsproblem zusammengestellt, die mir aus historischen oder systematischen Gründen als wichtig erscheinen; selbstverständlich kann dieses Kapitel nicht den Anspruch erheben, eine vollständige Geschichte der induktiven Logik zu bieten. Ich habe hier versucht, eine Deutung des Induktionsbegriffs des Aristoteles unter Berücksichtigung seiner recht vereinzelten und verstreuten Bemerkungen zum Induktionsproblem aus meiner Sicht zu geben; dabei bin ich mir des hypothetischen Charakters dieser Interpretation wohl bewußt. Weniger hypothetisch dürfte die Dar-

stellung der Theorie des induktiven Schließens der Epikureer sein, die im Grunde nur eine geschlossenere Wiedergabe des leider bis heute kaum beachteten Textes des Philodemus ist. Zum Zwecke eines besseren Verständnisses habe ich auch die Rechtfertigung der Schlüsse der induktiven Logik, so wie sie von Mill entwickelt worden ist, in umgekehrter Reihenfolge entwickelt als es der Autor getan hat; ich hoffe, daß damit die Mißverständnisse, die sich durch alle deutschen Philosophiegeschichten ziehen (soweit sie Mill betreffen), beseitigt werden können. Etwas abrupt scheint sich der Kreis zu schließen, wenn ich noch den mathematischen Wahrscheinlichkeitsbegriff darstelle. Indessen gehört auch die mathematische Wahrscheinlichkeitslehre zur Geschichte der induktiven Logik, und die Frage, inwieweit etwa die Laplacesche Sukzessionsregel zur Berechnung von Wahrscheinlichkeiten Carnap zur Definition seiner Bestätigungsmethode c^* inspiriert hat oder hätte inspirieren können, wird in diesem Buch mit genügender Ausführlichkeit beantwortet und zwar so, daß der Zusammenhang des philosophischen und des mathematischen Ansatzes klar zutagetritt.

Im III. Kapitel werden drei notwendige Bedingungen dafür angegeben, daß eine geeignete Funktion c als induktive Methode angesehen werden kann; es wird gezeigt, daß sie miteinander logisch gleichwertig sind und daß sie den logischen Wahrscheinlichkeitsbegriff nicht hinreichend charakterisieren. Im IV. Kapitel wird eine Gesamtheit von Bedingungen angegeben, die eine zwar unendliche, aber in gewisser Hinsicht überschaubare Menge von Bestätigungsmethoden charakterisieren; es wird gezeigt, daß die Auswahl einer bestimmten von ihnen mit gewissen Aprioriannahmen über das zugrundegelegte Universum verbunden ist. Daran anschließend wird die Frage untersucht, in welchem Sinn man jene Bedingungen als für den logischen Wahrscheinlichkeitsbegriff adäquat ansehen kann und unter welchen Voraussetzungen man behaupten darf, daß sie nicht als Aussagen über die Natur (über den Gegenstandsbereich und seine Eigenschaften) anzusehen sind.

Im VI. Kapitel wird untersucht, in welcher Hinsicht sich induktive Logik und mathematische Wahrscheinlichkeitslehre unterscheiden und in welcher Hinsicht sie als miteinander gleichwertig anzusehen sind, und es wird in diesem Zusammenhang angedeutet, wie der Be-

griff des induktiven Schließens für ausdrucksreichere Sprachen zu charakterisieren ist. Schließlich wird im Zusammenhang mit einem von Goodman aufgeworfenen Problem nochmals auf die Frage der Voraussetzungen des induktiven Schließens eingegangen, und es wird angedeutet, wie mit einem adäquaten System der induktiven Logik auch andere Probleme, wie etwa das der Dispositionsprädikate, zu lösen sind.

Das historische Kapitel dieses Buches kann nur beanspruchen, eine grobe Skizze der vielschichtigen und schwer zu durchdringenden Entwicklung des Induktionsbegriffs zu sein. Ich kann aber nicht leugnen, daß auch der systematische Teil unvollständig ist und vieles nicht enthält, was heute schon zum gesicherten Bestand der induktiven Logik gehört. Genauso, wie es heute bereits unmöglich ist, ein Buch zu schreiben, das alles in genügender Ausführlichkeit darstellt, was in den letzten hundert Jahren zur deduktiven Logik beigetragen worden ist, dürfte es ausgeschlossen sein, sämtliche Aspekte und Teilprobleme anzudeuten, geschweige denn auszuführen, die von Logikern und Mathematikern in den letzten Jahrzehnten zum Problem des induktiven Schließens analysiert worden sind. In diesem Buch möchte ich lediglich das herausgreifen, was für die *philosophische* Grundlegung des induktiven Schließens wichtig ist, oder genauer: was *mir* daran als wichtig *erscheint*, noch genauer: was mir daran *heute* als wichtig erscheint. Wenn ich dabei nicht zuviel übersehen habe, so ist der Zweck dieses Buches erreicht.

Schließlich möchte ich nicht versäumen, Herrn Professor Dr. Dr. Wolfgang Stegmüller für all die vielen Anregungen, die ich in Vorlesungen, Übungen und persönlichen Gesprächen während der letzten zehn Jahre direkt oder indirekt zum Induktionsproblem von ihm erhalten habe, zu danken, sowie dafür, daß er meine Bemühungen von Anfang an auf das Gebiet der exakten Philosophie gelenkt hat. Meinen Dank schulde ich ferner den Herren Dipl.-Ing. Dieter van Leyen und cand. phil. Christian Ullmann für ihre wertvolle Hilfe bei den Korrekturen sowie für viele inhaltliche wie auch formale Hinweise und Anregungen. Danken möchte ich auch meiner Frau Uta Essler, die mir geduldig und unermüdlich bei meinen ständigen Änderungen, Erweiterungen und Neufassungen des Textes geholfen hat.

Widmen möchte ich dieses Buch Herrn Professor Dr. Rudolf Carnap, der als erster ein System der induktiven Logik erarbeitet hat und dessen Analysen trotz der stürmischen Entwicklung aller Gebiete der Wissenschaftstheorie auch heute noch nicht überboten worden sind.

Günzburg, den 8. April 1969 *Wilhelm K. Essler*

I. Die geschichtliche Entwicklung des Induktionsbegriffs

1. Der Induktionsbegriff der antiken Philosophie

Wie in der deduktiven Logik ist Aristoteles (384/3—322/1 v. Chr.) auch in der induktiven Logik der erste gewesen, der über dieses Verfahren zur Gewinnung von Erkenntnissen aus anderen Urteilen reflektiert hat. Anders als in seiner Syllogistik hat er hier jedoch kein System von induktiven Schlüssen entwickelt, sondern sich auf gelegentliche Bemerkungen beschränkt. Wenn es also zwar nicht angeht, von einer aristotelischen *Theorie* der Induktion zu sprechen, so ist es doch sinnvoll, seine durchaus klaren *Vorstellungen* zu diesem Problem zusammenzustellen und zu zeigen, welche Gestalt eine derartige Theorie wohl angenommen hätte, wenn sie von ihm entwickelt worden wäre. Der Induktionsbegriff des Aristoteles wird verständlich, wenn man sich seine Logik und Wissenschaftstheorie vergegenwärtigt. Eine Wissenschaft, sofern sie wirklich zu Wissen und nicht nur zu Glauben führt, ist für ihn eine apodiktische Wissenschaft, die folgendermaßen aufgebaut ist:[1] Die meisten ihrer Sätze werden bewiesen. Ein Beweis, wenn er diesen Namen verdient, hat dabei die folgenden Eigenschaften: Er muß ein logischer Schluß sein, also die Form eines Syllogismus besitzen. Dies ist jedoch auch bei den sogenannten dialektischen Schlüssen der Fall und keine ausschließliche Eigentümlichkeit des wissenschaftlichen Beweises; vielmehr unterscheidet sich dieser von jenen dadurch, daß seine Prämissen wahre Aussagen sind, daß sie also bestehende Sachverhalte beschreiben, und daß diese Tatsachen im Sinne der Begründung dem durch den Schlußsatz ausgedrückten Sachverhalt vorangehen;[2] denn die Prämissen sollen die Ursachen oder die

[1] Seine Wissenschaftslehre hat Aristoteles hauptsächlich in an. post. I, 2 dargestellt; analysiert wurde sie erstmals von Heinrich Scholz, Die Axiomatik der Alten, in: Blätter für deutsche Philosophie 4 (1930/31) S. 259—278, Nachdruck in: *Mathesis Universalis*, Darmstadt 1961, S. 27—44. Vgl. hierzu auch Essler [1], S. 240 f. (Die in eckigen Klammern hinter den Autorennamen stehenden Zahlen in den Fußnoten sind die jeweiligen Titelnummern des Literaturverzeichnisses.)

[2] Vgl. dazu auch an. post. I, 72b 28—30, sowie an. pr. II, 68b 35 f.

Gründe des zu Beweisenden beschreiben, der Schlußsatz hingegen die Wirkung aus diesen Ursachen oder die Folge jener Gründe. Von jeder Wirkung kann man jedoch eindeutig zu den Ursachen aufsteigen und von jeder Folge zu den Gründen. Damit sind auch für jeden Beweis eindeutig die Prämissen des zu beweisenden Schlusses gegeben.

Nun gibt es Wirkungen; also gibt es auch Sätze, die bewiesen werden können. Folglich gibt es auch eine Wissenschaft davon. Eine solche könnte es jedoch nicht geben, wenn alle Sätze dieser Wissenschaft zu beweisen wären, wenn also keiner von ihnen ohne Beweis als gültig erkennbar wäre, da dann ein regressus ad infinitum vorliegen würde und überhaupt kein Satz bewiesen werden könnte. Wenn es also, wie gezeigt worden ist, eine Wissenschaft gibt, dann muß diese folgende Struktur haben: Sie geht von ersten Sätzen aus, also von Prinzipien, die jene Ursachen oder Gründe beschreiben, die der Sache nach zuerst sind; da diese also nicht selbst Wirkungen oder Folgen beschreiben, können sie auch nicht Schlußsatz eines Beweises sein. Sie sind vielmehr in diesem Sinn unbeweisbar. Auf Grund dieser Prinzipien können unter Benützung von Syllogismen andere Sätze bewiesen werden, mit diesen wiederum neue und so fort.

Hier wird man nun fragen: Auf welche Weise ist denn ein Prinzip überhaupt als gültig zu erkennen, wenn nicht durch einen Beweis? Die Antwort, die Aristoteles darauf gibt, ist: durch Induktion.[3] Während der Beweis davon ausgeht, was *der Sache nach* früher und einleuchtender ist, benützt die Induktion als Prämisse, was *uns* früher einleuchtet.[4] In diesem Sinn ist die Induktion das Gegenteil des deduktiven Schlusses, da sie von dem, was eigentlich zu beweisen ist, zu den Prinzipien führt.[5]

Die Induktion (*ἐπαγωγή*, Herbeischaffung) soll die Prinzipien *herbeischaffen*; sie kann also nicht diese Prinzipien *voraussetzen*. Andererseits gibt es keine Sätze, die diese Prinzipien logisch implizieren. Was also sind die Prämissen der Induktion? Die Antwort des Aristoteles lautet: Die Einzelfälle des allgemeinen Satzes;[6] sind mir alle Einzelfälle als wahr bekannt und gewiß, so kann ich auch die entsprechende

---

3 Vgl. an. post. II, 100b 3f., sowie an. pr. II, 68b 30—33.
4 an. post. I, 72b 28—30; an. pr. II, 68b 35f.
5 top. 105a; an. pr. II, 68b 33—35.
6 an. pr. II, 68b 15—21.

Allaussage als wahr und gewiß annehmen. Wie erlange ich jedoch Gewißheit von der Vollständigkeit der Aufzählung? Denn wenn ich einige Einzelfälle jener allgemeinen Behauptung als gewiß erkenne und überhaupt an keinem mir bekannten Einzelfall zweifle, so folgt daraus noch nicht die Gewißheit der Allaussage; es wäre ja denkbar, daß es Einzelfälle gibt, die diese Allaussage widerlegen, die mir aber unbekannt sind. Zu diesem Problem hat Aristoteles nicht eindeutig Stellung genommen. Man könnte ihn so verstehen, daß nach seiner Vorstellung die Welt nur aus endlich vielen unteilbaren Dingen, Individuen, besteht und daß also eine derartige vollständige Aufzählung im Prinzip stets möglich ist.[7]

Nach einer anderen Deutungsmöglichkeit ist diese Aufzählung überhaupt nicht nötig. Es genügt vielmehr, einen einzigen Einzelfall der allgemeinen Behauptung zu betrachten und eingehend zu analysieren, um jene Allaussage als gewiß einzusehen.[8] Es ist dann zwar wichtig, zur Kontrolle auch noch andere Einzelfälle zu betrachten, doch verhelfen diese nicht mehr zu neuer Einsicht, sondern haben vielmehr nur die Funktion, zu zeigen, daß keine Fehler in der Analyse des ersten Beispiels enthalten sind. Bei dieser Deutung müßte man auf die Frage, woher man die Gewißheit aus dem Einzelfall nimmt, wohl erwidern: Man erhält sie, wenn man die Begriffe, die in dem den Einzelfall beschreibenden Satz vorkommen, expliziert, analysiert und vergleicht. Auf die konkreten Individuen kommt es dann gar nicht an, sie dienen in den Überlegungen nur als Variable. Der Schluß auf die allgemeine Behauptung entspricht dann dem Allschluß in der deduktiven Logik.[9] Während bei jener Deutung der Allsatz nur empirische Gewißheit erhält, ist er bei dieser a priori gültig und sogar analytisch, da seine Wahrheit allein aus den Begriffen unter Verwendung der Regeln der deduktiven Logik folgt.

---

[7] Dafür spricht auch die Stelle in an. post. II, 92a 37—39, daß man auf Grund der Induktion erkennt, daß alles so ist, weil nichts anders ist, weil man also zumindest im Prinzip feststellen kann, daß nichts anders ist. Vgl. hierzu auch top. 108b 9—12.

[8] an. post. II, 100a 15 — b 1. Vgl. hierzu auch top. 158a 33 — b 4 und 158b 24 — 159a 2, wo er ausdrücklich sagt, daß Definitionen die Grundlagen der Prinzipien sind.

[9] Vgl. die Regel der Alleinführung im Kalkül des natürlichen Schließens von Gentzen, siehe auch Essler [1], S. 121.

Wie schon erwähnt, ist es nicht klar, an welche der beiden Deutungen Aristoteles gedacht hat. Die analytische Interpretation gibt den Prinzipien jenen Grad an Gewißheit, dessen sie bedürfen, um unmittelbar einsichtig zu sein. Andererseits ist es dann nicht sinnvoll, der Deduktion die Induktion gegenüberzustellen, da sie in diesem Fall nichts anderes ist als eine spezielle Art des Analysierens und Deduzierens.

Zu klären bleibt noch die Frage, worauf sich die Einzelfälle beziehen. An einigen Stellen scheint es, als ob Aristoteles auch dann noch an Allgemeinbegriffe, also an Klassen, und nicht an Individuen denkt.[10] Andere Stellen zeigen jedoch, daß er primär an Individuen denkt und erst in zweiter Linie an Gesamtheiten solcher Gegenstände.[11]

In welchem Verhältnis stehen nun in einem induktiven Argument Prämissen und Konklusion zueinander? Daß die Induktion die Umkehrung eines Beweises ist, hilft hier offenbar nicht weiter. Andererseits ist die Induktion aber ein Schluß, der zu Wissen und Gewißheit führen muß, da sie ja dazu dienen soll, die Prinzipien als gültig zu erkennen. Sie muß also in irgendeiner Beziehung zum Syllogismus stehen. Nun ist der Gehalt der Konklusion eines Syllogismus in dem seiner Prämissen vollständig enthalten; die Annahme, bei einer Interpretation der Sprache könnten die Prämissen wahr und die Konklusion falsch sein, schließt einen Widerspruch ein. Bei einem induktiven Argument ist dies nicht der Fall: Die Annahme der Falschheit der Konklusion und der Wahrheit der Prämissen sind miteinander verträglich. Logisch schlüssig wird das Argument erst, wenn man die Prämissen verstärkt, indem man noch gewisse Zusatzbedingungen hinzufügt.[12] Die Induktion ist also nach Aristoteles eine unvollständige logische Folgerung, oder, wie man auch sagen kann, eine partielle Implikation. Nach dem Grad dieses Einschlusses bemißt sich

---

[10] an. pr. II, 68b 15—21.

[11] an. pr. II, 69a 13—19; rhet. 1357b 10—13, 26—36; an. post. II, 100a 15 — b 5.

[12] Vgl. an. pr. II, 70a 16—35. In dieser Weise deute ich auch rhet. 1357b 10—21. So ist wohl auch an. pr. II, 68b 15—24 zu verstehen, da hier vorausgesetzt ist, daß der Begriff C nicht über den Mittelbegriff B hinausreicht. In top. 108b 12—19 charakterisiert er die hypothetischen Schlüsse als jene, bei denen es wahrscheinlich (wenn auch nicht sicher) ist, daß es sich so wie in den bisher beobachteten Fällen auch in den übrigen verhält; ein Teil der Allbehauptung folgt also logisch aus der Beobachtung, der andere hingegen nicht, weshalb der Allsatz nur partiell durch die Beobachtung impliziert wird.

dann auch der Grad der Gewißheit, mit der wir die Konklusion annehmen, falls wir von der Wahrheit der Prämisse überzeugt sind.
Auf den ersten Blick mag es scheinen, daß Aristoteles nur *eine* Art von induktiven Argumenten kennt, nämlich den sogenannten Allschluß, dessen Konklusion eine Allbehauptung ist und dessen Prämissen Einzelfälle eben dieses Allsatzes darstellen. Dieses Argument ist für seine Wissenschaftstheorie ja von grundlegender Bedeutung, da unter seiner Verwendung die Prinzipien als gültig erkennbar sein sollen, die stets generelle Aussagen sind. Daneben kennt er aber auch den Schluß von Einzelfällen auf einen anderen Einzelfall, etwa, daß ein Ding eine gewisse Eigenschaft besitzt, weil ein oder mehrere andere Individuen diese Eigenschaft ebenfalls haben (und keine oder praktisch keine Gegenbeispiele vorhanden sind);[13] dieses Argument, das er ebenfalls als nicht schlüssig nachweist, wird von Carnap „singulärer Voraussageschluß" genannt.
Interessant ist hier der folgende Zusammenhang zwischen dem singulären Voraussageschluß und dem Allschluß:[14] Angenommen, Peisistratos erstrebte die Gewaltherrschaft, als er eine Leibwache forderte, und desgleichen Theagenes in Megara, und angenommen, es gilt von allen uns bekannten Personen, daß sie entweder keine Leibwache forderten oder aber, falls sie dies taten, die Alleinherrschaft erstrebten. Dann werden wir sicher sein, daß auch Dionysios die Gewaltherrschaft erstrebt, wenn er eine Leibwache fordert, und zwar sind wir dies in dem gleichen Grad, in dem wir auf Grund jenes Datums dem Allgemeinsatz, daß der nach Gewaltherrschaft Strebende eine Leibwache fordert, vertrauen. Die Wahrscheinlichkeit eines Allsatzes im Hinblick auf ein Datum ist dann gerade die Wahrscheinlichkeit seiner Einzelfälle im Hinblick auf jenes Wissen, und wir werden deshalb von der Erfahrung über diese singulären Einzelfälle zu dem Allsatz *hingeführt*. Es wäre dies eine dritte Deutung der aristotelischen Induktion, soweit sie sich auf den Allschluß bezieht, eine Deutung, die übrigens vor kurzem auch Carnap vorgenommen hat.[15] Nach ihr ist es nicht

---

[13] Vgl. an. pr. II, 70a 24—28, ferner rhet. 1357b 13—21, 26—36; diese Arten von induktiven Argumenten nennt er „Zeichen", „Anzeichen" und „Beispiele".
[14] Vgl. rhet. 1357b 26—36; vgl. hierzu auch top. 156b 10—17 und 160a 35 bis b 2.

mehr, wie bei der ersten, notwendig, *sämtliche* Einzelfälle des Allsatzes zu kennen, um auf diesen schließen zu können.
Schließlich ist noch zu bemerken, daß Aristoteles bereits statistische Gesetze, d. h. objektive Wahrscheinlichkeiten und relative Häufigkeiten, gekannt hat; so sind z. B. die Sätze „Die Neider hassen meistens" und „Die Neider lieben meistens nicht" Wahrscheinlichkeitsaussagen, die nicht ein Verhältnis von Prämissen und Konklusion ausdrücken (und also kein Zeichen, kein induktives Argument sind), sondern die eine relative Häufigkeit beschreiben und folglich ein Satz sind. Sätze dieser Art können wiederum als Prämissen von deduktiven und induktiven Argumenten verwendet werden.[16]
Die analytische Deutung des aristotelischen Induktionsbegriffs liegt auch insofern nahe, als sie sich mit der Einstellung Platons deckt, daß Schlüsse aus Beobachtungen nur auf *Scheinbares* und *Geglaubtes* führen, während man durch Wesensanalysen der Begriffe zu den *wirklichen* (*wahren*) Ursachen der Dinge und damit zu *Wissen* gelangt.[17] Dieses ohne Zweifel an der Mathematik der Antike orientierte Wissenschaftsideal ist später insbesondere von der Stoa vertreten worden. So hat Dionysios von Cyrene[18] behauptet, daß man Sätze wie „Alle Menschen sind sterblich" durch die Methode der sogenannten Kontraposition (*ἀναστενή*) beweisen könne, indem man sich vergegenwärtigt, daß die Annahme, ein Mensch sei nicht sterblich, einen logischen Widerspruch in sich schließe. Aus den empirischen Daten, daß sich alle bisher beobachteten Menschen als sterblich erwiesen haben, könne man jene generelle Aussage jedoch nicht erschließen, auch durch Induktion nicht, da ohne zusätzliche Voraussetzungen nicht ausgeschlossen sei, daß die Konklusion trotz jener Erfahrungen falsch sei.
Er bezieht sich damit auf die epikureische Lehre von der adäquaten Methode der Wirklichkeitserkenntnis, die man als eine empirische Deutung des aristotelischen Induktionsbegriffs ansehen kann, die

15 Vgl. die Bemerkungen im Anschluß an Theorem 20 auf S. 219ff. dieses Buches.
16 Vgl. an. pr. II, 70a 2—11, sowie rhet. 1357a 32 — b 2.
17 Dies wird von Platon u. a. an folgenden Stellen ausgedrückt: Philebos 55e bis 56d, Phaidon 96b, 99d — 101e, Gorgias 465a, 500e — 501b, Phaidros 269e bis 272b, Politeia 516c—d, 533b—c.
18 Vgl. Philodemus [1], Col. I[a] 1 — VII 5, insbesondere IV 5—16.

jedoch vermutlich durch Reflexion über die in der Schule des Hippokrates benützten Verfahren zur Erforschung der Ursachen der einzelnen Krankheiten und deren Heilung gewonnen worden ist.
An dieser antiken Heilkunde und nicht an der Geometrie jener Zeit orientiert sich das Wissenschaftsideal der Epikureer, wie es in dem Lehrbuch des Philodemus durch die Theorien des Zenon von Sinon, der sogar die Sätze der Mathematik durch Induktion begründet,[19] des Bromius, des Demetrius des Lakoniers und eines vierten, dessen Namen verlorengegangen ist (Cleomedes?), geschildert wird. Diese von Philodemus im Jahre 54 v. Chr. oder 40 v. Chr. aufgezeichneten Ansichten lassen sich folgendermaßen zu einer epikureischen Lehre von der Induktion zusammenfassen:
Der induktive Schluß (*ἐπιλογισμός*) ist eine Folgerung, deren Prämisse Beobachtungen beschreibt und deren Konklusion sich auf nicht wahrgenommene Objekte bezieht. Da von uns und von anderen zu jedem Zeitpunkt stets nur endlich viele Dinge wahrgenommen worden sind, kann die Prämisse stets nur endlich viele Einzelbeobachtungen zusammenfassen und nicht *alle* Ereignisse einer gewissen Art beschreiben, insbesondere nicht die in der Zukunft liegenden. Die Konklusion hingegen ist entweder eine generelle Hypothese wie „Alle Menschen sind sterblich"[20] oder eine generelle Hypothese über das nicht Wahrgenommene wie „Alle nicht wahrgenommenen Menschen sind sterblich"[21] oder eine singuläre Hypothese wie „Jener Mensch ist sterblich"[22]. Diese drei Konklusionen werden als miteinander gleichwertig angesehen, d. h. wenn aus einem Erfahrungsdatum eine von ihnen induktiv erschlossen werden kann, so auch die beiden anderen.
Ein induktives Argument oder, wie statt dessen oft gesagt wird, ein Schluß aus Analogie (*ὁμοιότης*) kann Wahrscheinlichkeit (*εὐλογία*) oder Sicherheit geben. Ein Schluß mit Wahrscheinlichkeit wird aus einer Prämisse gezogen, die nicht ganz gleichförmig ist und also Ge-

[19] Vgl. Philodemus [1], XV 13 — XVI 1.
[20] Vgl. Philodemus [1], XVI 4 — XVII 3 und XXXI 26—36.
[21] Vgl. Philodemus [1], XXVIII 15—25; dieser Schluß wird von Carnap „Voraussageschluß" genannt, vgl. Carnap-Stegmüller [1], S. 81.
[22] Vgl. Philodemus [1], XXVI 32 — XXVII 9; diesen Schluß von einigen Instanzen einer Art auf ein anderes Ding der gleichen Art nennt Carnap, wie schon erwähnt, einen „singulären Voraussageschluß".

genbeispiele enthält, deren Gegenbeispiele jedoch nicht sonderlich ins Gewicht fallen, so wie etwa „Im Sommer kann man sicher segeln" auf Grund von „Im Sommer sind, soweit unsere Erfahrungen reichen, die Winde meist günstig", aber auch (wie noch hinzuzufügen wäre) aus einer Prämisse, die zwar gleichförmig ist und also keine Gegenbeispiele enthält, die jedoch Gegenbeispiele nicht vollständig ausschließt. Sicherheit liegt dagegen vor, wenn die Prämisse völlig gleichförmig ist und jedes Gegenbeispiel ausschließt. Sicherheit ist also ein Höchstmaß an Wahrscheinlichkeit, und Wahrscheinlichkeit eine relative Sicherheit.[23]

Wichtig für die Erkenntnis der Naturgesetze sind nach Ansicht der Epikureer jene Schlüsse aus Analogie, die Sicherheit geben. Nur solche Argumente nämlich geben die Gewähr, daß die Konklusion wahr ist, denn nur solche Argumente führen nicht von Wahrem zu Falschem, wobei die Prämisse stets eine Wahrnehmung beschreibt und demnach immer wahr ist. Auf den ersten Blick scheinen nur die Schlüsse der *deduktiven* Logik dieses Merkmal zu erfüllen, was ja auch von den Stoikern behauptet worden ist. Die Epikureer haben jedoch geglaubt, ein Kriterium für korrekte *induktive* Argumente zu besitzen, das *Sicherheit* gewährleistet, nämlich das der Unvorstellbarkeit (*ἀδιανόητος*). So ist es nach ihrer Ansicht auf Grund unserer Erfahrungen bezüglich der Sterblichkeit von Menschen unvorstellbar, daß ein (noch lebender oder noch nicht geborener) Mensch nicht sterblich ist, und dies besagt nichts anderes, als daß aus jenen Erfahrungen durch Analogie auf seine Sterblichkeit geschlossen werden kann, bzw. daß aus jenen Erfahrungen induktiv die Behauptung „Alle Menschen sind sterblich" ableitbar ist.[24] Unvorstellbar wäre auf Grund der Erfahrungen ebenfalls, daß ein in die Luft geworfener Stein nicht mehr zur Erde zurückkehrt, vorstellbar hingegen, daß ein Schwan schwarz ist, obwohl sich alle (zu jener Zeit) beobachteten Schwäne als weiß erwiesen haben, weil einige Vögel in ihren Farben nämlich nicht uniform

[23] Vgl. Philodemus [1], VII 32—38, XXV 24 — XXVI 9 sowie Frag. III 5—12. Die Rhetoriker und Ärzte jener Zeit kannten auch schon verschiedene Grade der Wahrscheinlichkeit, je nach der Beschaffenheit der ihnen zur Verfügung stehenden Daten; vgl. Sextus Empiricus, Adversus Mathematicos, VIII 175—184, in: Opera, hrsg. von H. Mutschmann, bearbeitet von J. Mau Band I—III, Leipzig 1954—58.

[24] Vgl. Philodemus [1], XIV 11—21 und XXVIII 15—25.

sind (weil z. B. einige Raben weiß sind). Ein solcher Schluß würde demnach nur mit Wahrscheinlichkeit, nicht jedoch mit Sicherheit gelten.
Ein induktives Argument ist inkorrekt, wenn *nicht* die gesamte zur Verfügung stehende und für die Konklusion relevante Erfahrung (das Gesamtdatum) berücksichtigt wird, gleichgültig, ob es Sicherheit oder nur Wahrscheinlichkeit nach sich zieht. Wenn man etwa aus der Tatsache, daß *bei uns* alle Menschen weiß sind, induktiv schließt, daß *alle* Menschen weiß sind, oder daraus, daß in Äthiopien jedermann schwarz ist, folgert, daß alle Menschen schwarz sind, so argumentiert man nicht korrekt; im allgemeinen wird sich der Schlußsatz dabei als falsch erweisen. Man kann nur dann sicher sein, daß es keine Dinge gibt, die unseren Erfahrungen widerstreiten, wenn man die Erscheinungen in ihren mannigfachen Verschiedenheiten sorgfältig beobachtet hat, und nur dann ist ein Schluß aus Analogie korrekt.[25] Ein induktives Argument führt demnach genau dann zu einer sicheren Konklusion, wenn das Gesamtdatum kein Gegenbeispiel enthält und wenn auf Grund von Analogiebetrachtungen bezüglich ähnlicher Eigenschaften kein Gegenbeispiel vorstellbar ist. Solche Analogiebetrachtungen sind jedoch, genau besehen, ein induktives Argument von höherer Art: Es wird aus der Homogenität (*ὁμογένεια*) einiger Eigenschaften einer gewissen Art auf die Homogenität einer anderen Eigenschaft der gleichen Art geschlossen. Daß gewisse Eigenschaften homogen sind, ist zwar im allgemeinen nicht beobachtbar, sondern kann nur aus Beobachtungen erschlossen werden, doch darf man den erschlossenen Tatsachen genauso trauen wie den beobachteten und kann sie folglich auch als Prämisse eines neuen Arguments verwenden.[26]
Daß die gesamte Erfahrung zu berücksichtigen ist, wird von den Epikureern manchmal auch in einem strengeren Sinn verstanden: Die Wahrnehmungen sind so zu ordnen, daß man sicher sein kann, alle Variationen berücksichtigt zu haben; die Dinge unserer Erfahrung müssen in jeder Hinsicht geprüft werden und dürfen nicht einfach wahllos hingenommen werden, wenn das Gegenteil der Konklusion unvorstellbar sein soll.[27] Der Schluß aus Analogie ist demnach korrekt

[25] Vgl. Philodemus [1], XXX 20—33 und XXXIII 9—15.
[26] Vgl. Philodemus [1], Frag. II.
[27] Vgl. Philodemus [1], XXXIII 9—15 und XXVIII 37 — XXIX 4.

und das Gegenteil des zu Erschließenden unvorstellbar genau dann, wenn man die wahrgenommenen Dinge *systematisch* überprüft und wenn man die Dinge in *allen* ihren Variationen betrachtet (z. B. wenn man große und kleine Menschen, schwarze und weiße, kluge und dumme auf ihre Sterblichkeit hin untersucht). Wenn man sich also nicht nur auf zufällige Daten verläßt, sondern in dieser Weise systematisch beobachtet, dann ist der Schluß sicher und folglich der Schlußsatz wahr, auch wenn man, was der Normalfall sein wird, nicht alle Dinge beobachtet hat.[28]

Diese Sicherheit besteht jedoch nur, wenn die Dinge nur endlich vieler Variationen fähig sind und damit nur endlich viele beobachtbare Eigenschaften haben können. Es darf angenommen werden, daß dies der epikureischen Lehre entspricht, da nach dieser jeder Gegenstand eine der Variationen besitzt, die wir beobachten, gleichgültig, ob er wahrgenommen worden ist oder nicht; sonst wäre er ja *unvorstellbar*.[29]

Gegen die Behauptung, daß ein induktives Argument zu Sicherheit führen könne, hat Dionysios folgenden prinzipiellen Einwand erhoben:[30] Wenn alle bisher wahrgenommenen Menschen sterblich sind, so ist dies damit logisch verträglich, daß einige oder alle der nicht wahrgenommenen Menschen unsterblich sind, es sei denn, man setzt zusätzlich voraus, daß die nicht wahrgenommenen Menschen den wahrgenommenen bezüglich der Eigenschaft *sterblich* ähnlich sind, daß diese Eigenschaft also im Bereich der Menschen *homogen* ist, daß der Bereich der Menschen in bezug auf sie *uniform* ist. Allein unter dieser Voraussetzung ist jener Schluß korrekt.

Zenon versucht diesen Einwand auf folgende Weise zu entkräften: In den induktiven Argumenten, die wir entwickeln, wird diese Zusatzannahme nirgendwo gemacht; vielmehr werden als Prämissen immer nur Beobachtungsaussagen benützt sowie Sätze, die daraus bereits erschlossen worden sind. Ob eine Eigenschaft homogen ist oder nicht, steht nicht a priori fest, sondern muß von Fall zu Fall aus der Erfahrung erschlossen werden, und zu einem derartigen Schluß

---

28 Vgl. Philodemus [1], XX 31 — XXI 16.
29 Vgl. Philodemus [1], XXI 22—29.
30 Vgl. Philodemus [1], II 26 — IV 16 und IV 30—34.

kann man nur induktive Argumente benützen, nicht jedoch (ausschließlich) deduktive. Wesentlich ist, daß die Erfahrung bezüglich der fraglichen Eigenschaften durchgehend uniform ist; dann darf man annehmen, daß der Gegenstandsbereich bezüglich dieser Eigenschaft auch außerhalb des Erfahrungsbereichs uniform sein wird.[31]

Man wird nicht sagen können, daß Dionysios damit durch Zenon widerlegt worden ist; denn um aus der Uniformität in der Wahrnehmung auf Uniformität außerhalb von ihr logisch schließen zu können, muß man ja wiederum Uniformität voraussetzen. Die Epikureer können demnach *nicht* voraussetzungslos aus der Erfahrung induktive Schlüsse ziehen, wie sie behaupten. Andererseits wird man Zenon zugeben müssen, daß die Frage, ob eine Eigenschaft in einem Bereich homogen oder nicht homogen ist, nicht analytisch, d. h. nicht durch Analyse der Bedeutungen der Begriffe allein, entscheidbar ist; ob etwa *sterblich* oder *intelligent* unter Menschen homogen ist, kann nicht a priori, sondern nur empirisch, also unter Zugrundelegung von Erfahrungen, festgestellt werden. In voller Klarheit und Deutlichkeit wird hier also ein Grundproblem allen induktiven Schließens formuliert und diskutiert.

Manche Eigenschaften haben in gewissen Bereichen einen maximalen Grad an Homogenität, so daß man bereits auf Grund einer einzigen positiven Instanz (bei Abwesenheit von negativen Instanzen) sicher sein kann, daß auch irgendein anderer Gegenstand diese Eigenschaft besitzt, und damit, daß überhaupt alle Dinge von dieser Eigenschaft sind. In anderen Fällen ist der Grad an Homogenität nur so hoch, daß man Sicherheit erst nach zwei, drei oder noch mehr positiven Instanzen (bei keiner Gegeninstanz) erhält. Und oft benötigt man eine große Anzahl von weitgestreuten Beobachtungen, um diese Sicherheit zu erlangen, weil der Grad an Homogenität wesentlich kleiner ist als in jenen Fällen, wenn auch noch groß genug, um Sicherheit und nicht nur Wahrscheinlichkeit zu erlangen.[32] Diese Homogenität erschließt man, indem man aus der beobachteten Uniformität einer Klasse von Dingen bezüglich einer Eigenschaft (z. B. der Klasse der Menschen

---

[31] Vgl. Philodemus [1], XVI 4 — XVII 3, XVII 15—28 und XXII 37 — XXIII 7, sowie XXXV 35 — XXXVI 11.

[32] Vgl. Philodemus [1], XXVI 32 — XXVII 9, sowie VIII 16—22.

bezüglich der Sterblichkeit) auf die Uniformität einer anderen Klasse induktiv schließt, deren Objekte den Gegenständen jener früheren Klasse ähnlich sind (etwa auf die der Klasse der Affen).

Die Epikureer kennen also zwei Arten von induktiven Argumenten, die mit Sicherheit gelten: den Schluß von einzelnen Dingen einer Klasse auf andere Einzeldinge der gleichen Klasse (den Schluß auf „identische Dinge") und den Schluß von einzelnen Klassen von Gegenständen, die sich in einer gewissen Hinsicht ähnlich sind, auf eine andere Klasse von Gegenständen, die mit jenen ebenfalls ähnlich sind (den Schluß auf „ähnliche Dinge").[33] Die Homogenität einer Eigenschaft wird durch Induktion der zweiten Art erschlossen; dies setzt jedoch voraus, daß man die Homogenität wenigstens einer Eigenschaft als eine gesicherte Tatsache ansehen darf. Um dies aber tun zu können, muß man alle Dinge auf diese Eigenschaft hin untersucht haben oder aber deren Homogenität mit schwächeren Voraussetzungen erschließen, wenn die Argumentation zirkelfrei sein soll. In den *uns erhaltenen* Fragmenten der epikureischen Lehre von den induktiven Argumenten wird diese Frage nicht gestellt und folglich auch nicht beantwortet.

Möglicherweise war diesen Philosophen das Problem jedoch bekannt, und man darf als ihren Lösungsvorschlag die epikureische Lehre von den konstanten und variablen Eigenschaften ansehen. Sie betonen nämlich ständig, daß man nur konstante Eigenschaften und keine zufälligen Gemeinsamkeiten der Dinge eines zugrundegelegten Bereichs bei induktiven Schlüssen verwenden dürfe. Damit ist folgendes gemeint: Über Heckenrosen, Veilchen, Birken und andere Pflanzenarten kann man nach jeweils wenigen Beobachtungen bereits sichere Generalisierungen vornehmen, da Pflanzen einer Art sich voneinander kaum unterscheiden, nicht jedoch bezüglich der Klasse der Pflanzen schlechthin, denn diese unterscheiden sich ja voneinander in den verschiedensten Hinsichten, so daß ihre Gesamtheit nur einen kleinen Grad an Uniformität besitzt. Will man durch induktive Argumente Sicherheit erlangen, so hat man daher stets die speziellsten Eigenschaften zu wählen, die überhaupt in Betracht kommen können; zwei verschiedene Dinge, die eine solche Eigenschaft besitzen, unterscheiden

---

[33] Vgl. Philodemus [1], XVIII 17 — XIX 4.

sich dann in keiner charakteristischen Weise mehr voneinander und werden dann von den Epikureern „identisch" genannt.[34]
Allerdings wird die Frage, *warum* dies so ist, in den uns erhaltenen Texten nicht klar beantwortet. Sagt man, dies sei eine Erfahrungstatsache, so besteht wiederum die Gefahr einer zirkelhaften Argumentation. Nimmt man hingegen ein Prinzip der allgemeinen Gesetzmäßigkeit der Natur an, so scheinen nur drei Möglichkeiten offenzustehen: Entweder man versucht, es durch Induktion zu begründen und setzt sich damit der Gefahr eines Zirkelschlusses aus; oder man versucht, es auf apriorischem Weg zu deduzieren und verläßt damit die empiristische Grundeinstellung der Epikureer; oder man nimmt diese Aussage zunächst nur als Hypothese an und ist bereit, sie zu verwerfen, sobald sich herausstellt, daß sie der Erfahrung und den daraus gewonnenen Konklusionen widerspricht. Vielleicht haben sie ein solches Gesetz hypothetisch angenommen und sodann versucht, es durch den Nachweis zu bestätigen, daß es mit der Erfahrung in Einklang steht und daß alle ihm widersprechenden Behauptungen auch der Erfahrung widersprechen; diese Methode der Bestätigung von Aussagen wäre dann eine dritte Art von induktiven Schlüssen, die sie gekannt haben. Tatsächlich haben sie in ihrer Naturphilosophie auch solche Argumente benützt und z. B. versucht, ihre Theorie von den Atomen durch den Nachweis zu begründen, daß sie widerspruchsfrei mit der Erfahrung sei, aber jede mit ihr unverträgliche Hypothese auch der Erfahrung widerspreche.[35]
Die Werke der Epikureer über das induktive Schließen, die an Gehalt den Aristotelischen Schriften zur deduktiven Logik kaum nachstehen, sind in der ausgehenden Antike vergessen worden und haben in den Werken der späteren Philosophen kaum Spuren hinterlassen. Erst 1865 wurden die Schriften des Philodemus über die Induktion von Theodor Gomperz veröffentlicht. Sie haben jedoch trotzdem noch nicht den Platz in der Geschichte der Philosophie und der induktiven Logik erhalten, der ihnen gebührt.

---

[34] Vgl. Philodemus [1], XIII 1 — XIV 2, XIV 39 — XV 13, XVII 30—36, XVIII 17 — XIX 4, XXV 24 — XXVI 9 und XXVIII 25—29. Eine induktive Methode, die Individuen (und analog auch Eigenschaften) auf diese Weise als gleich behandelt, wird auf S. 144f. dieses Buches „symmetrisch" genannt.
[35] Vgl. Philodemus [1], XXXVI 11—17 und XXXVII 24—29.

2. *Enumerative und eliminative Induktion*

Nach den Epikureern hat vor allem Lord Francis Bacon von Verulam (1561—1626) versucht, unabhängig von den diesbezüglichen Vorstellungen des Aristoteles und darüber hinaus sogar diesen entgegengesetzt eine Theorie der Induktion zu entwerfen. Nach seiner Ansicht gibt es höchstens zwei Wege, auf denen ein Naturgesetz als wahr erkennbar ist: Indem man entweder aus den Einzelfällen durch Induktion die Prinzipien gewinnt und aus diesen durch Deduktion dann die abgeleiteten Gesetze, wie es Aristoteles getan hat, oder aber, indem man durch Induktion zunächst zu den einfachsten Gesetzen und sodann schrittweise und wiederum auf induktivem Weg zu immer höheren Theorien aufsteigt, bis man auf diese Weise die höchsten Prinzipien erreicht.[36]

Es muß bemerkt werden, daß Bacon hier Aristoteles mißversteht. Nach Aristoteles *können* die vorgegebenen Sätze einer Wissenschaft so geordnet werden, daß aus einigen Prinzipien alle übrigen Urteile deduzierbar und damit begründbar sind; da die Prinzipien jedoch nicht auf diese Weise zu rechtfertigen sind, benötigt man induktive Argumente, um auch sie als gültig einzusehen. Er hat jedoch nirgends gesagt, daß dies der einzige Weg ist, um irgendwelche Gesetze als wahr zu erkennen. Zur *Erkenntnis* der Sätze wird man vielmehr auch nach seiner Ansicht eine schrittweise und auf immer höheren Stufen operierende induktive Methode benützen, die nicht den Weg einzugehen braucht, auf dem nachträglich die so gewonnenen Sätze geordnet werden.[37] Die Methode Bacons zur *Auffindung* der Gesetze ist also durchaus zu vereinbaren mit der Aristotelischen Methode zur *Begründung* dieser Aussagen.

Nach Bacon hat Aristoteles nur die folgende Methode zur Verfügung, um von Einzelfällen zu einer generellen Aussage zu gelangen: Die der *einfachen Aufzählung*. Es werden dabei die zur Verfügung stehenden Einzelfälle der Theorie mit dieser Theorie verglichen, wobei die Ein-

---

[36] Vgl. Bacon [1], I, 19. Die Gedanken Bacons sind übrigens in vielen Abschnitten von spätmittelalterlichen Autoren vorweggenommen worden, wie etwa von Dietrich von Freiberg; vgl. A. C. Crombie, Robert Grosseteste and the Origins of Experimental Science, 1100—1700, Oxford 1963. (Diesen Hinweis verdanke ich Herrn Professor N. Lobkowicz.)

[37] Vgl. an. post. II, 100a 15 — b 3.

zelfälle aber (normalerweise) weder den Gegenstandsbereich erschöpfen noch zur Einsicht verhelfen, daß die Allgemeinbehauptung notwendig ist. Die Konklusion eines solchen Arguments ist also auf Grund jener Beobachtungsbasis (noch) nicht sicher; es besteht immer die Gefahr, daß Gegenbeispiele gefunden werden, die jene Konklusion (jene Hypothese) als falsch erweisen.[38] Nach seiner Ansicht ist aber die Anzahl der positiven Instanzen für das Induktionsproblem unerheblich (wenn nur wenigstens eine positive Instanz vorhanden ist, wie noch hinzuzufügen wäre); doch genügt ein einziges Gegenbeispiel, um die Theorie zu widerlegen.[39] Man benötigt also, so Bacon, eine andere und für den wissenschaftlichen Fortschritt nützlichere Methode der Induktion, die die Erfahrungsdaten systematisch zusammenstellt und durch geeignete Verfahren alle in Frage kommenden Hypothesen bis auf eine eliminiert. Die Aristotelische Theorie der Induktion und ähnliche Systeme, die nur die Hypothese und das Datum miteinander vergleichen, werden seitdem oft „Systeme der enumerativen Induktion" genannt, während man das Verfahren Bacons zur Eliminierung miteinander konkurrierender Hypothesen durch ein Datum und ähnliche Methoden unter dem Begriff „Systeme der eliminativen Induktion" zusammenfaßt.

Die Wissenschaft hat die Aufgabe, die Ursachen der einzelnen Wirkungen[40] zu erforschen. Wie aber soll sie dabei verfahren? Nach Bacon muß dazu das vorhandene Beobachtungsmaterial in die drei Gruppen der Anwesenheit, der Abwesenheit und des Grades eingeteilt werden.[41] Zur Gruppe der Anwesenheit sind dabei alle Daten zu zählen, die in der Wirkung übereinstimmen, z. B. daß sie heiß sind, die sich jedoch sonst in allen erdenklichen Hinsichten voneinander unterscheiden; keine dieser verschiedenen Eigenschaften kann dann die Ursache der Hitze sein, da sie ja im einen Fall zwar zusammen mit Hitze vorkommen, im anderen jedoch nicht. In der zweiten Gruppe der Abwesenheit faßt man alle Daten zusammen, die von irgendwelchen Dingen sagen, daß sie jene Wirkung nicht besitzen (daß sie also

---

38 Vgl. Bacon [1], I, 105.
39 So ist wohl Bacon [1], I, 46 zu verstehen.
40 Ich interpretiere seine Begriffe „erzeugende Natur" und „erzeugte Natur" als *Ursache* und *Wirkung*.
41 Vgl. Bacon [1], II, 11—13.

nicht heiß sind), daß ihnen vielmehr gewisse andere Eigenschaften zukommen; auch hier kann keine dieser anderen Eigenschaften die Ursache der Hitze sein. In der dritten Gruppe werden schließlich alle Einzelfälle der Wirkung zusammen mit dem Grad der Wirkung und den Graden der anderen Eigenschaften aufgezählt; alle Eigenschaften, die nicht gleichmäßig mit der Wirkung variieren, scheiden als Ursache aus. Bacon kommt dann zu dem Schluß, daß Bewegung die Ursache der Hitze ist.

Er hat also scheinbar eine Methode zur Erzeugung von Theorien aus Beobachtungsdaten entdeckt.[42] Nun kann nicht geleugnet werden, daß Naturwissenschaftler in einzelnen Fällen zur Entdeckung von Hypothesen tatsächlich so oder so ähnlich vorgehen.[43] Falsch ist jedoch die Behauptung, daß man damit alle wissenschaftlichen Theorien finden könne, also insbesondere etwa die höheren Theorien der Physik; ob der Weg, auf dem Newton, Einstein und andere zu ihren Systemen gelangt sind, überhaupt rational beschreibbar ist, darf heute mehr denn je in Frage gestellt werden. Möglicherweise hat auch Bacon seine Methode nicht als vollständig angesehen, da nach ihm der Intellekt aus der nach diesen Eliminationen noch übrigbleibenden Mannigfaltigkeit von Gesetzesformen das wahre Gesetz herausfinden muß.[44]

Bacon hat weder bewiesen, daß bei seiner Methode der Eliminierung von konkurrierenden Hypothesen höchstens eine von ihnen übrigbleibt, die die Ursache jener Wirkung nennt, noch, daß damit auch wenigstens eine nicht eliminiert wird, und ich sehe auch nicht, wie dies zu beweisen wäre. Es würde dies ja bedeuten, daß 1. eine Wirkung, wie z. B. Fieber, nicht verschiedene Ursachen haben kann und daß 2. die Grundgesetze der Welt von deterministischer Art sind, was jedoch möglicherweise, wie etwa beim Atomzerfall, nicht zutrifft.

Bacons Methode wird so gut wie immer zu dem Ergebnis führen, daß alle in Frage kommenden Ursachen durch sein Verfahren eliminiert werden. Denn bei Messungen werden in gewissen Abständen immer wieder singuläre Effekte auftreten, die die allgemeine Hypothese widerlegen und die nicht reproduzierbar sind. Der Naturwissenschaft-

---

[42] Bacon [1], I, 130.
[43] Vgl. z. B. auch Carnap [20], S. 40—47.
[44] So kann man Bacon [1], II, 19, 20 verstehen.

ler vernachlässigt dann solche unwiederholbare und unter kein universelles Gesetz subsumierbare Messungen. Damit Bacons Methode überhaupt funktioniert, darf man sie also nicht auf die Beobachtungsdaten anwenden, sondern nur auf einfache Naturgesetze, unter die jene Daten subsumierbar sind. Dann setzt er jedoch zumindest teilweise das voraus, was er gewinnen möchte.

Mit dieser Kritik sollen seine Verdienste nicht geschmälert werden, die er sich bei der Klärung und Analyse der naturwissenschaftlichen Methoden erworben hat. Er hat ja als einer der ersten und lange vor dem naturwissenschaftlichen Aufschwung einen diskussionsfähigen Beitrag zur Frage geliefert, wie man in den Wissenschaften vorzugehen hat, um zu sicheren Ergebnissen zu gelangen, während die Naturphilosophen seiner Zeit die Welt mit zumeist mystischen Vorstellungen zu ergründen suchten.

Mit seinen Methoden zur Eliminierung von falschen naturwissenschaftlichen Hypothesen hat Bacon versucht, einen Weg zu ermitteln, auf dem 1. die wahren Gesetze der Natur zu erlangen und 2. diese Gesetze von ihrem hypothetischen Charakter zu befreien und in gesichertes Wissen zu überführen sind. Man kann die Metaphysik der Erfahrung Immanuel Kants (1724—1804) unter dem gleichen Gesichtspunkt betrachten, nämlich als einen Versuch, 1. die Gesamtheit dieser Methoden systematisch darzustellen und 2. diese Methoden auf apriorischem Weg zu rechtfertigen.[45] Eine *empirische* Rechtfertigung solcher Verfahrensweisen ist nach Kant deswegen ausgeschlossen, weil sie ja überhaupt erst die Erfahrung ermöglichen und deshalb zur Gewinnung *jener* Erfahrungen, durch die sie zu begründen wären, bereits vorausgesetzt werden müßten. Da er aber andererseits klar erkannt hat, daß sie nicht durch bloße Bedeutungsanalyse der Begriffe unter Verwendung der Gesetze der formalen Logik deduzierbar sind, hat er versucht, sie auf einem synthetisch-apriorischen Weg als gültig zu erweisen.[46]

Kants Ansicht, daß weder die Methoden zur Gewinnung von Wirk-

---

[45] Vgl. hierzu auch W. Stegmüller, Gedanken über eine mögliche rationale Rekonstruktion von Kants Metaphysik der Erfahrung, in: Ratio 9 (1967) S. 1 bis 30; 10 (1968) S. 1—31.

[46] Wichtig sind in diesem Zusammenhang insbesondere seine *Analogien der Erfahrung*, vgl. Kritik der reinen Vernunft, B 218 sowie B 252ff.

lichkeitserkenntnissen noch die Grundsätze der Mathematik analytisch sind, wird auch von John Stuart Mill (1806—1873) geteilt; dieser hält sie aber bereits deshalb für synthetisch, weil sie nach seiner Meinung empirische Aussagen sind. Zur Aufhellung von Mills Argumenten für diese These wäre es nützlich, wenn er sich mit den diesbezüglichen Behauptungen Kants gründlich auseinandergesetzt hätte; er scheint sie aber nicht gekannt zu haben, oder aber er hat ihre Bedeutung nicht richtig eingeschätzt. Es kann den Werken Mills auch nicht entnommen werden, ob er die seit 1831 in Oxford befindliche Kopie der in Herculaneum gefundenen Schriften des Philodemus gelesen hat; sicherlich hätte er sie einsehen können, wenn er sich darum bemüht hätte.

Mills Hauptproblem läßt sich folgendermaßen formulieren: Wie kann gezeigt werden, daß die Methoden einer eliminativen induktiven Logik, seien es die Regeln Bacons, sei es eine Modifikation derselben, korrekt und adäquat sind und zu sicheren Naturerkenntnissen führen? Eine apriorische Begründung hält er nicht für möglich (allerdings vertritt er diesen Standpunkt undogmatisch), desgleichen ist es für ihn ausgeschlossen, diese Regeln allein aus unseren Wahrnehmungsberichten zu deduzieren. Die eliminative Induktion kann deshalb nach seiner Ansicht nur unter Verwendung einer Induktion gerechtfertigt werden — selbstverständlich nicht durch die eliminative Induktion selbst, sondern, da für ihn keine andere Wahl übrigbleibt, durch die enumerative.

Ein enumeratives induktives Argument ist nach Mill in erster Linie ein Schluß von Besonderem auf Besonderes, also von Einzelinstanzen auf andere Einzelinstanzen, und zwar genauer von bekannten Einzeltatsachen zu unbekannten singulären Sachverhalten;[47] der Schluß von einer (endlichen) Gesamtheit einzelner Erfahrungsgegebenheiten, etwa der Sterblichkeit von Menschen, auf eine singuläre Behauptung wie „Lord Palmerston ist sterblich" ist genau dann gültig, wenn auch der Schluß aus jenem Datum auf einen Satz gilt, der aus diesem durch Ersetzung des Ausdrucks „Lord Palmerston" durch den Namen eines beliebigen anderen Menschen entsteht, wenn mit anderen Worten also der Schluß auf jeden beliebigen Menschen erlaubt ist und

---

[47] Vgl. Mill [1], Buch II Kap. III §§ 3, 4, Kap. IV § 3; Buch III Kap. II § 1.

wenn damit auf die Behauptung „Alle Menschen sind sterblich" korrekt geschlossen werden darf.[48]

Die Methode der enumerativen Induktion führt nach Mill in den meisten Fällen zu trügerischen Resultaten, deren vorläufige Gültigkeit nur darauf beruht, daß noch kein Gegenbeispiel entdeckt worden ist. Am unsichersten sind ihre Ergebnisse, wenn der zu erschließende Satz nur einen ganz beschränkten Anwendungsbereich besitzt, am sichersten hingegen, wenn der Anwendungsbereich unbegrenzt ist, wie es etwa für die Gesetze der Mathematik gilt. Die Sicherheit solcher Argumente steht also in umgekehrtem Verhältnis zur Allgemeinheit der zu erschließenden Behauptungen.[49]

Die Regeln der eliminativen Induktion werden nun auf folgendem Weg durch die Methode der enumerativen Induktion begründet und damit, grob gesprochen, die Baconsche Induktion durch die Aristotelische gerechtfertigt:

Unter Verwendung der enumerativen Induktion gewinnen wir einzelne Kausalgesetze wie „Feuer brennt", „Speise nährt", „Steine fallen" usw., ohne daß wir dazu ein allgemeines Kausalgesetz verwenden. Diese Generalisationen sind noch roh und müssen später eingeschränkt und modifiziert werden. Trotzdem bilden sie die Grundlage jeglicher Naturerkenntnis; ohne dieses vorwissenschaftliche Wissen gibt es keine Naturwissenschaft.[50]

Die Methode der einfachen Aufzählung (die enumerative Induktion) führt gewöhnlich zu Irrtümern, da sie, wie etwa im Fall der Generalisation „Alle Schwäne sind weiß", von raum-zeitlich beschränkten Beobachtungen ausgeht oder von einer gewissen Kombination von Umständen abhängt, die nicht notwendigerweise immer gegeben sein müssen. Ist aber eine Verallgemeinerung so weit, daß sie keine räumliche oder zeitliche Begrenzung hat und nicht von derartigen Kombinationen abhängen kann, so führt auch die enumerative Induktion zu Sicherheit. Eine Generalisation mit einer solchen Weite ist das

---

[48] Vgl. Mill [1], Buch II Kap. III §§ 4, 5, 7, 8; Buch III Kap. II § 2, Kap. XXI § 4.

[49] Vgl. Mill [1], Buch III Kap. XXI § 3, sowie Buch II Kap. XI § 1. Ich muß gestehen, daß es mir nicht geglückt ist, dem Begriff der Allgemeinheit einer Behauptung einen präzisen und den Absichten Mills entsprechenden Sinn zu geben.

[50] Vgl. Mill [1], Buch III Kap. XXI §§ 3, 4.

allgemeine Kausalprinzip; wie die Gesetze der Mathematik ist es umfangsgleich mit aller menschlichen Erfahrung und kann deshalb in seiner Gültigkeit nicht beschränkt sein. Durch eine enumerative Induktion höherer Art kann deshalb aus jenen einzelnen und vorwissenschaftlich gewonnenen Kausalgesetzen das allgemeine Kausalprinzip erschlossen werden, dessen Einzelfälle sie nach Mill sind.[51]

Seiner Ansicht nach ist dies die einzige Möglichkeit, jenes Prinzip als gültig zu erkennen; denn seine Negation ist denkbar und widerspruchsfrei, ja es ist sogar ein Prinzip der allgemeinen Ursachelosigkeit vorstellbar. Nur die Erfahrung kann lehren, was tatsächlich der Fall ist. Er versucht nun des weiteren zu zeigen, daß die Methoden der eliminativen Induktion genau dann gültig sind, wenn das allgemeine Kausalprinzip nicht verletzt ist.[52] Unter Benützung dieser Methoden wiederum können neue Kausalgesetze (neue Einzelfälle des allgemeinen Kausalprinzips) gefunden und alte modifiziert und berichtigt werden, wodurch jenes Prinzip eine zusätzliche Stütze erhält. Mills Eliminationsmethoden sind, entgegen seiner Versicherung, nicht neu, sondern im Vergleich zu den von Bacon formulierten lediglich etwas detaillierter.[53]

Ein dem allgemeinen Kausalprinzip (dem allgemeinen Sukzessionsgesetz) analoges allgemeines Koexistenzgesetz gibt es nach Mill nicht (was er allerdings nicht näher begründet). Deshalb können die strengen Regeln der Elimination nur auf kausale Abfolgen angewandt werden, während man spezielle Koexistenzgesetze wie „Alle Raben sind schwarz" nur durch die unsichere enumerative Induktion gewinnen kann.[54]

---

[51] Vgl. Mill [1], Buch III Kap. XXI §§ 2, 3, 4; siehe auch Kap. III § 1 und Kap. IV § 1.

[52] Vgl. Mill [1], Buch III Kap. XXI § 1. Im strengen Sinne des Wortes sind Mills Regeln zur Gewinnung von Naturgesetzen nicht eliminativ im Sinne Bacons, da sie nicht von einem Vorrat an Hypothesen ausgehen und daraus alle falschen zu eliminieren trachten. Üblicherweise rechnet man sie jedoch trotzdem dazu, da sie aus einem geeigneten (wenn auch vielleicht relativ kleinen) Vorrat an Beobachtungen (angeblich) wahre Hypothesen zu erschließen und (angeblich) falsche zu verwerfen gestatten und da zusätzliche Beobachtungen, die mit den früheren in Einklang stehen, an der Wahrscheinlichkeit der Hypothesen nichts ändern.

[53] Sie sind in Mill [1], Buch III Kap. VIII zusammengestellt.

[54] Vgl. Mill [1], Buch III Kap. V § 8 sowie Kap. XXII § 4.

Wenn sich die Wissenschaften bis zu einem gewissen Grad entwickelt haben, werden auch Gesetzesaussagen aufgestellt, die man durch die Eliminationsmethoden nicht *auffinden* kann. Für diesen Fall empfiehlt Mill die folgende *deduktive Methode*: Es werden zunächst derartige Aussagen hypothetisch angenommen und Folgerungen aus ihnen deduziert. Stimmen diese Konsequenzen mit der übrigen Erfahrung überein, so können wir jenen Annahmen auch in anderen Fällen vertrauen; liegt keine Übereinstimmung vor, so müssen wir sie, falls nicht andere Ursachen (etwa eine Störung im Meßgerät) dafür verantwortlich sind, verändern oder vervollständigen. Ihres hypothetischen Charakters können sie allerdings nur entkleidet werden, wenn man sie durch die Kriterien der strengen, d. h. eliminativen Induktion überprüft.[55] Wie eine Prüfung derartiger allgemeiner Theorien, die nur einen sehr indirekten Kontakt mit der Wahrnehmung besitzen, vorzunehmen ist und warum diese danach als absolut sicher gelten können, wird allerdings nicht gesagt.

Eine detaillierte Kritik des Rechtfertigungsversuches der eliminativen Induktion von Mill ist nicht einfach und wohl auch nicht möglich, solange seine Formulierungen des allgemeinen Kausalprinzips und seine Methoden nicht genügend präzisiert worden sind. Es scheint mir jedoch, daß die Eliminationsmethoden aus jenem Prinzip nicht logisch folgen, da das Prinzip durch sogenannte *okkulte Effekte* (singuläre Ereignisse, die sich in keine Theorie einfügen lassen, die nicht reproduzierbar sind und die der Naturforscher deshalb unberücksichtigt läßt) nicht widerlegt wird, während durch sie unter Verwendung jener Regeln vermutlich die tatsächlichen Ursachen eliminiert werden. Diese Regeln eliminieren also zuviel, aber gleichzeitig auch zuwenig: eine modifizierte Newtonsche Mechanik kann z. B. bezüglich der Beobachtungsaussagen mit der Einsteinschen Relativitätstheorie gleichwertig sein, ohne daß sie deshalb mit ihr logisch äquivalent ist; keine dieser Theorien wird durch die Verfahren Mills als falsch erkannt. Ferner ist seine Unterscheidung von größeren und geringeren Weiten von Generalisationen wohl nicht präzisierbar und deshalb logisch unhaltbar, womit ein wichtiges Glied aus seiner Argumentationskette herausgebrochen ist.

---

[55] Vgl. Mill [1], Buch III Kap. XIII § 7 sowie auch Kap. XI § 3.

Genauso problematisch ist seine Begründung des allgemeinen Kausalprinzips durch einzelne vorwissenschaftliche Kausalgesetze. Unsere vorwissenschaftliche Naturerkenntnis besteht nicht ausschließlich aus Kausalgesetzen, sondern kennt auch rudimentäre Gesetze der zeitlichen Aufeinanderfolge wie etwa „Wer aus dem fünften Stockwerk eines Gebäudes fällt, stirbt gewöhnlich innerhalb der nächsten Tage", und sie enthält auch Fälle, in denen Gesetzmäßigkeiten irgendeiner Art nicht zu entdecken sind und die also Beispiele für ein allgemeines Prinzip der Ursachelosigkeit sind. Schließlich ist nicht ganz klar, in welchem genauen Sinn man überhaupt sagen kann, daß die einzelnen Kausalgesetze Einzelfälle des allgemeinen Kausalprinzips sind, da jedes von ihnen mehr aussagt als dieses und also keines von ihnen aus diesem logisch folgt. Die Namen „spezielles Kausalgesetz" und „allgemeines Kausalprinzip" sind hier irreführend gewählt.

Der Ablauf des Naturgeschehens läßt sich nach Mill als ein Gewebe der verschiedensten Kausalgesetze auffassen, wobei sich die einzelnen Ursachen überlagern und sich dabei je nach Lage der Dinge erhalten, abschwächen oder ganz aufheben; auf diese Weise entsteht der Eindruck, daß in den einzelnen Gegenstandsbereichen bezüglich verschiedener Eigenschaften verschiedene Grade an Uniformität bestehen. Die eliminative Induktion liefert die Verfahren zur Auflösung solcher Überlagerungen.[56]

Wann Uniformität vorhanden ist und wann nicht, und ob in dem einen Fall die Eigenschaften ein größeres Maß an Homogenität besitzen als in einem anderen, kann nach Mill nur durch Erfahrung entschieden werden. Nur durch Erfahrung wissen wir etwa, daß die Tiere der einzelnen Arten in ihren Farben nicht so uniform sind wie im Bau ihrer Hauptgliedmaßen. Deshalb ist auf Grund unserer Erfahrungen das Zeugnis des Plinius, daß ein Mensch den Kopf zwischen den Schultern getragen hat, unwahrscheinlich, während man die Behauptung, daß es graue Raben gibt, selbst dann als wahrscheinlich ansieht, wenn all die vielen Raben, die man bisher beobachtet hat, schwarz gewesen sind. So kann man, wenn ein Bereich bezüglich einer Eigenschaft uniform ist, aus einer einzigen positiven Instanz (bei keiner negativen) auf die entsprechende universelle Behauptung

---

56 Vgl. Mill [1], Buch III Kap. IV § 1 und Kap. XVII § 3.

schließen, während man im anderen Fall eine unüberschaubare Menge von übereinstimmenden Instanzen besitzt und trotzdem die dazugehörige allgemeine Aussage als durch dieses Datum nicht gesichert ansieht. Der Wahrscheinlichkeitsgrad, mit dem man aus einer gegebenen Anzahl von (ausschließlich positiven) Fällen aus einer gewissen Menge auf ein entsprechendes allgemeines Gesetz schließen kann, hängt von dem Grad der Uniformität ab, den die Gesamtheit der Klassen, zu der auch diese Menge gehört, bezüglich der fraglichen Eigenschaften besitzt.[57] Die hier von Mill benützten Begriffe „wahrscheinlich" und „sicher" bezeichnen ohne Zweifel einen enumerativen Wahrscheinlichkeitsbegriff. Die Ähnlichkeit seiner Formulierungen mit denen der Epikureer ist frappant.

Ein Kausalgesetz kann als die vollständige Beschreibung der Umstände charakterisiert werden, unter denen ein gewisses Ereignis eintritt. Sind zwei Ereignisse hingegen durch Zufall miteinander verbunden, so ist damit gezeigt, daß die Umstände, die zu dem letzteren führen, nicht vollständig bestimmt oder bekannt sind. Dieser Zufall kann nach Mill dadurch *a posteriori* ausgeschieden werden, daß man eine gewisse Anzahl von Beobachtungen oder Versuchen anstellt; denn, wie die *Erfahrung* lehrt, wird nach einer *hinreichend großen Anzahl von Fällen* der Irrtum auf der einen Seite durch den Irrtum auf der anderen *ungefähr* ausgewogen. Nimmt man z. B. als langfristige Ursache die Höhe der Sonne über dem Horizont im Sommer und als kurzfristige Ursachen, die die langfristigen überdecken, warme oder kalte Winde, Wolken und Regen, so können diese das Ergebnis in der einen oder anderen Richtung verfälschen. Nimmt man hingegen den Mittelwert aus einer genügend großen Anzahl von Jahren, so werden damit diese Nebenursachen ausgeschieden; dies ist dann der Fall, wenn der Mittelwert praktisch unverändert bleibt (d. h. sich nur innerhalb eines kleinen Intervalls ändert), falls noch die Daten einiger weiterer Jahre (wievieler Jahre?) hinzugefügt werden. Dann ist die *Induktion* nach Mill zulässig.[58] Eine Beobachtung oder ein Experiment ist solange zu wiederholen, bis man diesen Mittelwert erreicht hat, wobei die Fehlergrenze durch die Zwecke der Untersuchung bestimmt ist.

---

[57] Vgl. Mill [1], Buch III Kap. III §§ 2, 3; Kap. IV §§ 1, 2; Kap. XVII § 3; Kap. XXIII §§ 1, 2.

[58] Vgl. Mill [1], Buch III Kap. XVII §§ 2, 3.

In diesem Sinn kann die Wahrscheinlichkeit dafür, daß beim Münzwurf die eine Seite (etwa „Kopf") oben liegt, bestimmt werden: Bei einer normalen Münze wird, wenn man nur lange genug wirft, etwa in der Hälfte der Fälle dieses Ergebnis eintreten. Umgekehrt kann man die Irregularität einer Münze (etwa daß eine Seite wesentlich schwerer ist) dadurch nachweisen, daß man nach einer genügend großen Anzahl von Würfen einen Mittelwert entdeckt, der verschieden ist von dem Wert $1/2$, den man ohne irgendein Wissen erwartet hätte.[59]
Trotz seiner sonst teilweise subtilen Untersuchungen hat Mill nicht bemerkt, daß neben der eliminativen Induktion auch der enumerative Induktionsschluß einer Rechtfertigung bedarf und daß er vor allem nicht voraussetzungslos gültig ist. Er hat die Äußerungen Humes zu diesem Punkt nicht gelesen oder aber deren Bedeutung für das Induktionsproblem nicht erkannt. Es ist daher in einer Darstellung der historischen Entwicklung des Induktionsbegriffs angezeigt, Hume *nach* Mill zu referieren.

3. Über die Begründung des induktiven Schließens

David Hume (1711—1776) ist der erste und über zwei Jahrhunderte lang auch der einzige[60] gewesen, der systematisch über die Begründung des induktiven Schließens reflektiert hat; deshalb ist es auch heute noch nützlich, sich seine diesbezüglichen Argumente zu vergegenwärtigen. Ich möchte versuchen, seine Grundlegung der Wahrscheinlichkeitslehre im Hinblick auf seine philosophischen Standpunkte (Empirismus, Phänomenalismus, Skeptizismus) zu schildern,

---

[59] Vgl. Mill [1], Buch III Kap. XVII § 5; Kap. XVIII § 3. In späteren Auflagen hat Mill die Wahrscheinlichkeit nicht mehr den *Ereignissen* zugesprochen, wie hier eindeutig zum Ausdruck kommt, sondern hat, im Anschluß an Laplace, die Wahrscheinlichkeit als *Grad der Gewißheit* eines Ereignisses *relativ zu einem Wissen* verstanden; vgl. Buch III Kap. XVIII § 1. In § 2 dieses Kapitels gibt er als Kriterium dafür, daß relativ zu einem Wissen ein Ereignis und dessen Gegenteil die gleiche Wahrscheinlichkeit besitzen, an, daß man Wetten mit gleichen Einsätzen für oder gegen das zu erwartende Ereignis annehmen kann.

[60] Erst Popper [1] und vor allem Carnap [11] haben wieder nennenswerte Beiträge zum Problem der Begründung des induktiven Schließens erbracht.

indem ich die syntaktische Struktur dieser Schlüsse und ihre Abgrenzung von deduktiven Argumenten schildere, die verschiedenen von ihm angegebenen Arten von Wahrscheinlichkeitsargumenten charakterisiere und zeige, auf welchen Voraussetzungen nach seiner Ansicht alles nichtdeduktive Schließen beruht und auf Grund welcher Prinzipien wir so schließen. Dabei möchte ich an einigen Stellen seine im „Treatise" gegebene Darstellung mit der des „Enquiry" vergleichen, die mit jener zwar in den wesentlichen Grundgedanken übereinstimmt, von ihr aber in einigen interessanten Details abweicht.

Um die Einflüsse seiner philosophischen Standpunkte klar voneinander trennen zu können, ist es notwendig, diese wenigstens in groben Umrissen zu charakterisieren. Empirist kann man in zweifacher Hinsicht sein: bezüglich der Erfahrungs*begriffe* und bezüglich der Erfahrungs*urteile*. Man kann demnach behaupten oder voraussetzen, daß alle Begriffe, die nicht der Logik oder der Mathematik angehören, aus der Erfahrung durch irgendeine Art der Abstraktion gewonnen worden sind oder gewonnen werden können, und man kann davon ausgehen, daß alle Aussagen über die Welt, alle Sätze also, die nicht in den Bereich der Logik und der Mathematik fallen, nur durch singuläre Erfahrungssätze begründbar sind. Beide Positionen sind voneinander logisch unabhängig, obwohl es natürlich naheliegt, beide gleichzeitig oder keine von beiden zu vertreten: Man kann, wie etwa Franz Brentano, der Ansicht sein, daß alle Erfahrungsbegriffe aus der Erfahrung gewonnen werden, obwohl es wahre Sätze über die Welt gibt, die man auch anders begründen kann als nur durch Bezugnahme auf singuläre Erfahrungstatsachen, und man kann widerspruchsfrei dazu annehmen, daß wahre synthetische Aussagen (die man also nicht durch Analyse der in ihnen vorkommenden Begriffe als wahr erkennen kann) nur durch Erfahrung begründet werden können, obwohl bestimmte Grundbegriffe der Erfahrung, wie etwa der Begriff der Wirkungsrelation, mit dem man die Eigenschaften des Raumes und der Zeit beschreiben kann[61], nicht durch Abstraktion aus der Erfahrung zu gewinnen sind.

Nach Ansicht der Phänomenalisten ist unsere Welt eine Welt der Erscheinungen, genauer: baut sich unsere Welt aus Erscheinungen auf,

---

61 Vgl. dazu Essler [1], S. 228—233 und die dort genannten Werke.

noch genauer: sind alle Behauptungen über unsere Welt Behauptungen über Erscheinungen, noch genauer: kann die Sprache, in der wir über die Dinge unserer Welt reden, sinnvoll nur durch Erscheinungen und deren Eigenschaften und Beziehungen interpretiert werden. Obwohl es naheliegend ist, mit dem Standpunkt des Phänomenalismus auch den des Empirismus in seinen zwei Bedeutungen anzunehmen, sind beide doch voneinander logisch unabhängig. Man kann ohne Schwierigkeit aus der Kantischen Metaphysik der Erfahrung durch Weglassung seiner Aussagen über das *Ding an sich* eine zwar phänomenalistische, aber weder bezüglich der Begriffe noch hinsichtlich der Urteile empiristische Philosophie gewinnen. Daneben hat es auch viele empiristisch eingestellte Philosophen gegeben, die nicht vom phänomenalistischen Ansatzpunkt ausgegangen sind.

Logisch unabhängig von all diesen Einstellungen ist schließlich die des Skeptikers, der es nicht für erwiesen hält, daß wir an irgendeinem Punkt unserer Bemühungen zu definitivem Wissen im Aristotelischen Sinn gelangen können. Auch hier genügt der Hinweis darauf, daß weder alle Empiristen noch alle Phänomenalisten Skeptiker in diesem Sinne sind, und daß im Prinzip auch ein Rationalist Skeptiker sein kann, der daran glaubt, daß alle wahren Behauptungen aus dem Satz vom ausgeschlossenen Widerspruch folgen, und der trotzdem davon überzeugt ist, daß sich der Mensch immer irren kann.

Hume hat alle drei Positionen vertreten. Sein Phänomenalismus und sein Skeptizismus spielen vor allem bei der *Begründung* der Wahrscheinlichkeitsschlüsse eine Rolle, während ihn sein Empirismus, soweit er die Frage der Rechtfertigung von Erfahrungsurteilen betrifft, zu der Folgerung veranlaßt, daß Wahrscheinlichkeitsschlüsse aller Art, einschließlich der sichersten, *keine deduktiven Argumente* sind. In seinem „Treatise" rechnet er die Grundsätze der Arithmetik zur Logik und verweist darauf, daß die Axiome der Geometrie nur durch Erfahrung begründet werden können, während er im „Enquiry" erklärt, daß die Grundsätze der Geometrie, der Arithmetik und der Algebra insgesamt analytisch wahre Behauptungen sind, da sie nichts über unsere reale Welt aussagen, sondern nur Beziehungen zwischen Figuren, zwischen Zahlen und zwischen Größen ausdrücken.[62]

---

[62] Vgl. Hume [1], S. 70f., und Hume [2], S. 35.

Hume teilt die Denkakte oder Argumente an einigen Stellen in *demonstrative*, die Beziehungen zwischen Vorstellungen betreffen, und in *moralisch-gewisse oder wahrscheinliche* ein, die Tatsachen und Dasein betreffen.[63] Demonstrative Gründe sind also logische Folgebeziehungen, und demonstrative Sätze Behauptungen, die aus den logischen Prinzipien (den *intuitiv*-gültigen Sätzen) bewiesen werden können. Sätze und Argumente sind demonstrativ oder logisch gültig dann und nur dann, wenn ihr Gegenteil einen Widerspruch enthält und somit nicht klar gedacht und vorgestellt werden kann.[64] Wahrscheinlichkeitsgründe umfassen dann jene Argumente, die nicht demonstrativ sind, deren Gegenteil klar vorstellbar ist und keinen Widerspruch enthält; ein solcher Schluß führt über den Gehalt der Prämisse hinaus und kann deshalb nicht jenen Grad an Sicherheit haben, der den demonstrativen Argumenten zukommt.[65] Diese Charakterisierung der Wahrscheinlichkeitsgründe als Schlüsse, die einen gewissen Grad an Sicherheit besitzen, ohne daß ihre Negation einen Widerspruch enthält, benützt Hume nicht sehr häufig.

Einige der nichtdemonstrativen Argumente können als sicher angesehen werden, andere nicht. In Anlehnung an den vorwissenschaftlichen Sprachgebrauch sagt Hume, daß jene *sicheren Gründe* ein *Nachweis* (proof) sind, während die übrigen nur in einem engeren Sinn als *wahrscheinlich* angesehen werden können. Ein Nachweis ist ein Argument, dessen Gegenteil zwar klar denkbar ist, aber nicht ernsthaft behauptet werden kann, ein Schluß also, der praktisch von Unsicherheit frei ist (wie z. B. „Steine, die man in einer gewissen Höhe über der Erdoberfläche losgelassen hat, sind immer zu Boden gefallen; folglich wird auch dieser Stein, wenn man ihn in einer gewissen Höhe über der Erdoberfläche losläßt, zu Boden fallen"). Der Nachweis hat also fast den gleichen (und für praktische Zwecke genau den gleichen) Grad an Sicherheit wie die Demonstration, deren Gegenteil einen Widerspruch enthält und nicht klar vorgestellt werden kann. Hinge-

---

63 Vgl. Hume [1], S. 31, und Hume [2], S. 35; in [1] benützt er auf den Seiten 88f., 103 und 108 den Ausdruck „Wahrscheinlichkeitsschluß" als gleichbedeutend mit „nichtdemonstratives Argument". Moralisch-gewiß ist ein Grund in diesem Zusammenhang offenbar, wenn mit der Behauptung, daß man ihm vertrauen dürfe, bewußt nichts Falsches gesagt wird.

64 Vgl. Hume [1], S. 161 und 463, sowie Hume [2], S. 18, 25f. und 35.

65 Vgl. Hume [1], S. 95, und Hume [2], S. 25f., 56, 58.

gen ist das Gegenteil eines Wahrscheinlichkeitsarguments (im engeren Sinn) nicht nur klar vorstellbar, sondern sogar *möglich*, da man, aus welchen Gründen auch immer, damit noch bis zu einem gewissen Grad rechnet.[66] In der Mehrzahl der Fälle verwendet er den Wahrscheinlichkeitsbegriff in diesem Sinn.

Syntaktisch gesehen sind nach Hume Wahrscheinlichkeitsschlüsse (einschließlich der sicheren Argumente, die ja Nachweise sind) stets Schlüsse von Bekanntem auf Unbekanntes, und zwar im allgemeinen von Vergangenem oder Gegenwärtigem auf ein singuläres zukünftiges Ereignis, etwa von der Tatsache, daß bisher jeden Tag die Sonne aufgegangen ist, auf die Behauptung, daß sie auch morgen aufgehen wird. An einigen wenigen Stellen schließt er jedoch auf eine größere Klasse von Ereignissen, etwa daraus, daß die Welt bisher gleichförmig war, auf die Hypothese, daß sie es auch in der Zukunft sein wird, oder auf alle Ereignisse einer bestimmten Art, etwa aus der Tatsache, daß bisher alle Menschen sterblich waren, auf das Gesetz, daß alle Menschen sterblich sind. Offensichtlich sind auch für ihn diese Schlüsse äquivalent, d. h. ist einer von ihnen berechtigt, so auch die anderen.

Während Hume in seinem „Enquiry" lediglich bemerkt, daß die Sicherheit der Wahrscheinlichkeitsargumente die verschiedensten Gradabstufungen haben kann,[67] kennt er in seinem „Treatise" fünf verschiedene Typen von nichtdemonstrativen Gründen:

(1) die Kausalschlüsse, die er nur gelegentlich unter die Wahrscheinlichkeitsargumente subsumiert: vorausgesetzt wird hier, daß in einer *genügend großen Anzahl* von Beobachtungen Ereignisse einer gewissen Art durchgehend und ausnahmslos von Ereignissen einer anderen Art begleitet waren, die sich mit jenen raum-zeitlich berührten und auf sie zeitlich folgten; daraus wird geschlossen, daß ein neues Ereignis, das jenen der ersten Art gleich ist, wiederum von einem der zweiten Art begleitet sein wird, wobei die Sicherheit dieses Schlusses einer Überzeugung, daß dies nicht eintreten könnte, keinen Raum läßt;[68]

(2) die Wahrscheinlichkeitsschlüsse (der ersten Art), die auf einem

[66] Vgl. Hume [1], S. 124 und 135, sowie Hume [2], S. 56; vgl. auch Hume [1], S. 133f., 135, 136f.

[67] Vgl. Hume [2], S. 110.

[68] Vgl. Hume [1], S. 170, sowie S. 75f., 86ff., 134, 163, 171, 173. In seinem „Essay" verzichtet er auf die Bedingungen des raum-zeitlichen Zusammenhangs, vgl. Hume [2], S. 77f.

Mangel an Erfahrungsdaten beruhen: wenn das Erfahrungsdatum nicht groß genug ist, kann der Schluß nur mit einer gewissen Wahrscheinlichkeit behauptet werden, die bei einer einzigen Beobachtung noch nicht wesentlich vom Nichtwissen verschieden ist und die sich bei hundert gleichförmigen Beobachtungen nicht mehr von jener Gewißheit unterscheidet, die den Kausalschlüssen zukommt;[69]
(3) die Wahrscheinlichkeitsschlüsse (der zweiten Art), die darauf beruhen, daß die Beobachtungen nicht uniform sind und Gegeninstanzen enthalten: auch hier sind die verschiedensten Gradabstufungen der Sicherheit möglich, je nach dem Verhältnis der positiven zu den negativen Instanzen; wenn unter hundert Fällen nur ein Gegenbeispiel vorkommt, so gewährt uns dieser Schluß fast Sicherheit, daß auch in der Zukunft keine Gegeninstanz vorkommen wird;[70]
(4) die Wahrscheinlichkeitsschlüsse (der dritten Art), die darauf beruhen, daß ein Ereignis vorliegt, das den vergangenen Ereignissen nur ähnlich, aber nicht völlig gleich ist: auch hier kann nur mit einer gewissen Wahrscheinlichkeit darauf geschlossen werden, daß es von einem Ereignis der anderen Art begleitet sein wird, und der Grad der Wahrscheinlichkeit ergibt sich aus dem Grad der Ähnlichkeit (Argumente dieser Art nennt Hume „Analogieschlüsse");[71] und schließlich:
(5) die Wahrscheinlichkeitsschlüsse (der vierten Art), die vollständig auf Zufall beruhen, so daß, wie etwa beim Würfelwurf, die bisherigen Beobachtungen nicht berücksichtigt zu werden brauchen: die einzelnen Möglichkeiten werden als gleich wahrscheinlich angesehen, da kein Grund vorhanden ist, eine von ihnen höher zu bewerten als die anderen; eine größere Wahrscheinlichkeit ergibt sich lediglich dann, wenn man verschiedene dieser Möglichkeiten zusammenfaßt und sie den restlichen gegenüberstellt, wenn man etwa von der Wahrscheinlichkeit spricht, daß eine der ersten vier Zahlen beim kubischen Würfel oben liegt. Auch hier sind die verschiedensten Gradabstufungen möglich; die Wahrscheinlichkeit etwa, daß bei 1000 gleichmöglichen Fällen ein Ereignis unter eine der ersten 999 Möglichkeiten fällt, ist kaum noch von der Sicherheit zu unterscheiden.[72]

---

69 Vgl. Hume [1], S. 130f.
70 Vgl. Hume [1], S. 131—135, und Hume [2], S. 57f.
71 Vgl. Hume [1], S. 142, und Hume [2], S. 104.
72 Vgl. Hume [1], S. 125—130, und Hume [2], S. 56—58.

Mischformen dieser fünf Arten von Argumenten kennt er nicht, obwohl es an einigen Stellen so scheinen will, als ob er die Wahrscheinlichkeit der zweiten und die der vierten Art zusammenfassen möchte.
Die Frage, ob es einen Zufall im Sinne der Regellosigkeit gibt, läßt er im „Treatise" offen, während er sie im „Enquiry" unmißverständlich verneint. Alle Ereignisse sind demnach unter ausnahmslos gültige Regelmäßigkeiten subsumierbar, und insbesondere können alle Aufeinanderfolgen von Ereignissen durch Kausalgesetze beschrieben werden. Der Eindruck des Zufalls entsteht lediglich dadurch, daß wir die möglichen Ursachen eines Ereignisses nicht kennen, und dieses Nichtwissen hat bezüglich unserer Wahrscheinlichkeitsschlüsse den gleichen Effekt wie die Annahme, daß das fragliche Ereignis durch Zufall entstanden ist.[73] Von Zufall kann dann, aber auch nur dann gesprochen werden, wenn die einzelnen in Frage kommenden Möglichkeiten gleich wahrscheinlich sind; ist dies nicht der Fall, so nehmen wir an, daß die Wirksamkeit einer Ursache durch die einer ihr widerstreitenden teilweise verhindert wird.[74] So wie man im Fall eines Wahrscheinlichkeitsschlusses der vierten Art von der Gleichwahrscheinlichkeit der einzelnen Möglichkeiten ausgehen muß, so ist es bei den übrigen Wahrscheinlichkeitsgründen einschließlich der Kausalschlüsse notwendig, daß man jeder der einzelnen Beobachtungen, auf die man Bezug nimmt, das gleiche Gewicht beimißt und die Wahrscheinlichkeit lediglich von der Anzahl der Fälle abhängen läßt, die für die Annahme, daß ein Ereignis einer gewissen Art eintritt, positiv sind.[75]
Die Wahrscheinlichkeitsschlüsse der vierten Art beruhen nach Hume also darauf, daß auf Grund von Naturgesetzen gewisse Experimente (wie etwa der Wurf eines Würfels) genau ein Ergebnis aus einer endlichen Anzahl von möglichen Ergebnissen haben, wobei wegen unserer völligen Unkenntnis der Bedingungen, unter denen das Experiment abläuft (unter denen der Würfel fällt), die einzelnen möglichen Ergebnisse als zufällig und miteinander gleichwahrscheinlich angesehen werden müssen. Da man die Abläufe des Geschehens als zufällig

---

[73] Vgl. Hume [2], S. 35, 56. Vgl. hierzu die vorsichtigen Bemerkungen in Hume [1], S. 130, 132.
[74] Vgl. Hume [1], S. 125f., 132, und Hume [2], S. 58.
[75] Vgl. Hume [1], S. 136.

voraussetzt, brauchen die vergangenen Erfahrungen bezüglich dieser Experimente nicht berücksichtigt zu werden.

Es erhebt sich nun die Frage, wie nach Hume die anderen Wahrscheinlichkeitsargumente einschließlich der Kausalschlüsse zu rechtfertigen sind und auf welchem Prinzip sie beruhen. Die demonstrativen Gründe beruhen ja darauf, daß ihr Gegenteil einen Widerspruch enthält; für die moralisch-gewissen oder wahrscheinlichen Gründe ist es jedoch wesentlich, daß ihr Gegenteil denkbar und also widerspruchsfrei ist. Was gibt uns dann das Recht, sie trotzdem zu benützen?

Mit diesen Argumenten wird nach Hume erschlossen, daß die Zukunft der Vergangenheit gleich oder ähnlich ist, je nachdem, ob die Anzahl der Beobachtungen genügend groß ist oder nicht. Dieser Schluß ist nicht demonstrativ und also nicht logisch gültig, denn es liegt kein Widerspruch in der Annahme, daß sich die Naturabläufe ändern und die Zukunft der Vergangenheit deshalb *nicht mehr* ähnlich ist. Mit der biologischen Entwicklung des Menschengeschlechts etwa ändern sich auch die Gesetze der Psychologie, und mit der kulturellen Entwicklung auch die der Soziologie und die der anderen Wissenschaften vom Menschen, und was in einigen Gebieten des menschlichen Wissens der Fall ist, ist in anderen zumindest möglich.[76] Damit der Schluß gültig wird, muß also vorausgesetzt werden, daß die Zukunft mit der Vergangenheit (bezüglich der Eigenschaften, auf die sich Prämisse und Konklusion des Schlusses beziehen) uniform ist; dann, aber auch nur dann, ist man berechtigt, ihn zu benützen.

Wahrscheinlichkeitsschlüsse sind also *Schlüsse aus Annahmen*,[77] denn sie beruhen teils auf der Annahme, daß die Ereignisse zufällig sind, daß wir bezüglich ihrer Ursachen in völliger Unkenntnis sind und die einzelnen Möglichkeiten daher als gleichwahrscheinlich betrachten, und teils auf der Annahme, daß die Zukunft der Vergangenheit ähnlich oder gleich ist, entsprechend der Größe unseres Erfahrungsdatums. Wie ist aber die Annahme der Uniformität der Natur und damit die Projektion der in der Vergangenheit vorgefundenen Ereignisse auf die Zukunft zu rechtfertigen? Denn der Schluß aus einer ungerechtfertigten Annahme führt nicht notwendigerweise zu einer sicheren Konklusion.

[76] Vgl. Hume [1], S. 92, 102, 104, 125, und Hume [2], S. 27—30, 33—38, 42f.
[77] Vgl. Hume [1], S. 124.

Hier ergibt sich nun für Hume ein Dilemma, das er folgendermaßen formuliert:[78] Ein Denkakt ist entweder demonstrativ oder wahrscheinlich. Durch ein demonstratives Argument kann jene Annahme aber nicht als wahr erwiesen werden, denn ihr Gegenteil enthält ja keinen Widerspruch und ist klar vorstellbar. Es kann sogar der Fall eintreten, daß wir gute Gründe dafür haben, sie zu *verwerfen*; in diesem Fall können wir aus den vergangenen Beobachtungen überhaupt nichts mehr erschließen. Rational ist diese Annahme also nicht zu begründen. Wahrscheinlichkeitsgründe hingegen beruhen zweifellos teilweise auf Erfahrung; denn ohne Erfahrungen können wir nicht wissen, welche Ereignisse aufeinanderfolgen und welche nicht. Weder die begriffliche Analyse der Dinge noch die ihrer Eigenschaften verhilft uns zu diesem Wissen, denn dadurch gewinnen wir ja nur eine detaillierte Kenntnis der Sachverhalte, nicht jedoch der Aufeinanderfolge von Sachverhalten. Andererseits genügt Erfahrung jedoch nicht, um auf die Zukunft zu schließen. Wir haben zwar in der Vergangenheit ständig oder häufig gewisse Ereignisse miteinander verknüpft vorgefunden, wir können daraus jedoch nicht *logisch* schließen, daß dies auch in der Zukunft so sein wird, da ja das Gegenteil möglich ist; wir können freilich mit einer *gewissen Wahrscheinlichkeit* darauf schließen, daß dies so sein wird, was uns aber nicht weiterhilft, da hierbei ja die zu erschließende Annahme *vorausgesetzt* werden muß.

Man könnte versucht sein, diesem Dilemma mit der Behauptung zu entgehen, daß man zumindest die Gültigkeit der Kausalschlüsse auf die Wirkung gewisser geheimer Kräfte in den Dingen zurückführen könne; die übrigen Wahrscheinlichkeitsschlüsse wären dann vielleicht durch die Annahme verschiedener Ursachen, die sich teilweise oder ganz aufheben, zu rechtfertigen. Dieser Ausweg ist nach Hume aus zwei Gründen nicht gangbar: Erstens ist eine solche geheime Kraft nicht beobachtbar und kann deshalb nicht in einer gesicherten Theorie vorausgesetzt werden; in jedem Einzelfall sehen wir nur die Aufeinanderfolge zweier Ereignisse, etwa den Zusammenstoß zweier Billardkugeln, nicht jedoch eine Kraft, die bewirkt, daß die eine Kugel stehenbleibt und die andere fortrollt, und wenn wir einen Kausalschluß

[78] Vgl. Hume [1], S. 89f., und Hume [2], S. 34, 37f. Hume war offenbar der erste, der dieses Dilemma klar gesehen hat, denn die Epikureer haben ja verzweifelt versucht, diesem zu entrinnen.

nach hundert derartigen Einzelbeobachtungen als gerechtfertigt ansehen, so kann er nicht auf der Annahme der Existenz einer solchen Kraft beruhen. Zweitens würde die Hypothese von der Existenz geheimer Kräfte aber selbst dann keine Lösung des gegenwärtigen Dilemmas bieten, wenn solche Kräfte beobachtbar wären. Denn wir könnten ja nur sagen, daß *bisher* an den Gegenständen derartige Kräfte gewirkt haben; daraus allein kann jedoch nicht gefolgert werden, daß dies auch *künftig* so sein wird.[79]

Weder Vernunftgründe noch Erfahrungen können also die Annahme stützen, daß die Zukunft mit der Vergangenheit uniform ist. Welches Prinzip, so fragt Hume, das genauso stark ist wie Verstand und Erfahrung, da es unseren Wahrscheinlichkeitsschlüssen doch ein so großes Maß an Gewißheit gibt, liegt ihr dann zugrunde? Seine Antwort lautet: die Gewohnheit.[80] Wir haben uns daran *gewöhnt*, daß gewisse Ereignisse ständig oder häufig miteinander verknüpft sind, und *glauben* und *erwarten* deshalb, diese ständige oder häufige Verknüpfung auch in der Zukunft wiederzufinden. Wenn unser Erfahrungsdatum genügend umfangreich und bezüglich gewisser Verknüpfungen durchgehend gleichförmig ist, so hat unsere Gewöhnung ein Höchstmaß erreicht, und wir erwarten mit einem nicht mehr zu überbietenden Grad an Gewißheit, auch in der Zukunft diese Verhältnisse wiederzufinden. Ist unser Erfahrungsdatum zwar genügend umfangreich, doch nicht ohne Gegeninstanzen, etwa daß von zwanzig ausgelaufenen Schiffen nur neunzehn wieder heimgekehrt sind, so erwarten wir dieses Verhältnis mit größtmöglicher Sicherheit auch für die nächsten zwanzig auslaufenden Schiffe. Für das nächste auslaufende Schiff allein allerdings kann ich bezüglich der Frage seiner Rückkehr nichts mit Gewißheit vorhersagen, sondern lediglich mit einer Wahrscheinlichkeit, deren Grad sich nach der beobachteten relativen Häufigkeit bemißt. Die Gewohnheit, mit der ich hier schließe, wird durch die Gegeninstanzen abgeschwächt, wobei es für den Grad der Abschwächung gleichgültig ist, ob ich dafür mir unbekannte Ursachen verantwortlich mache und an der Gültigkeit des allgemeinen Kausalprinzips festhalte oder ob ich dies nicht tue.[81]

---

79 Vgl. Hume [1], S. 81, und Hume [2], S. 42.
80 Vgl. Hume [1], S. 102, 118, 128, 134, und Hume [2], S. 28f., 43—46, 54, 58.
81 Vgl. Hume [1], S. 134, und Hume [2], S. 57f.

Daß nur die Gewohnheit die Grundlage dieser Annahme über die Gleichheit oder Ähnlichkeit von Vergangenheit und Zukunft sein kann, erhellt sich nach Hume auch daraus, daß jene Wahrscheinlichkeit der ersten Art, die auf einem Mangel an Beobachtungen beruht, durch zusätzliche Beobachtung in Sicherheit übergeht und daß diese verschiedenen Grade an Sicherheit nur auf verschiedene Grade der Gewöhnung zurückgeführt werden können. Diese Gewöhnung nimmt ab, wenn nicht mehr ein mit den vergangenen Ursachen *gleicher*, sondern nur noch *ähnlicher* Gegenstand vorliegt, und zwar nimmt sie im gleichen Grad ab wie diese Ähnlichkeit.[82]

Daß auch unsere Kausalurteile auf Gewohnheit beruhen, leuchtet nach Hume sofort ein, wenn man an Phänomene denkt, die man nicht erwartet hat, wie etwa an zwei glattgeschliffene Marmorplatten, die, aneinandergedrückt und waagrecht gehalten, aneinander haften und nur durch größere Anstrengungen zu trennen sind, während sie sich seitwärts leicht verschieben lassen. Bei den selbstverständlichen Kausalschlüssen ist die Gewöhnung nach seiner Ansicht nur deshalb nicht zu bemerken, weil sie tatsächlich im höchsten Grad vorhanden ist, so daß man sie nicht mehr wahrnimmt und nach anderen Grundlagen für die Kausalschlüsse sucht. Solche sind jedoch, bei genauer Analyse dieser Argumente, nicht zu entdecken.[83]

Nun schließen wir jedoch im Alltag wie auch in den Wissenschaften häufig bereits aus einer einzigen Beobachtung auf ein entsprechendes Kausalgesetz. Hier kann man von einer Gewöhnung als der Grundlage dieses Denkaktes scheinbar nicht sprechen. Tatsächlich ist jedoch nach Hume auch hier die Gültigkeit nicht anders als durch Gewohnheit zu rechtfertigen, und zwar auf folgendem Umweg: Für das allgemeine Prinzip „Gleiche Anfangsbedingungen führen unter gleichen Umständen zu gleichen Ergebnissen" haben wir Millionen von positiven Instanzen (und können, wie man hinzufügen müßte, bei scheinbar negativen Instanzen praktisch immer zeigen, daß nicht gleiche Umstände vorgelegen haben).[84] Hat man also ein Experiment sorgfältig vorbereitet und kann sicher sein, alle störenden Umstände beseitigt zu haben, so kann man mittels jenes Prinzips, das man als durch die

---

[82] Vgl. Hume [1], S. 142, und Hume [2], S. 104.
[83] Vgl. Hume [2], S. 28f.
[84] Vgl. Hume [1], S. 104f., 131, und Hume [2], S. 107.

Erfahrung gesichert annimmt und an das man sich im höchsten Grad gewöhnt hat, aus einer einzigen Instanz mit Sicherheit auf beliebige andere Instanzen und damit auf das entsprechende Gesetz schließen.
Aus Humes empiristischer Grundeinstellung heraus ergibt sich also, daß nur die Gewöhnung zur Rechtfertigung der Annahme der Gleichförmigkeit von Vergangenheit und Zukunft in Frage kommt, da alle anderen Versuche ihrer Begründung fehlschlagen müssen, insbesondere ihr logischer Beweis und ihre logische Ableitung aus den Aussagen über die bisherigen Beobachtungen. Eine eher positive Begründung dafür, daß Gewohnheit und nichts anderes den Wahrscheinlichkeitsschlüssen aller Art zugrundeliegt, hat er durch seine phänomenalistische Philosophie gegeben.
Begriffe bezeichnen nach Hume, wenn sie überhaupt eine Bedeutung haben, Bewußtseinsinhalte. Diese wiederum teilt er in Eindrücke (impressions) und Vorstellungen (ideas) ein, wobei die Eindrücke die lebhafteren und stärkeren und die Vorstellungen die weniger eindringlichen Bewußtseinsinhalte sind.[85]
Der Verstand kann im Bereich des logisch Möglichen beliebige Vorstellungen miteinander verknüpfen und sich etwa mit Leichtigkeit vorstellen, daß ein Stein, in einer gewissen Höhe über dem Erdboden losgelassen, sich schräg nach oben fortbewegt, in der Luft stehenbleibt oder nach unten fällt. Jede dieser Verknüpfungen ist gleich klar vorstellbar; die letztere jedoch wird mit größerer Lebhaftigkeit gedacht als die übrigen, da wir uns nur an Eindrücke dieser Art erinnern können und da wir uns deshalb an diese Verknüpfung gewöhnt haben. Unser Glaube daran, daß der Stein nach unten fallen wird, beruht also vollständig auf Gewohnheit, die durch eine lange Kette von ständigen gleichartigen Aufeinanderfolgen von Eindrücken erzeugt worden ist, und diese Gewohnheit ist wiederum nichts anderes als eine lebhaftere Verknüpfung von gewissen Vorstellungen.[86]
In einigen Fällen ist diese Verknüpfung, diese Assoziation der Vorstellungen, schon so stark, daß unser Geist im gewissen Sinn bereits dazu *determiniert* ist, diese Verknüpfung zu denken.[87] In solchen Fällen sagen wir dann, daß wir auf Grund unserer Erfahrung mit *Sicherheit*

[85] Vgl. Hume [1], S. 1, 157, und Hume [2], S. 18, 21f.
[86] Vgl. Hume [2], S. 46—50.
[87] Vgl. Hume [1], S. 125, 128, 132, 134, 165.

darauf schließen, diese Verknüpfung auch in der Zukunft wahrzunehmen.
In anderen Fällen hat sich die Gewohnheit nicht so stark herausgebildet, sei es, daß die Verknüpfung nicht *oft* genug wahrgenommen worden ist (Wahrscheinlichkeit der ersten Art), sei es, daß sie nicht *immer* wahrgenommen worden ist (Wahrscheinlichkeit der zweiten Art) oder sei es schließlich, daß der gegenwärtige Eindruck oder die Vorstellung den vergangenen Eindrücken nur in einem abgeschwächten Grad ähnlich ist (Wahrscheinlichkeit der dritten Art). Die Vorstellungen sind hier in der gleichen Weise miteinander verbunden wie im Falle des Kausalschlusses, nur nicht so lebhaft wie bei diesem.
Zur Unterstützung dieses Nachweises, daß die Grundlage eines Wahrscheinlichkeitsschlusses nur die Gewohnheit ist, die sich in einer größeren Lebhaftigkeit der Verknüpfungen der Vorstellungen ausdrückt, bringt Hume folgendes Gedankenexperiment: Angenommen, ein Wesen, das mit vollen Verstandeskräften und mit allen unseren Fähigkeiten zur sinnlichen Wahrnehmung ausgerüstet ist, das jedoch bisher noch keine Sinneswahrnehmung gehabt hat, wird plötzlich in unsere Welt gesetzt. Es beginnt wahrzunehmen, daß gewisse Eindrücke immer mit anderen verknüpft sind, und daß in anderen Fällen diese Verknüpfung nur häufig, selten oder überhaupt nicht zu beobachten ist. Nach zwei oder drei gleichförmigen Wahrnehmungen wird dieses Wesen noch nicht schließen, Entsprechendes auch in der Zukunft zu beobachten, wohl aber vielleicht nach hundert derartigen Beobachtungen. In den einzelnen Wahrnehmungen liegt nichts, was diesen Unterschied rechtfertigt; der einzige Unterschied ist, daß sich das Wesen noch nicht nach drei, wohl aber nach hundert gleichförmigen Beobachtungen an diesen Zusammenhang gewöhnt hat und ihn deshalb auch in der Zukunft erwartet.[88]
An diesem Argument ist wenig auszusetzen, außer daß man sich nicht gut vorstellen kann, wie ein konsequenter Phänomenalist so etwas überhaupt sagen kann, und was er mit den Begriffen „Wahrnehmungen dieses Wesens", „Eindrücke dieses Wesens" usw. hier überhaupt meint. Das gleiche gilt von dem folgenden, für sich ebenfalls überzeugenden Argument Humes: Zwischen dem Ablauf der Natur und

---

[88] Vgl. Hume [2], S. 42.

unseren Vorstellungen besteht eine prästabilierte Harmonie; unser Denken und insbesondere unsere Wahrscheinlichkeitsschlüsse stimmen mit dem Ablauf der Dinge der Außenwelt überein. Wäre dies nicht der Fall, so wäre unser Wissen auf das beschränkt, was uns die Sinne und das Gedächtnis liefern, wir könnten uns der Welt nicht anpassen und uns nicht auf die Zukunft einrichten. Durch die Gewohnheit wird bewirkt, daß zwischen unserem Denken und der Welt diese Übereinstimmung besteht, die nötig ist zur Erhaltung unserer Art.[89]

Schließlich möchte ich noch ein ganz andersartiges Argument Humes anführen, mit dem er zu zeigen versucht, daß der Verstand nicht in der Lage ist, die Wahrscheinlichkeitsschlüsse zu rechtfertigen; es stammt aus seiner skeptischen Philosophie und ist durchaus geeignet, dieser das Angsterregende zu nehmen, das ihr im allgemeinen philosophischen Bewußtsein anhaftet.[90] Der Verstand kann uns nämlich nach Hume selbst dann keine vollständige Sicherheit geben, wenn wir dies ohne Zögern erwarten, nämlich bei den demonstrativen Wahrheiten. Dazu genügt es, sich zu vergegenwärtigen, daß wir uns bei logisch-mathematischen Deduktionen schon einige Male geirrt haben und daß es deshalb jedenfalls *möglich* ist, daß wir uns auch diesmal bei der scheinbaren Erkenntnis einer logischen Wahrheit geirrt haben. Der Grad dieser Möglichkeit mag noch so klein sein, er darf jedoch bei einer vernünftigen Überlegung nicht vernachlässigt werden. Nun kann man sich in einem zweiten Schritt überlegen, wie groß der Grad an Gewißheit dafür ist, daß wir uns bei der Berechnung jenes Grades an Gewißheit für die Korrektheit der Argumentation nicht geirrt haben. Auch dieser Grad ist verschieden von dem der absoluten Gewißheit, mag er auch noch so nahe bei ihr liegen. Je öfter man dieses Verfahren wiederholt, desto unbestimmter wird die ursprüngliche Gewißheit und sie verschwindet ganz, wenn man mit diesen Iterationen nicht abbricht. Andererseits haben wir aber bei unseren Argumenten in vielen Fällen Gewißheit; folglich kann dies nicht (allein) auf der Tätigkeit des Verstandes beruhen, der ja, wenn

89 Vgl. Hume [2], S. 54f. Redewendungen, die ein Phänomenalist genauer erklären sollte, bevor er sie benützt, findet man auch in Hume [1], S. 2, und Hume [2], S. 17, 32, 107.

90 Vgl. Hume [1], S. 180—187.

er allein operiert, jede Gewißheit zerstört. Da uns auch die Erfahrung diese Gewißheit nicht geben kann, muß sie auf Gewohnheit beruhen.

Grundsätzliche Kritik gegen Humes Rechtfertigung der Wahrscheinlichkeitsgründe kann man meines Erachtens an folgenden vier Punkten üben:

(1) Er ist bei der Begründung dieser Schlüsse nicht ganz konsequent, da er zwar die Kausalschlüsse, für die er sich hauptsächlich interessiert, und die Wahrscheinlichkeitsschlüsse der ersten drei Arten durch partielle oder totale Gewöhnung an die Erfahrung rechtfertigt, nicht jedoch die Wahrscheinlichkeitsgründe der vierten Art, die völlig auf Zufall beruhen. Wir sind ja auch bei jedem neuen Experiment mit Billardkugeln und nicht nur beim Münzwurf in Ungewißheit über den möglichen Ausgang mit dem einzigen Unterschied, daß wir hier eine gleichförmige Erfahrung besitzen und deshalb auch einen ganz bestimmten Ausgang erwarten, während sich die Ergebnisse bei den bisherigen Würfelwürfen etwa gleichmäßig auf die sechs verschiedenen Möglichkeiten verteilen und wir deshalb auf Grund unserer Gewohnheit dies auch für die zukünftigen Würfe erwarten. Nun ist kein Würfel vollkommen regelmäßig, und man kann sich einen fließenden Übergang von einem (fast) regelmäßigen Würfel zu einem anderen vorstellen, der (fast) immer das Obenliegen einer gewissen Ziffer, etwa der „6", als Ergebnis hat. Wenn wir über den Würfel weiter nichts wissen, können wir eben *nicht* a priori annehmen, daß er vollkommen regelmäßig ist und die Ergebnisse beim Werfen deshalb vollkommen dem Zufall überlassen bleiben, sondern müssen auch hier auf die Erfahrung Bezug nehmen, sei es auf die Würfe mit diesem Würfel, sei es durch einen Analogieschluß auf die Ergebnisse mit anderen Würfeln ähnlicher Gestalt.

(2) Mit einer Selbstverständlichkeit, die Hume nicht weiter begründet, nimmt er an, daß die Gesamtheit der absolut sicheren Wahrscheinlichkeitsschlüsse von vergangenen Beobachtungen auf ein singuläres zukünftiges Ereignis bzw. auf ein allgemeines und ausnahmslos gültiges Gesetz mit der Klasse der Kausalschlüsse identisch ist,[91] mit Schlüssen also, die neben diesem Grad an Gewißheit auch noch die Eigenschaft

[91] Dies äußert sich auch darin, daß nach seiner Ansicht nur Kausalschlüsse aus dem bisherigen Wissen herausführen; vgl. Hume [2], S. 26.

haben, daß die Prämissen und Konklusionen Ereignisse beschreiben, die sich raum-zeitlich berühren und bei denen die Ursache der Wirkung zeitlich vorangeht (die ständige Verknüpfung und die hinreichend große Anzahl von Fällen erzeugen in uns eine uneingeschränkte Gewohnheit, die sich in jenem Grad an Sicherheit ausdrückt). Diese Annahme ist jedoch unberechtigt. Es gibt nicht nur im Alltag Koexistenzgesetze, die man mit Sicherheit behaupten kann, wie etwa „Alle Menschen sind kleiner als 10 m", sondern auch in den Erfahrungswissenschaften, wie etwa die Newtonschen oder die Einsteinschen Gesetze der Gravitation oder wie eine Reihe von Gesetzen der Biologie, etwa „Alle Lebewesen mit Herz sind Lebewesen mit Nieren".
(3) Die Bedingungen, die er für einen gültigen Kausalschluß angibt (ständige Verknüpfung von Ereignissen zweier Arten in einer genügend großen Anzahl von Beobachtungen, wobei die Ereignisse der einen Art den ihnen entsprechenden der anderen Art zeitlich vorangehen und sie raum-zeitlich berühren)[92], sind zwar ohne Zweifel notwendig, aber in ihrer Gesamtheit nicht hinreichend für ein solches Argument. Denn gewisse Symptomgesetze, wie etwa „Wenn bei einer Person auf der roten Wangenschleimhaut Kopliksche Flecken auftreten, so werden sich an ihr später auch noch andere Symptome von Masern zeigen" oder „Wenn an einem Ort das Barometer plötzlich stark fällt, so wird sich danach dort ein starker Sturm erheben", das Gesetz, daß auf den Tag immer die Nacht folgt und die Beobachtung mitsamt der daraus erschlossenen Verallgemeinerung, daß die Wahrnehmung eines in einem bestimmten Winkel in Wasser getauchten Ruders immer mit der Wahrnehmung begleitet ist, daß dieses an der Wasseroberfläche abgeknickt ist, werden zwar auf Grund vergangener Beobachtungen als gesichert angesehen, können jedoch nicht zu den Kausalgesetzen gerechnet werden, obwohl sie alle diesbezüglichen Bedingungen Humes erfüllen. Man kann ihm hier allerdings zugutehalten, daß dies ein grundsätzliches wissenschaftstheoretisches Problem ist, das nicht für seine Philosophie spezifisch ist und das man auch heute noch nicht als befriedigend geklärt ansieht.
(4) Ausgehend von seiner Definition des Begriffs der Ursache formuliert Hume acht Regeln zur Entdeckung von Ursachen und Wirkun-

---

[92] Vgl. Hume [1], S. 172; vgl. auch Hume [2], S. 77f.

gen und stellt diese „Logic" dem gegenüber, was die „scholastic logicians" an Regeln hervorgebracht haben.[93] Diese Regeln sind jedoch nicht alle Konsequenzen aus seiner Definition des Begriffs der Ursache, und es ist auch schwer einzusehen, wie man sie aus der Erfahrung ableiten kann, ohne einen Zi kelschluß zu begehen. Kant hat dies wohl klar gesehen und deshalb versucht, durch eine *transzendentale Logik*, die er der *formalen Logik* gegenüberstellt, ein System von solchen Regeln, mit denen wir Erfahrungskenntnisse gewinnen, als synthetisch-apriorische Prinzipien zu erweisen. Vom Standpunkt Humes aus wiederum muß man diesen Versuch als gescheitert ansehen, da diese Prinzipien nach Kant ja nicht analytisch und damit nicht demonstrativ sind und weil nach Hume ihre Rechtfertigung dann nur durch Wahrscheinlichkeitsschlüsse (im weitesten Sinn) unter Zugrundelegung von Erfahrungen erfolgen kann.

Trotz dieser Einwände bin ich davon überzeugt, daß die Humesche Analyse des *Problems* der Grundlegung des nichtdemonstrativen Schließens eines der wenigen Ergebnisse in der Philosophie ist, die man wohl nicht mehr wird umstoßen können. Seine Darstellung zeigt eine Klarheit der Gedankenführung, wie sie auch nach ihm lange Zeit nicht mehr anzutreffen war. Seiner *Lösung* des Problems ist meines Wissens bis heute kein Lösungsvorschlag entgegengesetzt worden, der einer Kritik besser standhalten könnte.
Seine Begründung des Wahrscheinlichkeitsschlusses ist, wie eine genauere Analyse zeigt, von seinem Phänomenalismus und seinem Skeptizismus relativ unabhängig, obwohl diese Standpunkte zusätzliche Stützen für seine These liefern. Wesentlich für seinen Entschluß, daß rationale Gründe und Erfahrungstatsachen nicht zur Rechtfertigung dieser Argumentationen ausreichen, ist vielmehr sein Empirismus gewesen. Zwar hätten im Prinzip auch Philosophen des Rationalismus des 18. Jahrhunderts zu diesem Ergebnis gelangen können, wenn sie die deduktive Logik korrekt angewandt hätten; doch da sie, wie Christian Wolff (1679—1754), der Ansicht gewesen sind, das allgemeine Kausalprinzip (den Satz vom zureichenden Grund) aus dem logischen Prinzip vom ausgeschlossenen Widerspruch deduzieren zu können, und da sie deshalb Zufälle als Unwissen über die wirk-

[93] Vgl. Hume [1], S. 173ff.

lichen (Kausal-) Zusammenhänge, bewirkt durch unbekannte und einander widerstreitende Ursachen, gedeutet haben, ist ihnen gar nicht zu Bewußtsein gekommen, daß diese Art des Schließens mit Voraussetzungen über den Ablauf der Natur verbunden ist, da sie das Problem der Rechtfertigung dieser Voraussetzungen schon gelöst glaubten, noch bevor es ihnen richtig bewußt wurde.

Als Beispiel kann man die im übrigen sehr interessante und auch heute noch lesenswerte „Logik des Wahrscheinlichen" von Ludwig Martin Kahlius ansehen, der, ausgehend von der Metaphysik Wolffs, Regeln für das nichtdeduktive Schließen formuliert und sie im rationalistischen Sinn begründet hat. Nach ihm sind Sicherheit und Glauben beides Relationen, die eine Aussage auf ihre Gründe beziehen; *sicher* ist eine Aussage, wenn die Gründe alle Bedingungen ihrer Wahrheit enthalten, *wahrscheinlich*, wenn sie nicht alle diese Bedingungen, wohl aber die meisten enthalten, und sonst *möglich* (falls sie nicht logisch falsch ist).[94] Unsere Unsicherheit hängt vom Defekt unseres Wissens ab; wir begnügen uns deshalb mit Wahrscheinlichkeit, in der Hoffnung, daß wir diese durch entsprechende Gründe einmal in Sicherheit überführen können.[95] Mit zwei Methoden gelangt man zur Erkenntnis der Dinge: Durch rationale Gründe aus dem Wesen des Gegenstandes heraus, und auf der Grundlage von Erfahrungen; Wahrscheinlichkeit gewinnt man, indem man in vielen Fällen beobachtet, daß Umstände miteinander verknüpft sind.[96] So sind Naturgesetze nur wahrscheinlich, weil nicht bekannt ist, wie und aus welchen Prinzipien heraus sie zu beweisen sind; ist dieser Beweis jedoch gelungen, so hat man die Wahrscheinlichkeit in Sicherheit übergeführt.[97] Eine Hypothese wird um so wahrscheinlicher, je mehr Erfahrungssätze man aus ihr ableiten kann, je mehr Erfahrungen sie zusammenfaßt, je mehr also mit Wenigem begründet werden kann, je mehr Bedingungen der Wahrheit anwesend sind.[98] Wahrscheinlichkeit und Annahmen (Hypothesen, coniecturae) sind die Ergänzungen für den Defekt

---

[94] Vgl. Kahlius [1], §§ XI, XVII, XX, XXII, XXV, XXVI, LXXIX, XCIV, CVII.
[95] Vgl. Kahlius [1], §§ XLII, LXII.
[96] Vgl. Kahlius [1], §§ LXIX, XCVII.
[97] Vgl. Kahlius [1], § XXXV.
[98] Vgl. Kahlius [1], § LXXIII.

unseres Wissens; Gott braucht deshalb keine Wahrscheinlichkeit und keinen Wahrscheinlichkeitsgrad.[99]
Man kann diese Ausführungen folgendermaßen zu deuten versuchen: Man geht von bestimmten Annahmen über den Zustand der Welt aus, die einander wechselseitig ausschließen, die jedoch in ihrer Gesamtheit erschöpfend sind, und versucht, aus jeder einzelnen die fragliche Aussage, über die man Gewißheit erlangen will, zu deduzieren. Je mehr Deduktionen gelungen sind, desto mehr Bedingungen der Wahrheit enthält die Aussage und desto wahrscheinlicher ist sie folglich; sind alle Deduktionen gelungen, so ist sie sicher. Eine andere und damit nicht leicht zu vereinbarende Bestimmung des Wahrscheinlichkeitsbegriffs läuft darauf hinaus, als die wahrscheinlichste Hypothese die einfachste anzusehen, die es gestattet, vieles mit Wenigem zu begründen, und die meisten oder alle Erfahrungstatsachen aus ihr zu folgern. Kahlius war sich offenbar nicht im klaren darüber, daß dies nicht äquivalent ist mit den Bedingungen der Wahrheit der Hypothese.

Systematisch mit dem Problem der Begründung des induktiven Schließens hat sich nach Hume erstmals wieder Sir Karl R. Popper (geb. 1912) auseinandergesetzt.[100] Er unterscheidet zwei Bedeutungen des Begriffs „Induktion" und versucht zu zeigen, daß beide für die Erweiterung unseres Wissens und den wissenschaftlichen Fortschritt ohne Relevanz sind.
Induktion kann nach Popper verstanden werden als eine Theorie der Auffindung erfahrungswissenschaftlicher Theorien bei Zugrundelegung gewisser Erfahrungsdaten. Eine solche Methode kann es aber nicht geben, weil, wie er nachweisen kann, die Wissenschaftler tatsächlich nicht von den Erfahrungen zu den Theorien, sondern umgekehrt von den Theorien zu den Erfahrungen fortschreiten: Zunächst werden gewisse primitive Theorien entworfen und mit der Erfahrung verglichen; werden sie der Erfahrung nicht gerecht, da gewisse

---

99 Vgl. Kahlius [1], § XCII.
100 Vgl. Popper [1] und [2]. Allerdings kann man wohl nicht behaupten, daß er seine Gedanken in diesen Werken systematisch *dargestellt* hat. Eine kurze Analyse seines Bestätigungsbegriffs, die den heutigen Anforderungen an Klarheit genügt, hat meines Wissens erstmals Vetter in [1], S. 42—52, gegeben. Eine prägnante Zusammenfassung seiner Argumente gibt Popper in [1] auf S. 32ff.

aus ihr logisch deduzierbare Folgerungen den Erfahrungstatsachen widersprechen, so werden sie ergänzt und verbessert oder durch neue Hypothesen ersetzt, die man wiederum der Erfahrung aussetzt, und so fort. Ein rationaler Weg von den Beobachtungen zu den Gesetzesaussagen ist demnach nicht möglich.

Nach Popper ist Induktion jedoch auch in dem Sinn unmöglich, daß Theorien nicht aus Erfahrungen *gewonnen*, sondern durch sie *begründet* werden. Denn Theorien sind Allaussagen, die über alle Dinge, über alle Orte im Universum und über alle Zeitpunkte sprechen, die also weitaus mehr behaupten als Sätze über die vergangenen Beobachtungen und die deshalb aus diesen nicht logisch folgen können. Während also beliebig viele positive Instanzen nicht genügen, um eine Theorie als gesichert anzusehen (ein Gegenbeispiel in der Zukunft ist ja immer möglich und widerspricht nicht logisch den vergangenen Erfahrungen), genügt eine einzige (reproduzierbare) Gegeninstanz[101], um sie zu widerlegen.

Nach Popper ist nicht nur eine definitive Deduzierung, sondern auch eine induktive Bestätigung von Theorien unmöglich. Denn zur Rechtfertigung solcher Bestätigungen benötigt man ein Induktionsprinzip, das besagt, daß die Zukunft mit der Vergangenheit gleich oder zumindest bis zu einem gewissen Grad ähnlich ist, und das weder logisch beweisbar noch aus den Aussagen über die bisherigen Beobachtungen logisch ableitbar ist, wie schon Hume gezeigt hat. Andererseits kann Popper die psychologische Begründung dieser Annahme, wie sie Hume gegeben hat, nicht akzeptieren, da sie nur eine Antwort auf die Frage gibt, *warum* wir *faktisch* so schließen, und nicht das Problem löst, *mit welchem Recht* wir dies tun. Diese Annahme ist weder logisch noch empirisch begründbar und deshalb überhaupt nicht zu rechtfertigen. Folglich ist auch das induktive Schließen weder logisch noch empirisch zu rechtfertigen.

---

101 Singuläre Effekte, oft auch „Ausreißer" genannt, braucht der Naturwissenschaftler nicht zu berücksichtigen, wie Popper gezeigt hat. Die Behauptung, daß ein gewisser Effekt reproduzierbar ist, stellt sich jedoch bei genauerer Analyse als logisch äquivalent mit einem gewissen einfachen Naturgesetz heraus. Folglich braucht man zur Überprüfung von Naturgesetzen nach Popper nicht elementare Beobachtungen, sondern selbst bereits Gesetze einer einfacheren Art. Diesen Sachverhalt hat wohl auch schon Hume gekannt, vgl. Hume [1], S. 5f., und Hume [2], S. 20f. und S. 109—116.

Bis jetzt kann man Popper noch ohne Bedenken zustimmen. Man wird dies jedoch bei genauerem Zusehen nicht mehr tun, wenn er scheinbar ganz korrekt weiter folgert, daß dann auf Schlüsse dieser Art verzichtet werden müsse, da sie auf einer unbegründbaren Annahme beruhten. Es liegt ja kein Widerspruch in jener Annahme, und sie kann sogar zufällig richtig sein, auch wenn man sie weder logisch noch empirisch beweisen kann; und selbst wenn sie nur annähernd richtig ist, kann es doch sein, daß die induktiven Schlüsse, die durch sie erlaubt werden, zu annähernd richtigen Ergebnissen führen und daß die Fehler, die man dabei macht, innerhalb einer Grenze liegen, die es gestattet, sie zu vernachlässigen.

Als Ersatz für den Begriff der induktiven Bestätigung schlägt Popper den der Bewährung vor. Eine Theorie bewährt sich an der Erfahrung, wenn sie allen Widerlegungsversuchen widersteht, wenn also keine Aussage, die aus ihr logisch ableitbar ist, den Erfahrungsbehauptungen widerspricht. Nun kann man weiterfragen, ob sich eine Theorie nicht besser bewähren kann als eine andere und ob man ihr nach einer großen Anzahl von derartigen Falsifikationsversuchen nicht *mehr* vertrauen kann als nach einer kleinen, ob man nach vielen derartigen Versuchen also nicht eher damit rechnet, daß auch der nächste Widerlegungsversuch scheitern wird. Poppers Antwort darauf ist nicht ganz eindeutig. Manchmal scheint es, als ob sich nach seiner Ansicht eine Theorie nach hundert Falsifikationsversuchen besser bewährt hat als nach einem, an anderen Stellen hingegen streitet er dies offenbar ab. Was er dazu aber auch sagen mag, er setzt trotzdem etwas über die Welt voraus: in dem einen Fall, daß man den vergangenen Versuchen bis zu einem gewissen Grad vertrauen darf und im anderen, daß die Ergebnisse dieser Experimente Zufälle sind, die keine weiteren Schlüsse über zukünftige Experimente gestatten.

Mit der Frage der Begründung des induktiven Schließens hat sich in letzter Zeit systematisch auch Rudolf Carnap (1891—1970) befaßt. Während er zunächst geglaubt hat, ein induktives Verfahren zur Begründung von Hypothesen aller Art durch Erfahrungsdaten charakterisieren zu können, bei dessen Verwendung wie beim deduktiven Schließen keine Voraussetzungen über die Natur gemacht werden,[102]

---

[102] Vgl. Carnap [2].

ist es ihm 1952 gelungen, zu zeigen, daß mit jeder Verwendung einer induktiven Methode eine Voraussetzung über die Welt verbunden ist. Allerdings ist es nicht möglich, *eine* Voraussetzung über die Welt für *alle* ihre Aspekte zu machen, sondern man muß zu *jedem* ihrer Aspekte nach *einer* Voraussetzung suchen, unter deren Verwendung man möglichst adäquate Ergebnisse erhält. Dieser Ansatz wird in den folgenden Kapiteln analysiert und systematisch dargestellt, wobei ich versuchen möchte, das Problem der Voraussetzungen des induktiven Schließens von allen Seiten zu beleuchten.

4. Die mathematischen Wahrscheinlichkeitsbegriffe

Wer den Zusammenhang zwischen der induktiven Methode c^* von Carnap und der Sukzessionsregel von Laplace nicht kennt, wird verwundert fragen, wieso innerhalb einer Untersuchung der induktiven Logik auch die mathematischen Wahrscheinlichkeitsbegriffe behandelt werden. Dieses Erstaunen ist durchaus berechtigt: Die induktiven Verfahren und die Methoden der mathematischen Wahrscheinlichkeitslehre haben sich nebeneinander und unabhängig voneinander entwickelt; ein nennenswerter Einfluß der einen auf die anderen hat nicht stattgefunden.

In dieser Arbeit habe ich mir unter anderem das Ziel gesetzt, zu zeigen, daß kein grundsätzlicher Unterschied zwischen den Wahrscheinlichkeitsschlüssen der Mathematiker und den Induktionsargumenten der Philosophen besteht (soweit es sich hierbei um enumerative Induktionen handelt), daß vielmehr die einen mittels der anderen begründet werden können. Obwohl ein derartiges Unterfangen billigerweise mit dem Interesse sowohl der Philosophen als auch der Mathematiker rechnen könnte, bin ich mir wohl bewußt, daß zumindest im deutschen Sprachraum die große Mehrheit der Philosophen behaupten wird, daß ich nichts anderes als einen mathematischen Wahrscheinlichkeitsbegriff anbiete, der nichts mit der sogenannten *eigentlichen* Induktion zu tun habe, und daß hierzulande leider auch die meisten oder alle Mathematiker der Ansicht sein werden, daß in diesem Buch nichts anderes als ein philosophischer Begriff der Induktion dargestellt

wird, der für die Mathematik ohne Relevanz ist, auch wenn er vielleicht präziser als sonst charakterisiert wird. Trotzdem hoffe ich, wenigstens bei *einigen* Philosophen und bei dem *einen oder anderen* Mathematiker eine Unvoreingenommenheit antreffen zu können, die groß genug ist, den hier vorgebrachten Argumenten überhaupt Gehör zu schenken.

Die Vorgeschichte der mathematischen Wahrscheinlichkeitslehre reicht bis ins Mittelalter zurück; sie hat ihre Wurzel in der Beschäftigung einiger Mathematiker mit den Erfolgsaussichten bei Glücksspielen, wie etwa den Würfelspielen,[103] deshalb hat sie auch gelegentlich die Bezeichnung „géométrie du hazard" bzw. „aleae geometria" erhalten. Zu erwähnen sind hier Luca Pacciola (ca. 1445—ca. 1514) mit seiner „Summa de arithmetica geometria proportioni et proportionalita" (Venedig 1494), und Geronimo Cardano (1501—1576) mit seiner „Practica arithmetica generalis" (1539), die mit einfachen mathematischen Hilfsmitteln Probleme der Wahrscheinlichkeitslehre zu lösen versuchten. Wenn auch die Lösungen nicht korrekt sind, so haben diese Werke doch nicht unwesentlich zur Ansicht unter den damaligen Gelehrten beigetragen, daß Fragen dieser Art mathematisch eindeutig beantwortet werden können.

Systematisch und mit umfangreicheren mathematischen Hilfsmitteln haben sich erstmals Blaise Pascal (1623—1662) und Pierre de Fermat (1601—1655) befaßt. Es wird berichtet, daß ein an mathematischen Fragen interessierter Glücksspieler de Méré mit zwei Fragen, die Würfelspiele betreffen, an Pascal herangetreten sei, die diesen zu mathematischen Untersuchungen über Probleme dieser Art veranlaßt haben. Gleichzeitig und unabhängig davon gelangte Fermat von einem allgemeineren Ansatz, der kombinatorische Probleme betraf, zu den gleichen Ergebnissen. Beide teilten sich ihre Resultate mit, hielten sie aber vor der Fachwelt geheim.

Christiaan Huyghens (1629—1695) hatte in Paris gehört, daß mathematische Lösungen dieser Probleme existierten. Da er jedoch die Ergebnisse und die Methoden Pascals und Fermats nicht kannte, sah er

---

103 Eine ausführliche Darstellung der Geschichte der Wahrscheinlichkeitslehre gibt Todhunter in [1]; eine gute Zusammenfassung derselben findet man im Anhang zu Bernoulli [2].

sich gezwungen, den ganzen Gegenstand systematisch von Anfang an zu entwickeln. Seine Ergebnisse veröffentlichte er in „De ludo aleae" (erstmals abgedruckt in: Franciscus Schotanus [= Franz Schooten], Exercitationum mathematicarum libri quinque, Leiden 1657, S. 519—534; mit Kommentar abgedruckt in Bernoulli [1]). Er kann als der eigentliche Begründer der mathematischen Wahrscheinlichkeitslehre angesehen werden. Jakob Bernoulli (1654—1705) kommentierte und ergänzte diese Ergebnisse in [1]. In der Folgezeit erlebte die Wahrscheinlichkeitslehre eine stürmische Entwicklung, zu der neben vielen anderen auch die Mathematiker Karl Friedrich Gauss (1777 bis 1855) und Thomas Bayes (1702—1761) beitrugen. Systematisch zusammengefaßt und ergänzt wurden diese Resultate durch Pierre Simon de Laplace (1749—1827) in [1].

Die Wahrscheinlichkeit eines Ereignisses ist von Laplace als das Verhältnis der Anzahl der für das Ereignis *günstigen Fälle* zur Anzahl aller *möglichen Fälle* definiert worden, wobei vorauszusetzen ist, daß die einzelnen *möglichen Fälle* jeweils *gleichmöglich* sind; trifft diese Voraussetzung nicht zu, so muß zunächst der Grad der Möglichkeit der einzelnen Fälle bestimmt werden, und die Wahrscheinlichkeit ist dann das Verhältnis der Summe der Grade der Möglichkeiten der für das Ereignis günstigen Fälle zur Summe der Grade der möglichen Fälle überhaupt.[104]

Sind zwei Ereignisse in dem Sinn voneinander unabhängig, daß das Eintreffen des einen ohne Relevanz für die Wahrscheinlichkeit des Eintreffens des anderen ist, so ist die Wahrscheinlichkeit dafür, daß beide Ereignisse gleichzeitig verwirklicht sind, gleich dem Produkt der beiden einzelnen Wahrscheinlichkeiten; besteht diese Unabhängigkeit nicht, so ist die Wahrscheinlichkeit dafür, daß beide gleichzeitig verwirklicht sind, gleich dem Produkt der Wahrscheinlichkeit des einen Ereignisses mit der Wahrscheinlichkeit dafür, daß das zweite verwirklicht ist unter der Voraussetzung, daß das erste stattgefunden hat.[105] Auf diese Weise können bedingte Wahrscheinlichkeiten eingeführt werden: Die Wahrscheinlichkeit dafür, daß ein Ereignis verwirklicht ist unter der Voraussetzung, daß ein anderes stattgefunden

---

104 Vgl. Laplace [2], S. 6 und 9.
105 Vgl. Laplace [2], S. 11.

hat, ist gleich dem Quotienten aus der Wahrscheinlichkeit dafür, daß beide Ereignisse verwirklicht sind, und der Wahrscheinlichkeit dafür, daß jenes andere stattgefunden hat (vorausgesetzt, daß diese zweite Wahrscheinlichkeit nicht 0 ist).[106] Ist jenes zweite Ereignis eines, das der Vergangenheit angehört, während das erste noch in der Zukunft liegt, so liefert dieser Quotient also die Wahrscheinlichkeit eines zukünftigen Ereignisses auf Grund einer vergangenen Beobachtung.

Scheinbar ist das Laplacesche System somit vollständig, insbesondere auch deshalb, weil es erlaubt, genaue numerische Werte für die Wahrscheinlichkeit zukünftiger Ereignisse auf Grund vergangener Beobachtungen anzugeben:[107] Hat man in k bisherigen Beobachtungen festgestellt, daß k′ der k Objekte von der Art F sind, dann ist der Bruch $(k' + 1)/(k + 2)$ die Wahrscheinlichkeit, daß ein bisher noch nicht beobachtetes Ding ebenfalls die Eigenschaft F hat, bezüglich jenes Erfahrungsdatums.

Die tatsächliche Unvollständigkeit dieses Systems hängt mit der Vagheit und Mehrdeutigkeit der Begriffe „möglicher Fall", „gleich möglich" und „Grad der Möglichkeit" zusammen. Unter dem Begriff „möglicher Fall" kann wenigstens dreierlei verstanden werden: (1) eine *mögliche Welt* im Sinne von Leibniz (was im Sinne der späteren Terminologie als eine *Zustandsbeschreibung* bzw. als eine *individuelle Verteilung* verstanden werden kann), (2) eine *mögliche Welt unter Absehung der tatsächlichen Individuen* (was später eine *statistische Verteilung* genannt wird), und (3) ein *Element aus einer endlichen Bezugsklasse G von beobachteten Gegenständen* (das genau dann ein *günstiger Fall* ist, wenn es von der Art F ist). Die Väter der mathematischen Wahrscheinlichkeitslehre, die sich fast ausschließlich mit Glücksspielen befaßten, hatten vermutlich die erste Deutung im Auge, während die Deduktion der Sukzessionsregel zeigt, daß Laplace an die zweite gedacht hat.

Zur dritten Deutung des Möglichkeitsbegriffs führt wohl am zwanglosesten der folgende Weg. Aus gewissen einfachen Eigenschaften des Wahrscheinlichkeitsbegriffs folgt der Lehrsatz: „r ist die Wahrscheinlichkeit dafür, daß ein Gegenstand a aus einer Gesamtheit G die Eigen-

---

106 Vgl. Laplace [4], S. 13.

107 Zur Deduktion der sogenannten Laplaceschen Sukzessionsregel vgl. Theorem 24 auf S. 256ff. dieses Buches.

schaft F hat, wenn bekannt ist, daß r die relative Häufigkeit von F in G ist."[108] Ist also etwa die relative Häufigkeit von Akademikern unter den Münchnern $1/10$ dann ist, bezüglich dieses Erfahrungsdatums die Wahrscheinlichkeit dafür, daß der Münchner Cyprian Zwiderl ein Akademiker ist, gerade $1/10$. Es liegt daher nahe, in diesem Fall die beobachtete relative Häufigkeit mit der Wahrscheinlichkeit zu identifizieren. Ist die Bezugsklasse G hingegen nicht, wie in diesem Fall, endlich, so bietet sich als Wahrscheinlichkeit dann der Grenzwert einer Folge von relativen Häufigkeiten von immer umfassenderen Teilklassen von G an; wenn dieser Grenzwert jedoch nicht existiert, wenn die Folge der relativen Häufigkeit also nicht gegen eine reelle Zahl r konvergiert, dann gibt man konsequenterweise überhaupt keinen Wahrscheinlichkeitswert an.

Einen derartigen Wahrscheinlichkeitsbegriff hat erstmals Mill in der ersten Auflage von [1] 1843 vertreten.[109] Ausführlicher und konsequenter hat John Venn (1834—1923) vorgeschlagen, den Wahrscheinlichkeitsbegriff in diesem Sinn zu verstehen. Charles Sanders Peirce (1839—1914) hat diese Deutung zeitweise übernommen; allerdings hat er gelegentlich, wohl beeinflußt von Kries, die Wahrscheinlichkeit einer Aussage auch als das Verhältnis der Anzahl der Wahrheitsmöglichkeiten, unter denen sie wahr ist, zur Anzahl der Wahrheitsmöglichkeiten überhaupt charakterisiert.[110] Eine mathematisch auswertbare Wahrscheinlichkeitstheorie in diesem Sinn haben allerdings weder er noch Venn geliefert.

In mathematisch einwandfreier Form sind diese Gedanken erstmals von Richard von Mises (1883—1953) und Hans Reichenbach (1891 bis 1953) entwickelt worden.[111] Mises legt seinen Wahrscheinlichkeiten sogenannte *Kollektive* zugrunde, d. h. unendliche Gesamtheiten, deren relative Häufigkeiten gegen einen bestimmten Grenzwert kon-

---

108 Vgl. Korollar 11 auf S. 233 dieses Buches.

109 Vgl. das in Fußnote 59 Gesagte.

110 Vgl. Charles Sanders Peirce, Collected Papers, vol. II, Cambridge, Mass. 1960; im ersten Sinn verwendet er den Wahrscheinlichkeitsbegriff z. B. in den §§ 650, 652, 658, 661, 683, 697, 698, 700, 702, 730, 733, 737, 741, 758, im zweiten Sinn z. B. in den §§ 649, 651, 695, 696, 744—747, 756, 757.

111 Der Kollektivbegriff, auf den sich die Misesschen Wahrscheinlichkeiten beziehen, erwies sich zunächst als unerfüllbar, wurde jedoch von Abraham Wald in [1] korrigiert; vgl. auch Mises [2], S. 118f.

vergieren und bei denen fast alle regellos ausgewählten Teilklassen ebenfalls gegen diesen Grenzwert konvergieren (die Begriffe „fast alle" und „regellos" haben dabei einen mathematisch eindeutigen Sinn). Reichenbach hingegen hat nicht an *faktisch* unendliche Gesamtheiten von dieser Struktur, sondern an *potentiell* unendliche derartige Klassen gedacht, so daß er auch sinnvoll über die Wahrscheinlichkeit singulärer Ereignisse sprechen kann, etwa über die Wahrscheinlichkeit dafür, daß dieses Flugzeug mit einem ganz bestimmten und bisher noch niemals beobachteten Motorschaden heil den Flughafen erreichen wird.[112]

Diese Deutung des Wahrscheinlichkeitsbegriffs hat den Nachteil, daß man nicht sehen kann, wie man aus endlich vielen Beobachtungen irgendwann einmal eine einzige derartige Wahrscheinlichkeit, einen solchen Grenzwert einer unendlichen Folge von relativen Häufigkeiten, gewinnen kann. Wenn bei einer Folge von 10000 Münzwürfen z. B. nur 100mal *Kopf*, hingegen 9900mal *Wappen* beobachtet worden ist, so ist diese relative Häufigkeit 0,99 mathematisch verträglich mit jedem Grenzwert zwischen 0 und 1 von Folgen von relativen Häufigkeiten, die mit ihr beginnt. Allerdings sind die meisten dieser Grenzwerte nicht sehr *wahrscheinlich*, und es wird sie „kein Einsichtiger vermuten", wie Mises zu Recht sagt.[113] Um dies aber überhaupt sagen zu können, braucht man einen Wahrscheinlichkeitsbegriff, der nicht als Grenzwert einer Folge von relativen Häufigkeiten definiert werden darf, da man sonst in einen unendlichen Regreß gerät.

Reichenbach hat eine sogenannte *Induktionsregel* aufgestellt, mit deren Hilfe man von beobachteten relativen Häufigkeiten zu Wahrscheinlichkeiten im Sinne von Grenzwerten derartiger relativer Häufigkeiten gelangt. Sie ist eine Verdeutlichung und genauere Formulierung der schon von Mill und später auch von Mises aufgestellten Behauptung, daß die relativen Häufigkeiten mit zunehmender Beobachtung im allgemeinen gegen einen Grenzwert konvergieren.[114] Nach Reichenbach sind zwei Fälle zu unterscheiden:

Entweder die Folge von relativen Häufigkeiten konvergiert nicht

---

[112] Vgl. Reichenbach [10], S. 366—378.

[113] Vgl. Mises [2], S. 19.

[114] Vgl. das in Fußnote 59 Gesagte, sowie Mises [2], S. 16ff., und Reichenbach [10], S. 444—451, insbesondere S. 446.

gegen einen Grenzwert; dann sind nach seiner Ansicht alle induktiven Methoden gleich gut, oder besser: gleich schlecht, da keine von ihnen die objektive Wahrscheinlichkeit liefern kann, weil eine solche ja gar nicht existiert. Oder aber die Folge konvergiert; dann erhält man diese Konvergenz durch folgende Regel: Uns sind stets nur endlich viele Beobachtungen gegeben. Es sei k die Anzahl der Dinge aus einer Gesamtheit G, die auf eine Eigenschaft F hin überprüft worden sind. r sei die relative Häufigkeit von F in der beobachteten Teilklasse von G. Dann wird r nicht unbedingt die *relative Häufigkeit von* F *in* G *auf lange Sicht*, d. h. der *Grenzwert dieser relativen Häufigkeiten* sein; dieser wird jedoch in einem Zahlenintervall [r — s, r + s] um r liegen. Eine solche reelle Zahl s sei fest vorgegeben. Dann bemißt sich der *Grad an Sicherheit* dafür, daß auch die zukünftigen relativen Häufigkeiten in dem Intervall [r — s, r + s] bleiben werden, danach, wie lange die relative Häufigkeit schon in diesem Intervall ist; sind erst die letzten drei relativen Häufigkeiten in ihm, so wird die Sicherheit dieses Induktionsschlusses gering sein, während sich diese der Gewißheit nähert, wenn bei vielen Überprüfungen (d. h. bei einem großen k) über die Hälfte der letzten relativen Häufigkeiten aus diesem Intervall sind. Den Grenzwert der relativen Häufigkeiten kann man nach Reichenbach nun dadurch beliebig approximieren, daß man s immer kleiner wählt und entsprechend eine immer größere Anzahl k von Dingen aus G auf die Eigenschaft F hin überprüft.

Nun ist es jedoch unmittelbar klar, daß dieses Induktionsverfahren niemals den *tatsächlichen* Grenzwert der relativen Häufigkeiten liefert, sondern bestenfalls eine gute *Schätzung* der relativen Häufigkeit von F in einer unendlichen Bezugsklasse G oder aber eine *endliche Folge* derartiger Schätzungen; denn gleichgültig, wie groß die Zahl unserer Beobachtungen ist, sie muß immer endlich sein. *Man erreicht mit diesem Verfahren also tatsächlich keinen einzigen Grenzwert der relativen Häufigkeiten*; man muß sich vielmehr stets mit deren Schätzungen begnügen.

Dieser prinzipiellen Schwierigkeit war sich Reichenbach wohl bewußt. Er hat sie folgendermaßen zu bagatellisieren versucht:[115] Man muß zwischen dem *theoretischen* und dem *praktischen* Grenzwert der

---

115 Vgl. Reichenbach [10], S. 347.

relativen Häufigkeiten unterscheiden. Der theoretische ist uns niemals gegeben, da niemand unendlich viele Experimente oder Überprüfungen vornehmen kann. Andererseits stehen uns gewöhnlich genügend viele Beobachtungen zur Verfügung, deren Anzahl wir für praktische Zwecke mit *unendlich* identifizieren können, ohne einen zu großen Fehler zu begehen. Wenn wir etwa mit einer Münze nur 1000000 Würfe machen können (sei es, daß wir danach sterben, oder sei es, daß die Münze danach für Würfe nicht mehr geeignet ist) und wenn wir bei 999000 Würfen eine relative Häufigkeit von $^5/_6$ festgestellt haben, daß das *Wappen* oben liegt, so wird sich nach seiner Ansicht bei den nächsten 1000 Münzwürfen an dieser relativen Häufigkeit nicht mehr viel ändern. Wenn die relative Häufigkeit der Münzwürfe *theoretisch* einen anderen Grenzwert hat, den wir wegen unseres vorzeitigen Ablebens leider nicht mehr erleben, und wenn sie überhaupt keinen Grenzwert besitzt, dann ist es in dieser Situation dennoch vernünftig, *so zu tun, als ob* er vorhanden und in etwa durch die Beobachtung erkannt worden wäre; der Erfolg, den man bei der Anwendung eines solchen Wahrscheinlichkeitsbegriffs (z. B. bei Wetten) hat, wird dieses Verfahren rechtfertigen.

II. Analyse des logischen Wahrscheinlichkeitsbegriffs

1. Subjektive, objektive und logische Wahrscheinlichkeitsbegriffe

Durch ein System der induktiven Logik sollen nicht neue Denkweisen geschaffen, sondern alte, bekannte und oft geübte Argumente präzisiert und expliziert werden. Dies gilt ja auch für die deduktive Logik: Aristoteles hat nur Schlüsse, die vor ihm schon Sokrates, Platon und die Eleaten stillschweigend benützt haben, expliziert und mit anderen in ein System gebracht. Er hat dabei erkannt, daß solche Schlüsse nicht nur Philosophen benützen, sondern alle, die beweisen und Schlußfolgerungen ziehen. Im Unterschied zur deduktiven Logik geht der Gehalt eines induktiven Schlusses allerdings über den der Prämisse hinaus: die Konklusion sagt mehr über die Welt aus als die Prämisse. Sie ist deshalb nicht unbedingt wahr, wenn die Voraussetzung wahr ist, sondern nur mit einer gewissen Wahrscheinlichkeit. Durch diese Wahrscheinlichkeit wird angezeigt, in welchem Grad die Konklusion, die Hypothese, durch den als bekannt angenommenen Satz, durch die Evidenz, gestützt wird. Die folgenden Sätze sind Beispiele von induktiven Argumenten:

„Daß morgen schönes Wetter ist, ist auf Grund der Wettersituation der vergangenen Woche in hohem Grad wahrscheinlich (im Grad r wahrscheinlich)." „Es ist auf Grund der gesammelten Erfahrungsdaten sehr wahrscheinlich, daß jene Hypothese richtig ist." „Auf Grund jener Stichprobe ist s die Schätzung der relativen Häufigkeit jener Eigenschaft bezüglich einer Gesamtheit von Gegenständen." Solche intuitiv angewandten Verfahren sollen durch ein System der induktiven Logik expliziert, gerechtfertigt und zu einer Theorie vereinigt werden.

Die *induktive Logik* ist also eine Theorie der *induktiven Methoden*. Unter dem Begriff „induktive Methode" kann man zweierlei verstehen und hat man in der Geschichte der Philosophie auch zweierlei verstanden, meist ohne sich des Unterschieds bewußt zu werden: Es kann damit ein Verfahren gemeint sein, das von einer Anzahl von

Beobachtungen zu einer Theorie führt, unter die diese Erfahrungen subsumierbar sind, was etwa die Verfahren Bacons und Mills zu leisten beanspruchen. Es kann damit aber auch ein Verfahren gemeint sein, das den Grad der Bestätigung oder Stützung einer *vorgegebenen* Theorie (bzw. Hypothese) durch ein *vorgegebenes* Erfahrungsdatum angibt; die Hypothese muß also bereits zur Verfügung stehen, bevor man eine derartige induktive Methode überhaupt anwenden kann.
Im folgenden soll der Begriff der induktiven Methode stets im zweiten Sinn verstanden werden. Es ist äußerst zweifelhaft, ob es generell einen rationalen Weg von Beobachtungen zu erfahrungswissenschaftlichen Theorien gibt; jedenfalls ist uns bis heute kein solcher bekannt. Die Hoffnung, einen derartigen Weg jemals zu finden, wird aber stark gedämpft durch die Tatsache, daß es nachweislich unentscheidbare Theorien gibt: Es gibt kein mechanisch anwendbares Verfahren, um zu entscheiden, ob eine vorgelegte Aussage der Sprache der engeren Quantorenlogik logisch wahr ist oder nicht.[1] Die erweiterte Quantorenlogik und die Arithmetik der natürlichen Zahlen sind darüberhinaus sogar unvollständig, d. h. es gibt kein endliches und mechanisch anwendbares Verfahren, das *alle* logisch wahren Sätze der erweiterten Quantorenlogik liefert, und es gibt keine derartige Methode, die *jeden* Lehrsatz der Arithmetik der natürlichen Zahlen aufweist.[2] Mit diesen skeptischen Bemerkungen soll nicht der Elan der Kybernetiker gemindert werden, nach Maschinen zu suchen, die

---

[1] Eine Aussage der Sprache der engeren Quantorenlogik ist aus Gegenstands-, Eigenschafts- und Beziehungsausdrücken sowie aus den junktorenlogischen Konstanten „nicht", „und", „oder", „wenn — dann" und „genau dann, wenn" sowie aus den Quantoren „es gibt ein Ding" und „für jedes Ding" zusammengesetzt. Eine genauere Charakterisierung dieses Begriffs wird auf S. 89 f. dieses Buches gegeben. Zur Unentscheidbarkeit der engeren Quantorenlogik vgl. z. B. Hans Hermes, Aufzählbarkeit, Entscheidbarkeit, Berechenbarkeit, Berlin — Göttingen — Heidelberg 1961, S. 165—172.

[2] Die Sprache der erweiterten Quantorenlogik enthält neben den Gegenstands-, Eigenschafts- und Beziehungsausdrücken auch Ausdrücke, die Eigenschaften von derartigen Eigenschaften oder Beziehungen oder aber Beziehungen zwischen derartigen Eigenschaften und Beziehungen bezeichnen usw., und deren Quantoren sich nicht nur auf Dinge, sondern auch auf alle diese Eigenschaften und Beziehungen erstrecken. Zur Unvollständigkeit der erweiterten Quantorenlogik und der Arithmetik der natürlichen Zahlen vgl. z. B. Hans Hermes, Einführung in die mathematische Logik, Stuttgart 1963, S. 152—157.

das menschliche Verhalten bei der Auffindung von geeigneten Hypothesen und Theorien simulieren; es soll damit nur behauptet werden, daß solche Maschinen, wenn sie einmal entwickelt worden sind, vermutlich mit Zufallsoperatoren arbeiten werden und jedenfalls nicht einen geraden Weg von der Erfahrung zu den Theorien beschreiten.

Es wird gelegentlich behauptet, daß eine induktive Methode den *Grad der Wahrheit* einer Aussage (hinsichtlich eines Erfahrungsdatums) angibt; dies ist gleichermaßen falsch und irreführend. *Es gibt keine Grade der Wahrheit*, sondern nur die Qualitäten *wahr* und *falsch* (= *nicht wahr*).[3] *Grade kennt allerdings unsere Gewißheit*, die wir einer Aussage, eventuell im Hinblick auf andere Sätze, die man als wahr akzeptiert hat, beimessen. Den höchsten Grad an Gewißheit haben ohne Zweifel die logisch wahren Behauptungen; es ist schlechterdings unmöglich, sich vorzustellen, was noch gewisser sein könnte. Die geringste Gewißheit wird den logisch falschen Aussagen zuerkannt. Der Grad der Gewißheit einer Aussage variiert dann zwischen diesen beiden Extremen. Wird die Gewißheit nicht als Eigenschaft einer Aussage, sondern als Relation zwischen einer Aussage und einem Erfahrungsdatum (einem Wissen) verstanden, so variiert der Grad dieser Gewißheit analog zwischen der Gewißheit, die der logischen Folgerung zukommt, und jener, die für die logische Unverträglichkeit eigentümlich ist; wenn aus einem Erfahrungsdatum eine Hypothese logisch folgt, so hat sie relativ zu diesem Datum die höchste Gewißheit, die ihr zukommen kann, und wenn sie mit diesem Wissen logisch unverträglich ist, so kommt ihr relativ zu diesem der geringste Gewißheitsgrad zu. Eine induktive Methode in diesem Sinn ist ein Gradmesser, der anzeigt, wie sehr die Hypothese durch ein Erfahrungsdatum impliziert wird bzw. wie sehr ihre Aussage bereits in der Erfahrung enthalten ist.

Es werden im folgenden deshalb die Aussagen „Der Satz Ψ ist im Grad r bestätigt" und „Der Satz Ψ ist im Grad r gewiß" sowie die

[3] Es gibt zwar sogenannte mehrwertige Logiksysteme, die nicht, wie die klassische Logik, auf dem zweiwertigen Wahrheitsbegriff (der nur die Wahrheitswerte *wahr* und *falsch* kennt), sondern auf einem mehrwertigen Wahrheitsbegriff aufgebaut werden (der daneben auch noch andere Wahrheitswerte kennt). Diese anderen Wahrheitswerte können jedoch kaum als Grade der Wahrheit aufgefaßt werden. Es ist meines Wissens bisher noch niemandem gelungen, genau zu sagen, was sie eigentlich bedeuten.

Behauptungen „Die Hypothese Ψ wird durch das Erfahrungsdatum Φ im Grad r gestützt", „Ψ wird durch Φ im Grad r impliziert" und „Ψ wird durch Φ im Grad r bestätigt" als miteinander gleichwertig angesehen. Untersucht werden muß noch, in welcher Beziehung sie zu Sätzen stehen, die den Wahrscheinlichkeitsbegriff enthalten.

Im Hinblick auf die unterschiedlichen Verwendungen des Wortes „wahrscheinlich" bei den einzelnen Philosophen und Mathematikern wird man berechtigterweise behaupten müssen, daß es nicht korrekt ist, von *dem* Wahrscheinlichkeitsbegriff zu sprechen. Es gibt demnach wenigstens drei Arten von Wahrscheinlichkeitsbegriffen: subjektive, logische und objektive.

Ein genügend entwickelter *subjektiver Wahrscheinlichkeitsbegriff* ist wohl erstmals von David Hume konzipiert worden.[4] Die *subjektive Bewertung* der Wahrscheinlichkeit einer Hypothese bezüglich einer Erfahrung und damit das *Subjekt, das diese Bewertung vollzieht*, eventuell noch relativiert auf den Zeitpunkt, zu dem es dies tut, ist damit *Gegenstand der Wahrscheinlichkeitsaussage*, die nach dieser Deutung deshalb in vollständiger Formulierung lautet: „Die Wahrscheinlichkeit für die Hypothese Ψ ist auf Grund des Erfahrungsdatums Φ für die Person x zur Zeit t die Zahl r." Der Wahrscheinlichkeitsbegriff ist hier also eine fünfstellige Relation zwischen zwei Aussagen, einer Person, einem Zeitpunkt (bzw. einem Zeitintervall) und einer (reellen) Zahl. Da zu seinen Argumenten Sätze einer Sprache zählen, die durch die Buchstaben „Ψ" und „Φ" angedeutet werden, gehört er der Sprache an, in der *über* diese Sprache gesprochen werden kann, also der *Metasprache* dieser Sprache.

Der *logische Wahrscheinlichkeitsbegriff* ist dem subjektiven sehr ähnlich (und wird von Mathematikern mit diesem auch meist identifiziert). Er unterscheidet sich von ihm nur dadurch, daß er, wie der Begriff der logischen Folgerung, *nicht auf Personen und Zeitpunkte relativiert ist.* Er ist also, wie dieser, ein Ausdruck der Metasprache einer vorgegebenen Sprache. Wahrscheinlichkeits*aussagen*, in denen er vorkommt, nehmen nicht auf Personen und Zeitpunkte Bezug und haben demnach die Form: „Die Wahrscheinlichkeit einer Hypothese Ψ auf

---

[4] Zu den gegenwärtigen Hauptvertretern von subjektiven Wahrscheinlichkeitsbegriffen zählen Bruno de Finetti und Leonard J. Savage.

Grund (oder bezüglich) eines Erfahrungsdatums Φ ist r." In diesem Sinn ist der Wahrscheinlichkeitsbegriff bewußt und ausdrücklich erstmals von John Maynard Keynes und, systematisch, von Rudolf Carnap entwickelt worden.

Der *objektive Wahrscheinlichkeitsbegriff* ist der von Mill und Venn konzipierte und von Reichenbach und Mises entwickelte Begriff des Grenzwertes der relativen Häufigkeit.[5] Er wird in den Aussagen „Die Wahrscheinlichkeit der Eigenschaft F in der Bezugsklasse G ist r" und „Der Limes der relativen Häufigkeiten von F in G ist r" verwendet. Die Eigenschaft F, die Klasse G und die reelle Zahl r sind keine sprachlichen Entitäten und können deshalb in einer Sprache benannt werden, in der nicht über eine andere Sprache gesprochen wird, also in einer sogenannten Objektsprache.[6] Folglich kann der *objektive* Wahrscheinlichkeitsbegriff als ein *objektsprachlicher* Ausdruck eingeführt werden.

Der objektive Wahrscheinlichkeitsbegriff ist von dem des Grades an Gewißheit verschieden, da die beiden Aussagen „Die Hypothese Ψ ist bezüglich der Erfahrung Φ im Grad r gewiß" und „Die Wahrscheinlichkeit der Eigenschaft F in der Bezugsklasse G ist r" nicht miteinander gleichwertig sind: Die Wahrheit der ersteren hängt davon ab, welches Erfahrungsdatum Φ gewählt wird; Φ wird in der letzteren jedoch überhaupt nicht erwähnt, und deren Wahrheit ist demnach unabhängig von der Wahl irgendeines Φ.

Andererseits liegt auf der Hand, daß der logische Wahrscheinlichkeitsbegriff mit dem des Grades an Gewißheit zusammenfällt; die Aussage „Ψ ist bezüglich Φ im Grad r wahrscheinlich" besagt ja das gleiche wie „r ist der Grad an Gewißheit, den Ψ bezüglich Φ besitzt". Der subjektive Wahrscheinlichkeitsbegriff schließlich kann lediglich

---

[5] Eine neuere Darstellung dieses Begriffs findet man in Fisher [1].

[6] Diese Aussage muß in zweifacher Hinsicht modifiziert werden: 1. Jede objektsprachliche Aussage kann natürlich auch in der Metasprache dieser Objektsprache formuliert werden, aber nicht umgekehrt; entscheidend ist daher, daß Aussagen, die den *objektiven Wahrscheinlichkeitsbegriff* enthalten, in der *Objektsprache* formulierbar sind, während Aussagen, die den *subjektiven* oder den *logischen Wahrscheinlichkeitsbegriff* enthalten, nicht zu ihr gehören. 2. Wenn eine Sprache genügend ausdrucksreich ist (was insbesondere von der Sprache der erweiterten Quantorenlogik gilt), so ist sie bereits offen oder versteckt eine Metasprache, da es in ihr möglich ist, axiomatisch gewisse einfachere und ausdrucksärmere Sprachen zu charakterisieren. Vgl. hierzu auch Fußnote 15 auf S. 79.

mit einem subjektiven Begriff des Grades an Gewißheit identifiziert werden.

Hume unterscheidet zwischen einem philosophischen und einem unphilosophischen Wahrscheinlichkeitsbegriff.[7] Der philosophische trifft auf Personen x (zur Zeit t) zu, die die Wahrscheinlichkeit von Hypothesen auf Grund von Beobachtungen bis zu einem gewissen Grad rational beurteilen, die sich nicht durch Sympathien, Antipathien und ähnliche irrationale Momente bei der Berechnung der Wahrscheinlichkeit beeinflussen lassen und denen bei dieser Berechnung auch keine logisch-mathematischen Fehler unterlaufen. Der unphilosophische Wahrscheinlichkeitsbegriff hingegen trifft auf Personen x (zur Zeit t) zu, die diesen Bedingungen nicht genügen.

Der Vertreter der subjektiven Wahrscheinlichkeitslehre steht demnach vor folgender Wahl: Entweder er untersucht, wie und nach welchen Regeln eine Person (oder eine Gruppe von Personen) den Wahrscheinlichkeitsbegriff zu einer gewissen Zeit (oder in einer gewissen Zeitspanne) verwendet; dann gehören seine Untersuchungen zur empirischen Psychologie und nicht zur Wahrscheinlichkeitslehre, so wie diese üblicherweise verstanden wird. Es ist auch ausgeschlossen, aus der *tatsächlichen Verwendung* des Wahrscheinlichkeitsbegriffs durch einige Menschen auf seine *korrekte Verwendung* zu schließen, wie man ja auch in der deduktiven Logik nicht aus der Tatsache, daß jemand eine logische Folgerung *behauptet*, darauf schließen kann, daß diese *gilt* und er sich nicht geirrt hat. Oder aber der Anhänger des subjektiven Wahrscheinlichkeitsbegriffs gibt gewisse Rationalitätskriterien an, denen die Personen zu genügen haben; dann ist es sinnvoller, das subjektivistische Gewand überhaupt abzustreifen und mit diesen Kriterien unter Abstrahierung von Personen und Zeitpunkten den Wahrscheinlichkeitsbegriff axiomatisch zu charakterisieren.

Daß in der induktiven Logik die ausdrückliche Bezugnahme auf Personen und Zeitpunkte unnötig und damit der subjektive Wahrscheinlichkeitsbegriff durch den logischen zu ersetzen ist, sieht man besonders deutlich, wenn man sich die analogen Verhältnisse in der deduktiven Logik vergegenwärtigt:

Die logische Folgebeziehung wird üblicherweise als eine Beziehung

---

[7] Vgl. Hume [1], S. 143—155.

zwischen Sätzen verstanden. Die Frage, ob sie zwischen gewissen Sätzen besteht oder nicht, ist unabhängig davon, welche Sachverhalte diese Sätze beschreiben, sondern hängt nur von der logischen Struktur, von der logischen *Form* der Sätze ab, von der Art und Weise also, wie sie mittels logischer Konstanten aus elementaren Sätzen zusammengesetzt sind. Man sagt deshalb, daß die Logik eine *formale* Wissenschaft ist. Sie ist außerdem eine *objektive* und keine subjektive Wissenschaft; denn ob eine logische Folgebeziehung zwischen zwei Sätzen besteht oder nicht, ist *unabhängig* davon, ob jemand den einen Satz aus dem anderen schon gefolgert hat oder jemals deduzieren wird, ob jemand glaubt, daß die Folgebeziehung zwischen den Sätzen besteht oder nicht, und was die einzelnen Menschen oder die Mehrzahl aller Menschen dazu denken. *Nicht unabhängig* von Personen ist natürlich die *Erkenntnis* des Vorliegens einer logischen Folgebeziehung. Bei der Analyse des Problems, ob eine logische Folgebeziehung vorliegt oder nicht, werden einzig und allein die in den Sätzen vorkommenden Ausdrücke in Erwägung gezogen, nicht jedoch, welcher Denkprozeß in demjenigen vorgeht, der untersucht, ob die Folgebeziehung zwischen den Sätzen besteht. Die Definition des Logikbegriffs als der Lehre vom richtigen Denken ist also irreführend und falsch. Bei jeder wissenschaftlichen Betätigung muß gedacht werden, jedoch nur in der Psychologie ist das Denken Gegenstand der Untersuchung.

Analog kann man die logische Deutung des Wahrscheinlichkeitsbegriffs gegen den Ansatz der Subjektivisten rechtfertigen. Um *erkennen* zu können, ob eine Hypothese auf Grund eines Erfahrungsdatums einen gewissen Wahrscheinlichkeitsgrad hat, ist selbstverständlich eine Person notwendig, nicht jedoch dafür, ob die Hypothese auf Grund dieses Wissens diesen Wahrscheinlichkeitsgrad *hat*. Eine Relativierung des Wahrscheinlichkeitsbegriffs auf Personen und Zeitpunkte ist genauso wenig notwendig wie eine entsprechende Relativierung des Begriffs der logischen Folgerung.

Selbstverständlich kann niemand daran gehindert werden, den Wahrscheinlichkeitsbegriff unter Relativierung auf Personen und Zeitpunkte einzuführen, doch stellt sich auch hier wie beim analogen Vorgehen in der deduktiven Logik sofort die Frage, ob dabei nicht die *Definition* des Wahrscheinlichkeitsbegriffs mit seiner *Anwendung* durch

eine Person zu einem bestimmten Zeitpunkt verwechselt wird. Untersuchungsgegenstand der Wahrscheinlichkeitstheorie ist ja nicht der Grad des Glaubens einer *bestimmten Person* an eine Hypothese, sondern der Glaubensgrad einer *streng rationalen Person*, die den Wahrscheinlichkeitsgrad einzig und allein auf Grund der Hypothese und der Prämisse ermittelt. Das ist jedoch gleichwertig damit, daß für die Errechnung des Wahrscheinlichkeitsgrades nur die beiden Sätze, nicht jedoch die Individualität der Person wichtig sind, und daß der Wahrscheinlichkeitsbegriff deshalb von Person und Zeit unabhängig ist.[8]

Es ist, nebenbei bemerkt, irreführend, von *dem* subjektiven Wahrscheinlichkeitsbegriff zu sprechen. Da tatsächlich jeder Mensch den Wahrscheinlichkeitsbegriff verschieden verwendet, zumindest in einigen Nuancen, ja, weil sogar ein und derselbe Mensch ihn zu verschiedenen Zeitpunkten verschieden benützt, gibt es eine unendliche Vielfalt von subjektiven Wahrscheinlichkeitsbegriffen, die von Fall zu Fall durch größere oder geringere Rationalität ausgezeichnet sind. Der subjektivistische Wahrscheinlichkeitstheoretiker erhält außerdem eine Vielfalt von Wahrscheinlichkeitsbegriffen je nachdem, wie er seine Rationalitätskriterien abstuft: ob er nur ein gewisses Maß an innerer Widerspruchslosigkeit verlangt, oder ob er Bedingungen stellt, die nur ein Laplacescher Weltgeist erfüllen kann.

Analog scheint es zunächst unangebracht, von *dem* logischen Wahrscheinlichkeitsbegriff zu sprechen, gibt es doch, wie Carnap gezeigt hat, so viele adäquate induktive Methoden wie positive reelle Zahlen existieren.[9] Andererseits versuche ich in dieser Arbeit zu begründen, daß es ein widerspruchsfreies und vollständiges System von Kriterien und Rationalitätsforderungen gibt, denen eine jede induktive Methode zu genügen hat. Man könnte, wenn man diese Argumente akzeptiert, den Wahrscheinlichkeitsbegriff mit dem System dieser Kriterien oder, was damit äquivalent ist und daher das gleiche besagt, mit der Gesamtheit der Methoden, die allen diesen Kriterien genügen, identifizieren; dann wäre man wiederum berechtigt, den bestimmten Artikel zu verwenden und von *dem* logischen Wahrscheinlichkeitsbegriff zu reden.

---

[8] Zu diesen und ähnlichen Argumenten für den logischen Wahrscheinlichkeitsbegriff vgl. auch die Ausführungen von Carnap [9], S. 37—51.

[9] Vgl. Carnap [11] sowie das Theorem 14 auf S. 149 dieses Buches.

Es bleibt noch die Frage zu diskutieren, ob die Einzigkeitsbedingung auch bezüglich des objektiven Wahrscheinlichkeitsbegriffs erfüllt ist, ob es also wenigstens und höchstens einen derartigen Begriff gibt, so daß man auch hier von *dem* objektiven Wahrscheinlichkeitsbegriff sprechen kann. Dazu ist folgendes zu sagen: (1) Es ist bisher noch überhaupt nicht untersucht worden und folglich auch nicht geklärt, ob die verschiedenen Ansätze zur Charakterisierung des objektiven Wahrscheinlichkeitsbegriffs miteinander logisch-mathematisch äquivalent sind; wenn sich erweisen würde, daß dies nicht der Fall ist, so wäre zusätzlich noch zu untersuchen, ob einige dieser Ansätze Teiltheorien der anderen sind, bzw. ob einige von ihnen inkorrekt sind. (2) Es ist bis heute noch nicht geklärt, ob es in der Natur Entitäten wie objektive Wahrscheinlichkeiten überhaupt gibt. Wenn es sie nicht gibt, dann ist zwar der Begriff der objektiven Wahrscheinlichkeit deshalb noch nicht logisch unerfüllbar, da zumindest einige dieser Theorien widerspruchsfrei sind,[10] aber sie sind dann nichts anderes als nutzlose Spielereien.

Daß überhaupt das Problem entstehen kann, ob objektive Wahrscheinlichkeiten existieren, wird den erstaunen, der bedenkt, daß halbwegs rationale Glücksspieler stets zu schätzen versuchen, wie groß die relative Häufigkeit einer Eigenschaft in einer hinreichend großen Bezugsklasse *auf lange Sicht* ist (etwa der Eigenschaft *Wappen liegt oben* bezüglich der Würfe mit einer gewissen Münze), daß sie also im Grunde nichts anderes als den *Grenzwert* der relativen Häufigkeiten schätzen. Also, wird er argumentieren, muß es doch in der Natur so etwas wie Grenzwerte der relativen Häufigkeiten geben.

Dieser Schluß ist jedoch nicht korrekt, wie die zwei folgenden Gründe zeigen. (1) Es liegt kein Widerspruch in der Annahme, daß es so etwas wie eine unbegrenzte und unabhängige Wiederholbarkeit ein und desselben Experimentes in unserer Welt nicht gibt. Jede zunächst normale Münze wird mit genügend oftmaligen Würfen so deformiert, daß sich dabei unsere *Erwartung* des Grenzwertes beliebig und in der einen oder anderen Richtung von $1/2$ entfernt, und ähnlich kann es sich im Prinzip auch bei anderen Experimenten verhalten.[11] (2) Der

[10] Vgl. Wald [1].

[11] Man kann zwar statistische Gesetze über das Verhalten von Menschen in bestimmten Situationen angeben; doch werden sich diese Gesetzmäßigkeiten im

Glücksspieler berechnet nicht die *objektive Wahrscheinlichkeit*, sondern die *Schätzung dieser objektiven Wahrscheinlichkeit*. Diese Schätzung kann er jedoch auch dann vornehmen, wenn die objektive Wahrscheinlichkeit gar nicht existiert, wenn sich z. B. die Folge der Münzwürfe nicht auf einen Grenzwert einpendelt. Daraus folgt, daß die Existenz einer Schätzung der objektiven Wahrscheinlichkeit nicht die Existenz dieser objektiven Wahrscheinlichkeit selbst nach sich zieht.
Dieser Begriff der Schätzung von objektiven Wahrscheinlichkeiten ist ein Begriff der Metasprache, wie man folgendermaßen sieht. Eine Schätzung muß stets im Hinblick auf ein gewisses Erfahrungsdatum, etwa auf eine bisher beobachtete Regelmäßigkeit oder auf eine statistische Erhebung, vorgenommen werden; wenn kein derartiges Datum zur Verfügung steht, so wird die Schätzung im Hinblick auf das rein logische Wissen um die verschiedenen Möglichkeiten gemacht, also im Hinblick auf eine Prämisse, die mit der Tautologie logisch äquivalent ist. Korrekt wird der Begriff der Schätzung folglich in Sätzen der Form „Die Schätzung der relativen Häufigkeit der Eigenschaft F in der (endlichen) Bezugsklasse G bezüglich des Erfahrungsdatums Φ ist r" und „Die Schätzung der objektiven Wahrscheinlichkeit der Eigenschaft F in der (unendlichen) Bezugsklasse G bezüglich des Erfahrungsdatums Φ ist r" verwendet, die unter gewissen elementaren Voraussetzungen äquivalent sind mit „Die logische Wahrscheinlichkeit des Satzes, der besagt, daß ein noch nicht beobachtetes Ding aus G die Eigenschaft F hat, im Hinblick auf das Beobachtungsdatum Φ ist r".[12] Während der Begriff „objektive Wahrscheinlichkeit" also der Objektsprache angehört, ist „Schätzung der objektiven Wahrscheinlichkeit" ein metasprachlicher Ausdruck; jener kann aus diesem genauso wenig herausgespalten werden wie „alt" aus „kalt" und ist kein selbständiger Bestandteil dieses Begriffs.
Die Existenz objektiver Wahrscheinlichkeiten könnte meines Erachtens höchstens auf einem der beiden folgenden Weisen gezeigt werden: (1) durch Angabe eines Verfahrens, das eine Folge von relativen

Laufe der Jahrtausende sicher in unvorhersagbarer Weise ändern, und wenn die Perioden, in denen sich diese Änderungen vollziehen, immer länger werden, konvergieren die beobachteten relativen Häufigkeiten gerade nicht gegen einen Grenzwert. Ähnliches kann theoretisch für alle Gesetze einer jeden Wissenschaft (und sogar der Physik) gelten.

[12] Vgl. Theorem 23 auf S. 241 f. dieses Buches.

Häufigkeiten erzeugt, wobei nachzuweisen ist, daß diese Folge konvergiert, oder (2) durch effektive Angabe einer solchen unendlichen Folge von relativen Häufigkeiten. Der erste Ausweg ist nicht nur durch den Misesschen Kollektivbegriff verbaut, sondern auch durch die Tatsache, daß es ein solches Verfahren gar nicht geben kann, da wir ja höchstens die vergangenen relativen Häufigkeiten genau kennen, nicht jedoch die, die sich in der Zukunft bei Fortsetzung des Experiments einstellen werden. Aus dem gleichen Grund ist jedoch auch der zweite Ausweg nicht gangbar, denn diese unendliche Folge wäre uns eben nur dann gegeben, wenn wir neben den vergangenen auch die zukünftigen relativen Häufigkeiten kennen würden. Der Misessche Versuch, die Existenz von objektiven Wahrscheinlichkeiten dadurch als empirische Tatsachen zu begründen, daß auf den Erfolg der Spielbanken hingewiesen wird, bei denen die relativen Häufigkeiten *bisher* konvergiert[13] sind, ist unhaltbar, weil, wie schon Hume erkannt hat, aus dieser Beobachtung nicht deduktiv geschlossen werden kann, daß dies auch in der *Zukunft* so sein wird.

Schließlich kann aus der Tatsache, daß einige Kalküle der objektiven Wahrscheinlichkeitstheorie *widerspruchsfrei* sind, nicht geschlossen werden, daß es in *unserer faktischen Welt* ein System gibt, das diesen Kalkül erfüllt; die Widerspruchsfreiheit ist keine hinreichende Bedingung für die Wahrheit von Aussagen.

Vom theoretischen Standpunkt aus unterscheiden sich Theorien der objektiven und der logischen Wahrscheinlichkeit wesentlich, und zwar weniger in den Axiomen, die in beiden Theorien formal ähnlich oder sogar gleich sind, sondern dadurch, daß durch die eine Theorie (die der objektiven Wahrscheinlichkeit) Aussagen über die Welt gemacht werden, durch die andere hingegen (die der logischen Wahrscheinlichkeit) nur über partielle logische Implikationen. Vom praktischen Gesichtspunkt aus ist der Unterschied wiederum gering, da die Vertreter der objektiven Wahrscheinlichkeitslehre ja tatsächlich nicht von objektiven Wahrscheinlichkeiten ausgehen, wie sie vorgeben, sondern von deren Schätzungen, und da sie Axiome und Schlußregeln verwenden, die man ohne Mühe auch als Axiome und Schlußregeln der logischen Wahrscheinlichkeitstheorie auffassen kann.

---

13 Vgl. Mises [2], S. 18f.

Versteht man den Wahrscheinlichkeitsbegriff *objektiv*, so ist die Aussage „Die Wahrscheinlichkeit dafür, daß ein Ding aus G die Eigenschaft F hat, ist r" ein vollständiger Satz, der besagt, daß r die relative Häufigkeit von F in G ist, falls G endlich ist, sowie, daß r der Limes der relativen Häufigkeiten von F in einer Folge von Teilklassen von G ist, falls G unendlich ist und der Limes existiert. Ob eine derartige Aussage wahr oder falsch ist, kann, wenn überhaupt, nur empirisch entschieden werden. Versteht man den Wahrscheinlichkeitsbegriff hingegen *logisch*, so ist jene Aussage genauso unvollständig wie die Wortfolge „Ψ folgt logisch", da in beiden Fällen die Prämissen des Schlusses nicht genannt werden. Erst die Bezugnahme auf das Erfahrungsdatum gewährleistet eine vollständige und logisch determinierte Wahrscheinlichkeitsaussage. Behaupte ich etwa: „Die Wahrscheinlichkeit dafür, daß es morgen regnen wird, ist $^1/_6$", so muß ich mich, damit dieser Satz sinnvoll wird, direkt oder indirekt auf ein Wissen beziehen, etwa auf die Wettersituation der letzten Stunden oder der letzten Tage, auf Wetterabläufe in vergleichbaren Jahreszeiten früherer Jahre usw. Wenn ich hingegen noch kein relevantes Wissen zur Verfügung habe und mich also auf kein solches beziehen kann, so muß ich deutlichmachen, daß dies der Fall ist und meine Aussage den Charakter einer Aprioriwahrscheinlichkeit hat, einer Wahrscheinlichkeit, die sich lediglich auf das Wissen um die rein logisch möglichen Fälle bezieht.

Auf das zur Verfügung stehende Wissen wird scheinbar am einfachsten durch eine Konditionalaussage der folgenden Form Bezug genommen: „Wenn Φ wahr ist, dann ist die Wahrscheinlichkeit von Ψ gleich r" oder, um ein konkretes Beispiel zu benützen: „Wenn es wahr ist, daß die relative Häufigkeit der Katholiken in Schweden $^1/_{20}$ beträgt und Sven Svensson ein Schwede ist, dann ist die Wahrscheinlichkeit, daß Sven Svensson ein Katholik ist, gleich $^1/_{20}$."

Der Wahrscheinlichkeitsbegriff bezieht sich bei dieser Formulierung nur auf Ψ, nicht jedoch auf Φ; die Wahrscheinlichkeit ist hier also nur eine Funktion von Ψ und nicht von Φ abhängig. Uns möge nun folgendes Wissen zur Verfügung stehen: „Es ist wahr, daß die relative Häufigkeit der Katholiken in Schweden $^1/_{20}$ ist, daß Sven Svensson

ein Schwede ist, und daß er im vergangenen Monat nach Lourdes gepilgert ist."

Diese beiden Aussagen implizieren logisch den Satz: „Die Wahrscheinlichkeit, daß Sven Svensson ein Katholik ist, ist $1/20$." Dieser Satz ist jedoch auf Grund des gegebenen Wissens nicht gerechtfertigt; denn aus der Tatsache, daß Sven Svensson nach Lourdes gepilgert ist, folgt mit hoher Wahrscheinlichkeit, daß er ein Katholik ist. Die konditionale Form der Wahrscheinlichkeitsaussagen führt also in bestimmten Fällen zu inadäquaten Resultaten, und es liegt deshalb nahe, den Wahrscheinlichkeitssätzen folgende Form zu geben: „Die Wahrscheinlichkeit dafür, daß Sven Svensson ein Katholik ist, ist im Hinblick darauf, daß er ein Schwede ist und die relative Häufigkeit der Katholiken unter den Schweden $1/20$ ist, $1/20$." Oder allgemeiner formuliert: „Die Wahrscheinlichkeit einer Hypothese Ψ im Hinblick auf (relativ auf) ein Wissen Φ ist r."

Der Wahrscheinlichkeitsbegriff bezieht sich bei dieser Formulierung also nicht auf Ψ allein, sondern auf die Relation, in der Ψ zu Φ steht; sie gibt den Grad an, in dem Ψ durch Φ impliziert wird. Es kann nun aus dem Wissen, daß $1/20$ der Schweden Katholiken sind, Sven Svensson ein Schwede ist und daß er nach Lourdes gepilgert ist, nicht mehr darauf geschlossen werden, daß die Wahrscheinlichkeit dafür, daß er ein Katholik ist, $1/20$ ist. Denn die Wahrscheinlichkeit gilt nicht im Hinblick auf dieses Wissen, sondern im Hinblick darauf, daß $1/20$ der Schweden Katholiken sind und Sven Svensson ein Schwede ist.

Wenn r die Wahrscheinlichkeit von Ψ bezüglich des Wissens ist, das aus Φ und Γ besteht, und s die Wahrscheinlichkeit von Ψ relativ zum Wissen Φ allein, so wird gewöhnlich r nicht mit s identisch sein. Die induktive Logik unterscheidet sich nun von der deduktiven dadurch wesentlich, daß sie einen Grundsatz für die *Anwendung* von induktiven Argumenten kennt, der dieser fremd ist, nämlich die *Forderung des Gesamtdatums*: Zwar kann die Wahrscheinlichkeit einer Hypothese Ψ bezüglich eines jeden möglichen Erfahrungsdatums Φ *berechnet* werden, jedoch darf nur jene Wahrscheinlichkeitsaussage in einem konkreten Fall *angewandt* werden, in der Φ das *gesamte* zur Verfügung stehende Wissen (oder, falls der Relevanzbegriff schon zur Verfügung steht: in der Φ das *gesamte für* Ψ *relevante* Wissen) beschreibt. Tut man dies nicht, so erhält man in folgender Hinsicht falsche Resultate: Es

gibt Fälle, in denen dann die mittels einer induktiven Logik gewonnenen Wahrscheinlichkeitswerte von denen, die man durch intuitive Überlegungen erhält, in einem Grad abweichen, der nicht mehr toleriert und als Fehleinschätzung der Intuition entschuldigt werden kann.

Dies wird unmittelbar klar, wenn man sich vergegenwärtigt, daß zwar in der *deduktiven* Logik beim Auftreten von Zusatzerfahrungen logische Folgebeziehungen erhalten bleiben, daß also, falls Ψ aus dem Wissen Φ logisch folgt, Ψ auch aus Φ und dem Zusatzwissen Γ logisch folgt, daß in der *induktiven* Logik jedoch ein derartiges Zusatzwissen Γ den Wahrscheinlichkeitsgrad stark verändern kann, und zwar in beiden Richtungen, je nachdem, ob es die Hypothese Ψ plausibler oder unplausibler macht (ob es für Ψ bezüglich Φ positiv oder negativ relevant ist).

Folgendes Beispiel möge das soeben Gesagte verdeutlichen: Die Hypothese Ψ besage, daß morgen in München schönes Wetter ist; Φ sei das Wissen, daß in München seit zehn Tagen schönes Wetter ist und daß *bisher* (soweit man sich entsinnen kann) in 80 von 100 Fällen auf zehn schöne Tage in München ein elfter schöner Tag gefolgt ist; und Γ sei schließlich die Zusatzinformation, die beinhaltet, daß heute in München das Barometer stark gesunken ist, verbunden mit der Erfahrung, daß *bisher* (soweit man sich entsinnen kann) in 90 von 100 Fällen in München am nächsten Tag schlechtes Wetter eingetreten ist. Dann wird die Wahrscheinlichkeit von Ψ relativ zu Φ allein in der Nähe von 0,8 liegen (und zwar wird sie um so näher bei ihr liegen, je umfangreicher die in Φ beschriebene Gesamtheit von Wetterbeobachtungen ist), während sich die Wahrscheinlichkeit von Ψ bezüglich Φ *und* Γ vielleicht in der Nähe von 0,35 bewegt (da man ja in München in solchen Fällen immer noch auf einen Föhnsturm hoffen kann, der einen blauen Himmel bringt).

Daß die meisten der sogenannten Objektivisten der Ansicht gewesen sind, ihre Interpretation des Wahrscheinlichkeitsbegriffs sei die einzig richtige, hat vielleicht auch folgende Ursache: Zunächst haben sie die Wahrscheinlichkeitsaussagen in jener unkorrekten konditionalen Weise formuliert. Da das Erfahrungsdatum vorgegeben ist, haben sie mittels *modus ponens* den Satz „Die Wahrscheinlichkeit von Ψ ist r" erschlossen. Nun ist die Wahrscheinlichkeit dafür, daß ein beliebig

vorgegebenes Ding aus einer Gesamtheit G eine gewisse Eigenschaft F hat, identisch mit der Schätzung der relativen Häufigkeit von F in G; also haben sie jenen Satz gelesen als: „Die Schätzung der relativen Häufigkeit von F in G ist r". Nun ist die Frage naheliegend, was unter dem Begriff der Schätzung zu verstehen sei; sicherlich könne als Schätzung *der* relativen Häufigkeit nicht eine zufällig beobachtete relative Häufigkeit genommen werden, vielmehr sei dazu jener Wert zu verwenden, auf den sich die relativen Häufigkeiten einpendeln, wenn die Bezugsklasse immer mehr vergrößert wird, also die relative Häufigkeit *auf lange Sicht* bzw. den *Limes* der relativen Häufigkeiten. Daher wird die *Schätzung* der relativen Häufigkeit mit dem Limes der relativen Häufigkeiten identifiziert.[14]

Daß dies nicht korrekt ist, sieht man sofort, wenn man sich die Formen der beiden Begriffe vergegenwärtigt. Der objektive Wahrscheinlichkeitsbegriff, der in Aussagen der Gestalt „Die Wahrscheinlichkeit von F in G ist r" verwendet wird, ist eine dreistellige Relation zwischen zwei Klassen und einer reellen Zahl; er ist also, wie schon bemerkt worden ist, in einer genügend ausdrucksreichen Objektsprache ausdrückbar.[15] Der Begriff der Schätzung ist unter Zugrundelegung des logischen Wahrscheinlichkeitsbegriffs definierbar. Grob gesprochen ist die auf Grund von Φ vorgenommene Schätzung der relativen Häufigkeit von F in G gleich der Wahrscheinlichkeit, daß ein Ding aus G die Eigenschaft F hat, relativ auf das Datum Φ. Nun ist der logische Wahrscheinlichkeitsbegriff eine dreistellige Beziehung zwischen zwei Sätzen und einer reellen Zahl und der Schätzungsbegriff eine vierstellige Beziehung zwischen zwei Klassen (Eigenschaften), einem Satz und einer reellen Zahl. Da diese Beziehungen und die damit formulierbaren Aussagen beide Male auf Sätze der Objektspra-

---

14 Dies ist neben der auf S. 60 ff. angeführten eine zweite Motivation für die Ansicht von Reichenbach, Mises und anderen, daß der Wahrscheinlichkeitsbegriff in ihrem Sinn als eine in der Objektsprache ausdrückbare Relation verstanden werden müsse. Auf eine dritte Motivation dafür, die sich aus einer mißverständlichen Formulierung des Theorems von Bernoulli ergibt, wurde von Keynes [1], S. 92f., und von Carnap-Stegmüller [1], S. 62f., hingewiesen.

15 In der Logik wird die Umgangssprache in Sprachstufen eingeteilt. Zur Objektsprache werden jene Sätze gerechnet, die nur über außersprachliche Entitäten sprechen. Sätze der Metasprache können darüber hinaus auch Ausdrücke und Sätze der Objektsprache zum Gegenstand haben, Sätze der Metametasprache auch Ausdrücke und Sätze der Metasprache, und so fort.

che bezogen sind, gehören sie der Metasprache an. Der objektive Wahrscheinlichkeitsbegriff ist also nicht eine bessere Explikation des vorwissenschaftlichen Ausdrucks als der logische, sondern ein Begriff mit völlig andersartiger Struktur. Dies kommt darüberhinaus auch dadurch zum Ausdruck, daß objektiv gedeutete Wahrscheinlichkeitsaussagen empirisch wahr oder falsch, logisch gedeutete hingegen logisch wahr oder falsch sind.

Der logische Wahrscheinlichkeitsbegriff kann als dreistellige Beziehung zwischen zwei Sätzen und einer reellen Zahl eingeführt werden, die bezüglich der dritten Argumentstelle eindeutig ist, so daß also zwei vorgegebene Sätze nur mit einer reellen Zahl in dieser Beziehung stehen, und er kann als Funktion charakterisiert werden, die einem geeigneten Paar von Sätzen eine reelle Zahl als Wert zuordnet (wobei im allgemeinen noch vorausgesetzt wird, daß jener Satz, der das Erfahrungsdatum darstellt, nicht logisch falsch ist). Beide Möglichkeiten sind miteinander logisch äquivalent. Da die funktionale Schreibweise größere praktische Vorteile bietet, wird sie im folgenden verwendet.

Vom Begriff der Wahrscheinlichkeit[16] ist der des Zuwachses an Wahrscheinlichkeit zu unterscheiden.[17] Die Begriffe „Bewährung" und „Bestätigung", die anfänglich von Carnap, Hempel, Popper und anderen anstelle des Wahrscheinlichkeitsbegriffs benützt worden sind, verleiten dazu, diesen Unterschied zu verwischen, was Widersprüche nach sich ziehen kann. Der Zuwachs an Wahrscheinlichkeit, den eine Hypothese Ψ durch eine neue Erfahrung Γ bezüglich des schon bekannten Wissens Φ erhält, kann identifiziert werden mit der Differenz zweier Wahrscheinlichkeiten, nämlich mit der Wahrscheinlichkeit von Ψ bezüglich Φ und Γ abzüglich der Wahrscheinlichkeit von Ψ bezüglich Φ allein. Wenn die Wahrscheinlichkeitsfunktion Werte von 0 bis 1 annimmt, so liefert diese Zuwachsfunktion Zahlenwerte

---

[16] Mit den Ausdrücken „Wahrscheinlichkeit" und „Wahrscheinlichkeitsbegriff" ist im folgenden immer, wenn nicht ausdrücklich etwas anderes gesagt wird, die *logische* Wahrscheinlichkeit bzw. der *logische* Wahrscheinlichkeitsbegriff gemeint.

[17] Vgl. Carnap [9], 2. Auflage, S. XVff. Auf die Wichtigkeit dieser Unterscheidung hat erstmals Yehoshua Bar-Hillel hingewiesen, als er zeigte, daß Poppers Behauptung, Carnaps Wahrscheinlichkeitstheorie sei widerspruchsvoll, falsch ist. Vgl. Bar-Hillel [4].

von —1 bis +1. Ist der Zuwachs identisch mit 0, so ist Γ für Ψ bezüglich Φ *irrelevant*; ist der Zuwachs größer als 0, so ist Γ für Ψ bezüglich Φ *positiv relevant*, und ist er kleiner als 0, so ist Γ *negativ relevant*. Γ ist für Ψ bezüglich Φ *relevant*, wenn es für Ψ bezüglich Φ positiv oder negativ relevant, also nicht irrelevant ist. Die Begriffe der Relevanz und der Irrelevanz sind also auf den Wahrscheinlichkeitsbegriff zurückführbar; dieser sollte daher zweckmäßigerweise für möglichst alle Satzpaare definiert sein und nicht nur für solche, bei denen das Datum für die Hypothese relevant ist.

3. Qualitative, komparative und quantitative Wahrscheinlichkeit

Zur Formulierung von Wirklichkeitserkenntnissen werden qualitative (= klassifikatorische), komparative (= topologische) und quantitative (= metrische) Begriffe verwendet. Diese Reihenfolge der Begriffsarten scheint auf den ersten Blick natürlich zu sein, da doch die komparativen Begriffe ein größeres Maß an Information liefern als die qualitativen und die quantitativen wiederum ein größeres als die komparativen. So erlaubt etwa die Begriffsgruppe „sehr kalt", „kalt", „lauwarm", „warm", „heiß" und „sehr heiß" eine vollständige Einteilung des Gegenstandsbereichs in einander ausschließende Teilklassen. Innerhalb der einzelnen Klassen hingegen, etwa der Gesamtheit der warmen Gegenstände, ist eine weitere Differenzierung nicht möglich, es sei denn, man ersetzt diese Einteilung durch eine andere, detailliertere. Das Begriffspaar „wärmer als" und „gleich warm wie" bewirkt eine genaue Ordnung der Gegenstände nach ihrer Temperatur, und der Begriff „die Temperatur von" ordnet den Gegenständen darüberhinaus Zahlenwerte als Temperaturen zu. Es scheint also natürlich zu sein, bei einer Explikation eines erfahrungswissenschaftlichen Begriffs zunächst seine qualitative Form zu analysieren, sodann, wenn dies geglückt ist, seine komparative Form genau zu charakterisieren und diese schließlich auf eine der üblichen Weisen zu metrisieren.

Bezüglich des Wahrscheinlichkeitsbegriffs hat Hempel versucht, diesen Weg zu beschreiten. Da jedoch bereits über die Adäquatheit sei-

nes Explikationsvorschlags für den qualitativen Begriff eine heftige Diskussion entbrannt ist, ist sein Vorhaben schon an diesem Punkt abgebrochen worden.

Andererseits sind unsere Vorstellungen von qualitativen und komparativen Begriffen häufig recht unbestimmt und teilweise sogar widerspruchsvoll, wie John Locke an einem Beispiel gezeigt hat.[18] Erst unter Verwendung des metrischen Temperaturbegriffs können die entsprechenden qualitativen und komparativen Begriffe scharf umrissen werden. Carnap hat daher den umgekehrten Weg eingeschlagen und versucht, zuerst den quantitativen Wahrscheinlichkeitsbegriff adäquat zu charakterisieren und sodann die anderen Begriffe auf diesen zurückzuführen. Unter Benützung der Abkürzung „c" für „confirmation" („Bestätigung", „logische Wahrscheinlichkeit") gibt er die Aussagen „Ψ wird durch Φ im Grad r bestätigt" und „Die Wahrscheinlichkeit von Ψ bezüglich Φ ist r" durch „$c(\Psi, \Phi) = r$" wieder.

Der qualitative Begriff, der in Sätzen der Art „Ψ ist bezüglich Φ wahrscheinlich" verwendet wird, kann dann durch „$c(\Psi, \Phi) > 1/2$" charakterisiert werden. Der komparative Wahrscheinlichkeitsbegriff wird in drei Arten benützt, je nachdem, was miteinander verglichen wird. So findet man ihn in Sätzen der Form „Ψ ist bezüglich Φ wahrscheinlicher als Δ es bezüglich des gleichen Datums ist", was durch „$c(\Psi, \Phi) > c(\Delta, \Phi)$" symbolisch wiedergegeben werden kann. Der Naturwissenschaftler kann sich jedoch auch fragen, wie sich die Wahrscheinlichkeit einer Hypothese ändert, wenn er plötzlich dies und jenes beobachten sollte. Er benützt dann Sätze der Art „Die Wahrscheinlichkeit von Ψ ist auf Grund von Φ größer als auf Grund von Γ", was durch „$c(\Psi, \Phi) > c(\Psi, \Gamma)$" abgekürzt werden kann. Beide Verwendungen sind Spezialfälle des Satzes „Die Wahrscheinlichkeit von Ψ bezüglich Φ ist größer als die Wahrscheinlichkeit von Δ bezüglich Γ", symbolisch wiedergegeben durch „$c(\Psi, \Phi) > c(\Delta, \Gamma)$." Ein Spezialfall des Begriffs der komparativen Wahrscheinlichkeit ist der der Relevanz. Es sei „Λ" eine Abkürzung für den objektsprach-

---

[18] Gegeben seien drei mit kaltem, warmem und heißem Wasser gefüllte Eimer. Stecken wir zunächst die eine Hand in das kalte Wasser und die andere in das heiße und darauf alle beide in das warme, so empfinden wir dieses mit der einen Hand als heiß und mit der anderen als kalt.

lichen Ausdruck „und" und „$\Phi \wedge \Gamma$" daher ein Mitteilungszeichen für die Konjunktion zweier Sätze der Objektsprache, für die die Buchstaben „Φ" und „Γ" Mitteilungszeichen sind. Daß Γ für Ψ bezüglich Φ *positiv relevant* ist, besagt ja soviel wie, daß die Wahrscheinlichkeit von Ψ auf Grund von $\Phi \wedge \Gamma$ größer ist als auf Grund von Φ allein, daß also $c(\Psi, \Phi \wedge \Gamma) > c(\Psi, \Phi)$, was gleichwertig damit ist, daß $c(\Psi, \Phi \wedge \Gamma) - c(\Psi, \Phi) > 0$. Andererseits besagt die Behauptung, daß Γ für Ψ bezüglich Φ *negativ relevant* ist, das gleiche wie, daß die Wahrscheinlichkeit von Ψ auf Grund von $\Phi \wedge \Gamma$ kleiner ist als auf Grund von Φ allein, daß also $c(\Psi, \Phi) > c(\Psi, \Phi \wedge \Gamma)$ und daß somit $c(\Psi, \Phi \wedge \Gamma) - c(\Psi, \Phi) < 0$. Schließlich wird durch die Aussage, daß Γ für Ψ bezüglich Φ *irrelevant* ist, dasselbe behauptet wie dadurch, daß weder $c(\Psi, \Phi \wedge \Gamma) > c(\Psi, \Phi)$ noch $c(\Psi, \Phi) > c(\Psi, \Phi \wedge \Gamma)$, daß also $c(\Psi, \Phi \wedge \Gamma) = c(\Psi, \Phi)$ und folglich $c(\Psi, \Phi \wedge \Gamma) - c(\Psi, \Phi) = 0$.

4. Über Möglichkeit und Grenzen der induktiven Logik

Die Frage der Möglichkeit einer induktiven Logik als Wissenschaft und der Grenzen ihrer Anwendbarkeit hat Carnap systematisch und in allen Einzelheiten behandelt und beantwortet.[19] Er hat dabei untersucht, welche Einwände gegen die Explizierbarkeit dieses Begriffs vorgebracht worden sind oder werden können und hat gezeigt, wie sie zu entkräften sind.

So ist behauptet worden, es sei unmöglich, ein System der quantitativen induktiven Logik zu entwerfen. Wenn damit gesagt werden soll, daß jede derartige Theorie widerspruchsvoll ist, dann ist diese Aussage sicherlich falsch. Denn es hat sich erwiesen, daß es möglich ist, den Begriff der induktiven Bestätigung als quantitativen Begriff widerspruchsfrei zu definieren, womit der Einwand in dieser Form widerlegt ist.

Die Behauptung kann jedoch auch so verstanden werden, daß ein derartiger quantitativer Begriff künstlich sei, da im Alltag der Wahr-

---

[19] Vgl. Carnap [9], II. und IV. Kapitel.

scheinlichkeitsbegriff bzw. der Begriff des Bestätigungsgrades klassifikatorisch oder höchstens komparativ, jedoch nicht quantitativ verwendet werde; er könne deshalb nicht das Explikat irgendeines intuitiven Begriffs sein. Dem ist folgendes zu entgegnen: Selbst wenn dies der Fall wäre, so würde dies nichts gegen den quantitativen Wahrscheinlichkeitsbegriff aussagen. Der Temperaturbegriff ist ebenfalls sehr lange nur in klassifikatorischer und in komparativer Form benützt worden. Obwohl seine metrische Form zunächst nur zum Vokabular einer künstlich geschaffenen Wissenschaftssprache gehört hat, ist sie mit der allgemeinen Verwendung des Thermometers doch auch ein Ausdruck der Alltagssprache geworden. Es gibt jedoch gute Gründe dafür, daß der Begriff der logischen Wahrscheinlichkeit (oder, was im folgenden das gleiche bedeuten soll, der Begriff der induktiven Bestätigung) auch im Alltag in quantitativer Form benützt wird. Dies ist insbesondere dann der Fall, wenn man beurteilt, ob eine angebotene Wette günstig, ungünstig oder für beide Partner fair ist. Ein anderes, viel ernsteres Problem ist es, daß es nicht nur *eine* reguläre Bestätigungsfunktion, sondern *unendlich viele* gibt, und daß bisher keine allgemein akzeptierten Kriterien bekannt sind, die die Auswahl einer einzigen induktiven Methode aus dieser Gesamtheit rechtfertigen würden. Dieses Problem soll an späterer Stelle ausführlich erörtert werden.

Unter einer induktiven Methode wird im folgenden eine Funktion verstanden, die angibt, in welchem Grad ein Satz Ψ durch einen nicht logisch falschen Satz Φ bestätigt wird, die also jedem Paar, bestehend aus einem Satz und einem nicht logisch falschen Satz, eine reelle Zahl zuordnet. Diese Funktion ist dabei normiert, sie nimmt also nur Werte zwischen 0 und 1 an. Eine induktive Methode in diesem Sinn ist kein Verfahren, das von einem gegebenen Erfahrungsdatum zu einer wissenschaftlichen Theorie führt, unter die alles Erfahrungswissen subsumierbar ist. Zur Auffindung einer derartigen Hypothese sind stets neben rationalen Faktoren wie Wissen und Fähigkeit zum Schlußfolgern auch irrationale Komponenten wie Glück, Intuition oder Geschicklichkeit in der Anwendung der korrekten Argumentationsmöglichkeiten nötig. Erst wenn eine Hypothese und ein Erfahrungsdatum zur Verfügung stehen, kann durch eine induktive Methode angegeben werden, in welchem Grad die Hypothese durch das

Wissen bestätigt bzw. gestützt wird. Eine so verstandene induktive Methode ist also kein *Entdeckungsverfahren* von Hypothesen auf der Grundlage eines Erfahrungsdatums, sondern ein *Bestätigungsverfahren* (eine Bestätigungsmethode) für Hypothesen bezüglich eines derartigen Datums; die Hypothese muß bereits gefunden und formuliert sein, damit das Verfahren zur Ermittlung des Bestätigungs- und Stützungsgrades angewandt werden kann.

Die Methoden der mathematischen Statistik (z. B. Schätzung von relativen Häufigkeiten, Kurvenanpassungen, Bedeutungstests), die teilweise unabhängig voneinander entwickelt worden sind, können bei einer hinreichend entwickelten Theorie der induktiven Logik auf den logischen Wahrscheinlichkeitsbegriff zurückgeführt werden. Durch die induktive Logik sollen also nicht nur die Grundbegriffe der Wahrscheinlichkeitslehre analysiert und expliziert werden, es soll durch sie auch eine systematische Einheit in die Vielzahl der Wahrscheinlichkeitsmethoden gebracht werden.

Nicht nur für die Mathematik, sondern auch für die Philosophie und insbesondere für die Wissenschaftstheorie ist eine Explikation des Begriffs der induktiven Bestätigung von Aussagen durch andere dringend erforderlich. Die Begriffe der statistischen Voraussage und Erklärung und andere Zentralbegriffe dieser Disziplin können nur unter Bezugnahme auf eine bestimmte induktive Methode adäquat definiert werden.

Bei der Anwendung der induktiven Logik stützt man sich selten auf ein System von exakt formulierten Regeln und Gesetzen, sondern argumentiert meist intuitiv und je nach Geschick mit wechselndem Erfolg. Das Analoge gilt für die deduktive Logik. Hier allerdings stehen schon seit langem exakt formulierte und als adäquat bewiesene Verfahren zur Verfügung, mit deren Hilfe die Intuition überprüft und kontrolliert werden kann. Wie das formale induktive Argumentieren hat auch das deduktive formale Schließen Grenzen der Anwendbarkeit. Wenn die Prämissen und die Konklusion eines Schlusses nicht in eine Wissenschaftssprache übersetzt werden können, ist eine strenge Überprüfung von vornherein ausgeschlossen, und selbst wenn sie in einer solchen Sprache formuliert worden sind, führt diese Methode wegen der Unentscheidbarkeit der formalen Logik nicht mechanisch zu einem Ergebnis. Trotzdem hat sich die formale Logik

bei der axiomatischen Darstellung von einzelwissenschaftlichen Theorien und bei methodologischen Untersuchungen als nützliches und unentbehrliches Hilfsmittel erwiesen, da in der Mehrzahl der Fälle keine derartigen praktischen Schwierigkeiten ihrer Anwendung im Wege stehen. Das gleiche gilt auch für die induktive Logik. Wie die deduktive Logik ist sie eine intuitiv ständig benützte Methode, die deshalb keine im Prinzip neuen Verfahren zur wissenschaftlichen Erkenntnis liefert, sondern alte überprüft, sie gegebenenfalls modifiziert und ihre Adäquatheit erweist.

In den beiden folgenden Kapiteln werden Kriterien entwickelt, von denen man mit Recht behaupten darf, daß sie notwendige Bedingungen für korrekte Wahrscheinlichkeitsaussagen sind. Dabei werden sich zwei überraschende Resultate ergeben: 1. Es gibt unendlich viele voneinander verschiedene induktive Methoden, die diesen Kriterien genügen, und es ist ausgeschlossen, durch zusätzliche Adäquatheitsforderungen alle bis auf eine von ihnen auszuschließen. 2. Obwohl man diese Kriterien als adäquat ansehen kann und wohl auch muß, führen sie zu Konsequenzen, die den intuitiven Vorstellungen, die man zunächst vom Wahrscheinlichkeitsbegriff hat, offensichtlich widersprechen; es erweist sich nämlich, *daß in unendlichen Gegenstandsbereichen Allbehauptungen bezüglich endlicher Erfahrungsdaten stets den Wahrscheinlichkeitswert 0 erhalten.*[20]

Auf dieses Ergebnis sind, logisch gesehen, zwei Reaktionen möglich: (a) Man kann diese intuitiven Vorstellungen als für den Wahrscheinlichkeitsbegriff wesentlich und unumstößlich ansehen und mit *modus tollens* also darauf schließen, daß nicht alle der Adäquatheitsbedingungen tatsächlich adäquat sind; die wohl kaum zu lösende Schwierigkeit besteht dann darin, anzugeben, welche dieser Bedingungen fallenzulassen sind. (b) Man kann andererseits auch an den Adäquatheitsbedingungen festhalten und mit *modus ponens* darauf schließen, daß unsere intuitiven Vorstellungen bezüglich des Wahrscheinlichkeitsbegriffs in diesem Punkt korrigiert werden müssen; es ist dies der Schluß, den ich zu ziehen vorschlage.

Daß die intuitiven Vorstellungen, die einen vorwissenschaftlichen Begriff betreffen, bei dessen Überführung in ein wissenschaftliches

---

[20] Vgl. Theorem 20 auf S. 218 dieser Arbeit.

System teilweise korrigiert werden müssen, daß also der durch ein solches System explizierte wissenschaftliche Begriff von dem zu explizierenden vorwissenschaftlichen Begriff in gewissen Hinsichten stark abweicht, ist durchaus kein seltenes Phänomen, sondern eine Begleiterscheinung des wissenschaftlichen Fortschritts. So kann die Intension des vorwissenschaftlichen Begriffs „Fisch" etwa durch „ständig im Wasser lebendes Tier" wiedergegeben werden, während der biologische Begriff „Fisch" als „durch Kiemen atmender Kaltblütler" definiert wird; eine derartige Abweichung hat sich als zweckmäßig erwiesen, da man dann mehr und einfachere Gesetzesaussagen formulieren kann.

Bezüglich des Wahrscheinlichkeitsbegriffs ist eine Abweichung vom vorwissenschaftlichen Verständnis in der einen oder anderen Richtung nicht nur aus *praktischen* Gründen angebracht, sondern auch aus *logischen* Gründen notwendig: Das vorwissenschaftliche Verständnis dieses Begriffes, so wie es sich im Laufe der Zeit entwickelt hat, ist nicht konsistent, und man kann nicht gleichzeitig an den Adäquatheitsbedingungen festhalten und von den durch sie charakterisierten induktiven Methoden erwarten, daß sie in unendlichen Gegenstandsbereichen Allsätzen bezüglich endlicher Erfahrungsdaten positive Wahrscheinlichkeitswerte zuordnen.

III. Logische Grundlagen der Wahrscheinlichkeit

1. Reguläre Bestätigungsmethoden

Die induktive Logik ist von jeher als eine Disziplin verstanden worden, die der deduktiven Logik als der Logik im engeren Sinn gegenübersteht und sie gleichzeitig ergänzt. Zur Klärung der Frage, was ein induktiver Schluß sei, ist es daher nützlich, sich dasselbe bezüglich des deduktiven Schlusses zu fragen. Nun charakterisieren manche Autoren im Anschluß an Aristoteles den deduktiven Schluß als ein Argument, das vom Allgemeinen zum Besonderen führe, und den induktiven Schluß dann folgerichtig als ein Verfahren, das vom Besonderen zum Allgemeinen führe. Es kann nicht geleugnet werden, daß ein deduktiver Schluß *oft* diese formale Gestalt hat, aber es entspricht nicht den Tatsachen, daß dies *immer* so ist. Dann wird man folgern müssen, daß der induktive Schluß zwar manchmal, aber nicht immer vom Besonderen zum Allgemeinen führt und daß seine Charakterisierung anders zu erfolgen hat.

Wird aus einem Satz eine andere Aussage deduziert, so ist a priori gewiß, daß die Konklusion wahr ist, falls die Prämisse gilt, daß also das gleichzeitige Wahrsein der Prämisse und Falschsein der Konklusion nicht möglich ist. Dieser Folgerungsbegriff wird gewöhnlich so charakterisiert:[1] Die zugrundegelegte Sprache enthält neben den logischen Konstanten (der Kopula „ist", den aussagenlogischen Konstanten „nicht", „und", „oder", „wenn — dann", „genau dann, wenn", und den Quantoren „alle", „es gibt") Gegenstands-, Eigenschafts- und Beziehungsausdrücke. Bei einer Interpretation J dieser Sprache über einem vorgegebenen Bereich B bezeichnen die Gegenstandsausdrücke Dinge aus B, die Eigenschaftsausdrücke Eigenschaften von Dingen aus B und die Beziehungsausdrücke Beziehungen zwischen Dingen aus B. Ein Satz ist wahr unter einer Interpre-

---

1 Vgl. z. B. Hans Hermes, Einführung in die mathematische Logik, Tübingen 1963, S. 22 und 78, oder Essler [1], S. 31ff. und 197ff.

tation J über einen Bereich B dann und nur dann, wenn der Sachverhalt, den er unter J beschreibt, besteht, wenn also das, was er unter J aussagt, der Fall ist. Die Wahrheit ist demnach eine dreistellige Relation, die erst dann als Eigenschaft (als einstellige Relation) angesehen werden kann, wenn Gegenstandsbereich und Interpretation fest vorgegeben sind, wie etwa bei den natürlichen Sprachen. Ein Satz Ψ folgt ferner logisch aus einem Satz Φ genau dann, wenn Ψ bei allen Interpretationen über jedem Bereich[2] wahr ist, bei denen Φ wahr ist, wenn es also keine Interpretation über einen Bereich gibt, bei der zwar Φ, nicht aber Ψ wahr ist.

Versteht man unter dem *Spielraum* von Δ [kurz: unter dem sr (Δ)] die Gesamtheit der Interpretationen über irgendwelche Bereiche, unter denen Δ wahr ist, unter dem *Gehalt* von Δ hingegen die Klasse der Interpretationen, bei denen Δ falsch ist (intuitiv: die Gesamtheit dessen, was Δ ausschließt), so folgt Ψ logisch aus Φ genau dann, wenn der Spielraum von Φ im Spielraum von Ψ enthalten ist; dies ist wiederum genau dann der Fall, wenn der Gehalt von Ψ echter oder unechter Teil des Gehalts von Φ ist. Die Gesamtheit aller Interpretationen sei in Fig. 1a und b jeweils durch eine Anzahl von Punkten innerhalb der Quadrate charakterisiert. Dann kann die logische Folgerung so veranschaulicht werden:

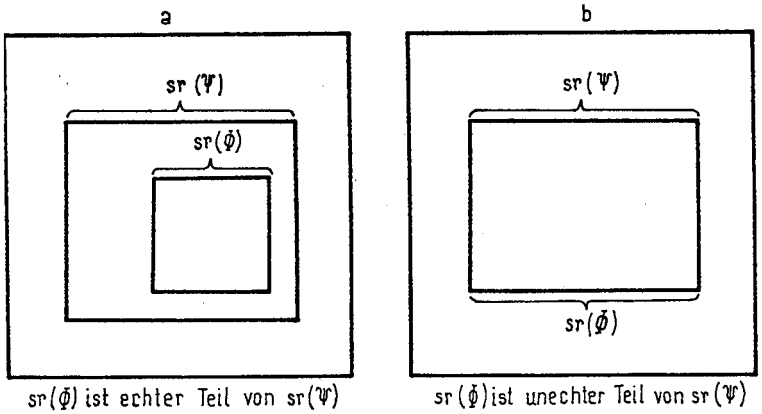


Fig. 1

[2] Der Interpretationsbegriff sei so eingeführt, daß er eine Funktion charakterisiert, deren Argumentbereich die Klasse jener Ausdrücke und deren Wert-

Eine Anzahl von Aussagen ist simultan erfüllbar genau dann, wenn es eine Interpretation über einen Bereich gibt, bei der jede dieser Aussagen wahr ist. Ein Satz ist erfüllbar dann und nur dann, wenn er simultan erfüllbar ist. Folgt Ψ logisch aus Φ, so sind Φ und die Negation von Ψ [kurz: ¬ Ψ] nicht simultan erfüllbar, und umgekehrt. Φ und Ψ sind also genau dann simultan erfüllbar, wenn ¬ Ψ nicht logisch aus Φ folgt. Sind die Sätze Φ, Ψ und Δ nicht simultan erfüllbar, so ist dies gleichbedeutend damit, daß aus zweien von ihnen die Negation des dritten logisch folgt.

Sind zwei Sätze miteinander logisch unverträglich, also nicht simultan erfüllbar, so daß es keine Interpretation gibt, bei der beide wahr sind, dann haben sr (Ψ) und sr (Φ) kein Element gemeinsam; sr (Φ) ist dann also Teil der Gesamtheit jener Interpretationen, die nicht in sr (Ψ) liegen, d. h. Teil des Komplements von sr (Ψ), und Φ impliziert daher logisch ¬ Ψ. Die logische Unverträglichkeit kann demnach als ein spezieller Fall der logischen Folgerung angesehen und graphisch folgendermaßen veranschaulicht werden:

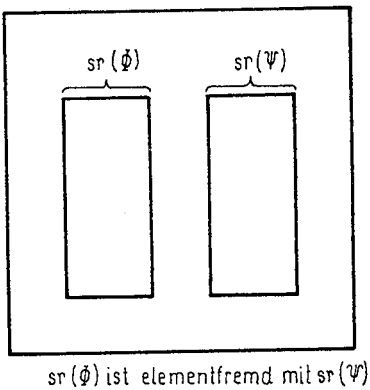


Fig. 2

bereich B die Gesamtheit der Teilklassen von B und für jedes $r > 1$ die Gesamtheit der Klassen von geordneten r-Tupeln von Dingen aus B enthält. Eine Interpretation J über B ordnet den Gegenstandsausdrücken Elemente aus B, den Eigenschaftsausdrücken Teilklassen von B und den r-stelligen Beziehungsausdrücken ($r > 1$) Klassen von geordneten r-Tupeln zu. Da die Sprache nach Voraussetzung Gegenstandsausdrücke enthält, kann der Gegenstandsbereich B einer solchen Interpretation J nicht leer sein; es ist daher überflüssig, von B noch ausdrücklich zu verlangen, daß es Elemente enthält.

Bei einem echten induktiven Schluß ist nun der Spielraum der Prämisse Φ teilweise, aber nicht ganz in dem der Konklusion (der Hypothese) Ψ enthalten, so daß es also Interpretationen gibt, bei denen zwar Φ wahr, Ψ jedoch falsch ist; in der deduktiven Logik ist ein derartiger Fall ausgeschlossen. Durch die folgende Figur wird diese partielle Implikation veranschaulicht:

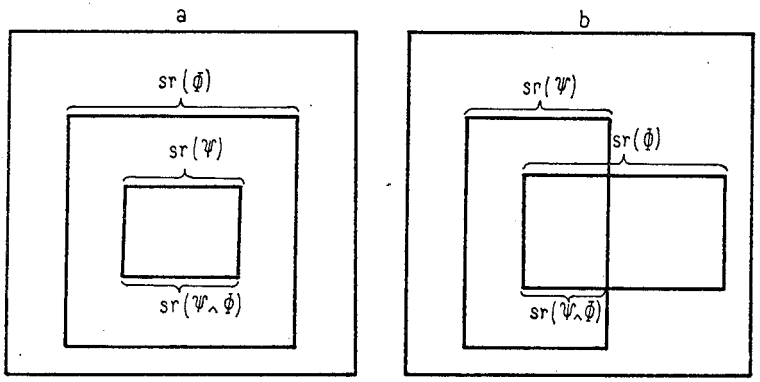


Fig. 3

Es erweist sich als zweckmäßig, die Fälle der logischen Folge und der logischen Unverträglichkeit als Grenzfälle des induktiven Schließens zu betrachten; ihnen werden dann die Zahlenwerte 1 bzw. 0 zugeordnet.

Angenommen, die Anzahl der Interpretationen ist endlich und von der Anzahl m; wenn man dann weiterhin voraussetzt, daß jede Interpretation für die Berechnung des Wahrscheinlichkeitswertes gl ichwertig ist und daß noch kein Tatsachenwissen über die Welt zur Verfügung steht, so wird man jeder Interpretation das Gewicht 1/m zuordnen (so daß die Summe aller Gewichte von Interpretationen 1 ist). In diesem Fall kann die Wahrscheinlichkeit von Ψ auf Grund von Φ identifiziert werden mit dem Verhältnis der Anzahlen der Interpretationen, die sowohl in sr (Ψ) als auch in sr (Φ) sind, zur Anzahl der Interpretationen, die in sr (Φ) sind, oder, wenn man $\Psi \wedge \Phi$ als die Bezeichnung der Konjunktion von Ψ mit Φ nimmt, mit dem Verhältnis der Anzahl von sr ($\Psi \wedge \Phi$) zur Anzahl von

sr (Φ). Es kann sich jedoch als zweckmäßig erweisen, die einzelnen Interpretationen nicht als gleichwertig anzusehen, sondern einigen ein größeres Gewicht beizumessen als anderen. Dann sind diese Gewichte für die Berechnung des Wahrscheinlichkeitswertes maßgeblich und dürfen nicht weggekürzt werden. Nennt man die Summe der gewogenen Interpretationen aus sr (Δ) das Maß von Δ, so ist die Wahrscheinlichkeit von Ψ auf Grund von Φ identisch mit dem Verhältnis des Maßes von $\Psi \wedge \Phi$ zum Maß von Φ.

Diese Begriffe sollen nun in einer einfachen Modellsprache präzisiert werden.[3] Es wird dazu angenommen, daß der Bereich B, über den die zugrundegelegte Sprache interpretiert ist, endlich ist und etwa die Anzahl n besitzt. Diese Einschränkung ist zwar vom theoretischen Standpunkt aus gravierend, für die praktische Anwendung des Wahrscheinlichkeitsbegriffs jedoch ohne größere Bedeutung; denn n kann ja jede beliebige endliche Anzahl sein, also auch etwa 10^{10}, eine Zahl, die in vielen Fällen schon als gute Approximation für ∞ gelten kann. Des weiteren wird vorausgesetzt, daß in der Sprache nur endlich viele in ihr nicht reduzierbare Eigenschaften und Beziehungen ausdrückbar sind. Auch diese Bestimmung ist für die praktische Anwendung weniger einschränkend als es zunächst scheinen mag. Soll etwa die Länge eines Körpers a gemessen werden, so muß hierzu ein Maßstab b zur Verfügung stehen. An diesem sind aber, wie lange und wie fein unterteilt er auch ist, nur endlich viele Meßintervalle markiert. Bei der Verwendung von b kann der Körper a daher nur eine der *endlich vielen* Längeneigenschaften haben. Selbstverständlich wird es uns in praktisch allen Fällen möglich sein, einen längeren oder besser gegliederten Maßstab zu schaffen, aber niemand hindert uns daran, in diesem Fall eine Sprache mit entsprechend mehr Eigenschaftsausdrücken zu verwenden.

Diese Vereinfachungen bewirken, daß die Interpretation J der Sprache in der Sprache selbst ausdrückbar ist, und zwar auf folgende Weise: Es wird zusätzlich angenommen, daß jedes Ding aus B unter J durch einen Gegenstandsausdruck der Sprache bezeichnet wird. Wird ein Ding durch mehr als einen Ausdruck bezeichnet, so wird einer davon

---

[3] Dies ist erstmals von Carnap getan worden, vgl. Carnap [9], III. und V. Kapitel, sowie Carnap-Stegmüller [1], S. 136ff.

ausgewählt, während man alle restlichen aus der Sprache entfernt.[4] Die Sprache besitzt dann so viele Ausdrücke, wie B Elemente enthält; $\alpha_1, \alpha_2, \ldots, \alpha_n$ seien diese Ausdrücke. $\Pi_1, \Pi_2, \ldots, \Pi_p$ seien die p Eigenschafts- und Beziehungsausdrücke der Sprache, die in ihr nicht mehr mittels anderer Begriffe definierbar sind. Neben diesem außerlogischen Vokabular enthält sie noch die Kopula „ε" („ist" bzw. „ist Element von"), durch die diese zu atomaren Sätzen verknüpft werden, sowie die junktorenlogischen Konstanten „¬" („nicht"), „∧" („und"), „v" („oder" im nichtausschließenden Sinn), „→" („wenn — dann") und „↔" („genau dann, wenn"), unter deren Verwendung aus atomaren Sätzen molekulare gebildet werden können.

Ein Basissatz enthält an logischen Konstanten höchstens das Negationszeichen, er ist also entweder ein atomarer Satz oder die Negation eines solchen. Ein Basissatzpaar besteht aus einem atomaren Satz und dessen Negation. Unter der Standardkonjunktion einer Anzahl von Sätzen wird die Konjunktion dieser Sätze in lexikographischer Anordnung verstanden. m sei die Anzahl der atomaren Sätze und damit auch der Basissatzpaare; dann gibt es 2^m verschiedene Mengen von Sätzen, von denen jede aus jedem Basissatzpaar genau einen Satz als Element enthält. Die Standardkonjunktion einer derartigen Menge von Sätzen ist eine Zustandsbeschreibung.

Enthält die Sprache lediglich drei Gegenstandsausdrücke und zwei Eigenschaftsbegriffe, so daß die atomaren Sätze die Form $\alpha_i \varepsilon \Pi_j$

---

[4] Diese Forderung kann umgangen werden, indem man Sätze, die besagen, daß die betreffenden Dinge miteinander identisch sind, als Axiome in der Sprache formuliert. Die Reduktion der Sprache auf ihr wesentliches Vokabular stellt keine echte Einschränkung ihrer Ausdrucksfähigkeit dar; es kann also unter der vorgegebenen Interpretation J in ihr jeder Sachverhalt beschrieben werden, der in einer Sprache mit zusätzlichen Konstanten ausdrückbar ist. Der Begriff des Sachverhaltes (bzw. des Ereignisses) wird in der Wahrscheinlichkeitstheorie im extensionalen Sinn verwendet, und zwar so, daß die Aussagen „Der Abendstern leuchtet hell" und „Der Morgenstern leuchtet hell" unter der gegebenen Interpretation der Sprache den gleichen Sachverhalt beschreiben, falls „der Morgenstern" und „der Abendstern" bei dieser Interpretation das gleiche Ding bezeichnen. Einige Philosophen verwenden den Begriff „Sachverhalt" in einer anderen Bedeutung, oder sie machen Unterschiede, wo für die Wahrscheinlichkeitstheorie keine bestehen. — Es ist übrigens für das folgende gleichgültig, ob die Gegenstandsausdrücke Konstanten oder Variablen sind. Carnap hat sich auf Konstanten festgelegt, vermutlich, weil damit leichter zu operieren ist.

haben, so sind die beiden folgende Sätze Beispiele von Zustandsbeschreibungen:

$\alpha_1 \varepsilon \Pi_1 \wedge \neg (\alpha_1 \varepsilon \Pi_2) \wedge \alpha_2 \varepsilon \Pi_1 \wedge \alpha_2 \varepsilon \Pi_2 \wedge \neg (\alpha_3 \varepsilon \Pi_1) \wedge \neg (\alpha_3 \varepsilon \Pi_2)$.
$\alpha_1 \varepsilon \Pi_1 \wedge \alpha_1 \varepsilon \Pi_2 \wedge \alpha_2 \varepsilon \Pi_1 \wedge \neg (\alpha_2 \varepsilon \Pi_2) \wedge \neg (\alpha_3 \varepsilon \Pi_1) \wedge \alpha_3 \varepsilon \Pi_2$.

Eine solche Zustandsbeschreibung ist erfüllbar und damit konsistent: Sie enthält ja keinen Satz zusammen mit seiner Negation, sondern nur entweder einen atomaren Satz und dann nicht dessen Negation oder aber seine Negation und dann nicht ihn selbst; zwei atomare Sätze sind jedoch voneinander logisch unabhängig, d. h. es ist nicht der Fall, daß einer von beiden den anderen oder dessen Negation logisch impliziert. Andererseits ist eine derartige Zustandsbeschreibung maximal-konsistent in folgendem Sinn: Wird ihr ein Basissatz, den sie nicht als Konjunktionsglied enthält, konjunktiv angefügt, so ist der dadurch entstehende Satz inkonsistent und also logisch falsch. Denn wenn sie diesen Basissatz nicht als Konjunktionsglied enthält, so kommt in ihr, je nachdem, ob er atomar oder negiert ist, dessen Negation oder der atomare Satz vor, aus dem er durch Negation hervorgegangen ist. Wird er nun jener Zustandsbeschreibung konjunktiv angefügt, so enthält der dabei entstehende Satz als Konjunktionsglieder sowohl einen atomaren Satz als auch dessen Negation, also einen Widerspruch.

Zu jeder Zustandsbeschreibung Σ gibt es wenigstens und höchstens, also genau eine Interpretation J′ über B, bei der Σ wahr ist und folgendes gilt: J′ unterscheidet sich von J nicht bezüglich der Gegenstandsausdrücke $\alpha_1, \ldots, \alpha_n$, sondern höchstens bezüglich der Eigenschafts- und Beziehungsausdrücke $\Pi_1, \ldots, \Pi_p$, d. h. sie ordnet den Gegenstandsausdrücken die gleichen Dinge zu wie J, sie ordnet jedoch den Eigenschafts- und Beziehungsausdrücken eventuell andere Begriffsumfänge (andere Teilklassen von B bzw. andere Klassen von geordneten r-Tupeln von Dingen aus B) zu. Umgekehrt ist bei jeder derartigen Interpretation J′ über B genau eine jener Zustandsbeschreibungen Σ wahr. Jede dieser Interpretationen J′ wird also eindeutig durch eine Zustandsbeschreibung Σ charakterisiert, und jedem derartigen Satz Σ entspricht eindeutig ein J′ der beschriebenen Art. Es kann deshalb der Satz „Φ ist bei einer dieser Interpretation J′ über B

wahr" ersetzt werden durch „Φ gilt bei einer Zustandsbeschreibung Σ (die jenem J' entspricht)". Σ ist jedoch ein Satz der Objektsprache, und der Begriff des Geltens bei einer Zustandsbeschreibung kann daher als ein rein syntaktischer Begriff definiert werden.
Da außerdem keine zwei Zustandsbeschreibungen simultan erfüllbar sind, während bei jeder derartigen Interpretation wenigstens eine von ihnen wahr ist, ist die Standardadjunktion aller Zustandsbeschreibungen (die Adjunktion dieser Sätze in lexikographischer Anordnung) logisch wahr. Diese Zustandsbeschreibungen sind also miteinander logisch unverträglich und in ihrer Gesamtheit logisch erschöpfend, daher ist jeder Satz der Sprache logisch äquivalent mit einer Adjunktion von Zustandsbeschreibungen (nämlich mit der Adjunktion jener, die ihn logisch implizieren), wie durch mathematische Induktion (= Rekursion) gezeigt werden kann.
Die Beschränkung auf einen endlichen Gegenstandsbereich B bewirkt weiterhin, daß die zugrundegelegte Sprache an logischen Termini nur das junktorenlogische Vokabular benötigt. Der Allsatz $\Lambda\, \xi\, \Phi\, (\xi)$ kann durch die Konjunktion $\Phi\, (\alpha_1) \wedge \Phi\, (\alpha_2) \wedge \ldots \wedge \Phi\, (\alpha_n)$ und der Existenzsatz $V\, \xi\, \Phi\, (\xi)$ durch die Adjunktion $\Phi\, (\alpha_1) \vee \Phi\, (\alpha_2) \vee \ldots \vee \Phi\, (\alpha_n)$ wiedergegeben werden.
Diesen Zustandsbeschreibungen werden nun Gewichte (oder Maße) zugeordnet. Da jede von ihnen eine Interpretation und damit eine mögliche Welt charakterisiert, muß das Gewicht größer als 0 sein; die Summe aller Gewichte normieren wir mit 1. Weitere Bedingungen werden an die Bewertung von Zustandsbeschreibungen und den durch sie beschriebenen Interpretationen vorläufig nicht gestellt. Für das folgende wird stets vorausgesetzt, daß „c" eine Funktionsvariable ist, die den Typus des Wahrscheinlichkeitsbegriffs besitzt, also für Satzpaare definiert ist und reelle Zahlen als Werte annimmt. „T" sei die Bezeichnung einer Tautologie, etwa des *tertium non datur*. Der Begriff des regulären Maßes für Zustandsbeschreibungen kann dann wie folgt definiert werden:

D 1 c ist ein reguläres Maß für die Zustandsbeschreibungen (kurz: für Zb) genau dann, wenn für alle Σ aus Zb: $c\, (\Sigma, T) > 0$, sowie $\sum_{\Sigma \in Zb} c\, (\Sigma, T) = 1$.

„Σ e Zb" ist dabei eine Abkürzung für „Σ ist eine Zustandsbeschreibung", und „Σ e Zb^{Δ}" für „Σ ist eine jener (ein Element jener Klasse von) Zustandsbeschreibungen, die Δ logisch implizieren". Zb^{T} ist also die Klasse aller Zustandsbeschreibungen und daher mit Zb identisch, und $Zb^{\neg T}$ ist die leere Klasse, die kein einziges Element enthält. Wenn „c" als adäquate Formalisierung des intuitiven Wahrscheinlichkeitsbegriffs angesehen werden kann, so ist „c (Δ, T)" zu lesen als „die Wahrscheinlichkeit von Δ auf Grund des Wissens, das nur aus der Tautologie besteht", „die Anfangswahrscheinlichkeit von Δ", „die Aprioriwahrscheinlichkeit von Δ" und als „das Maß von Δ". Ist Δ eine Zustandsbeschreibung, so ist sein Maß gerade das Gewicht, das ihr auf Grund von Aprioriüberlegungen beigemessen wird.

Das Maß eines beliebigen Satzes Δ ist dann die *Summe* der Maße (Gewichte), die jene Zustandsbeschreibungen besitzen, bei denen Δ gilt, mit deren Standardadjunktion Δ also logisch äquivalent ist; dies kommt durch die folgende Definition zum Ausdruck:

D 2 c ist ein reguläres Maß genau dann, wenn für jeden Satz Δ gilt: $c(\Delta, T) = \sum_{\Sigma \, e \, Zb^{\Delta}} c(\Sigma, T)$, wobei c ein reguläres Maß für die Zustandsbeschreibungen ist.

Es sei Φ wahr, so daß die tatsächliche Welt durch einen Satz aus Zb^{Φ} beschrieben wird; zu ermitteln ist, in welchem Grad Ψ durch Φ impliziert wird. Dazu genügt es, das Verhältnis der Summe der Gewichte jener Zustandsbeschreibungen aus Zb^{Φ}, die auch in Zb^{Ψ} sind, zur Summe der Gewichte aller Zustandsbeschreibungen aus Zb^{Φ} festzustellen, oder, was per definit onem damit identisch ist, das Verhältnis der Summe der Gewichte der Zustandsbeschreibungen aus $Zb^{\Psi \wedge \Phi}$ zur Summe der Gewichte der Zustand beschreibungen aus Zb^{Φ}. Werden alle Zustandsbeschreibungen mit gleichen Gewichten bewertet, so ist dieses Verhältnis identisch mit dem Verhältnis der Fläche von sr ($\Psi \wedge \Phi$) zu sr (Φ) in Figur 3a und b. Werden die Zustandsbeschreibungen nicht gleich bewertet, so hängt das Verhältnis außerdem noch von der Verteilung der Gewichte ab. Dies führt zu der folgenden Definition:

D 3 c ist eine reguläre Bestätigungsmethode genau dann, wenn für jeden Satz Ψ und für jeden erfüllbaren[5] Satz Φ gilt:

$$c(\Psi, \Phi) = \frac{c(\Psi \wedge \Phi, T)}{c(\Phi, T)}, \text{ wobei c ein reguläres Maß ist.}$$

Angenommen, c ist eine reguläre Bestätigungsmethode und Φ ist ein erfüllbarer Satz, der Ψ logisch impliziert; dann ist jede Zustandsbeschreibung, bei der Φ gilt, eine Zustandsbeschreibung, bei der Ψ gilt. Es ist also Zb^{Φ} in Zb^{Ψ} enthalten und Zb^{Φ} mit $Zb^{\Psi \wedge \Phi}$ identisch. Dann ist $c(\Psi, \Phi) = 1$, gleichgültig, wie die Zustandsbeschreibungen bewertet werden. Angenommen hingegen, Φ ist mit Ψ logisch unverträglich; dann ist $Zb^{\Psi \wedge \Phi}$ leer, so daß $c(\Psi \wedge \Phi, T)$ und damit auch $c(\Psi, \Phi)$ mit 0 identisch sind. Die Wahrscheinlichkeit liegt also zwischen der logischen Folgerung und der logischen Unverträglichkeit.

Soll Ψ aus Φ logisch gefolgert we den, so darf zur Ableitung von Ψ kein anderes Wissen als Φ benützt werden; denn sonst dürfte nicht mehr von einer rein logischen Ableitung gesprochen werden. Das gleiche gilt von jeder der regulären Bestätigungsmethoden: Steht neben Φ noch ein zusätzliches für Ψ relevantes Wissen Γ zur Verfügung, so muß dies bei Berechnung des c-Wertes benützt werden, will man nicht falsche Ergebnisse erhalten. Während allerdings in der deduktiven Logik Ψ aus $\Phi \wedge \Gamma$ folgt, falls Ψ aus Φ folgt, kann in der induktiven Logik der Fall eintreten, daß Ψ im Hinblick auf Φ einen anderen Wahrscheinlichkeitswert erhält als im Hinblick auf $\Phi \wedge \Gamma$.

Der Wahrscheinl chkeitswert bezieht sich nicht auf die Hypothese Ψ, sondern auf die partielle Folgerungsbeziehung zwischen Φ und Ψ. Auch in der deduktiven Logik bezieht sich ja in den Sätzen „aus Φ folgt logisch Ψ" und „aus Φ folgt notwendig Ψ" das „logisch" bzw.

---

[5] Während der Begriff der logischen Folgerung auch für kontradiktorische Prämissen definiert ist, haben reguläre Bestätigungsmethoden, die ja das induktive Schließen repräsentieren sollen, für solche Prämissen zunächst keinen Zahlenwert. Für die Anwendung der induktiven Logik hat dies keine weiteren Konsequenzen, da ja in keinem korrekten Argument von einem kontradiktorischen Wissen ausgegangen wird. Über die Probleme, die bei der Definition des Wahrscheinlichkeitsbegriffs auch für logisch falsche Prämissen auftreten, vgl. Carnap [9], S. 295f.

„notwendig" nicht auf „Ψ", sondern auf „folgt". Gilt also Φ, so darf nur auf Ψ, nicht aber auf notwendig-Ψ geschlossen werden.
Die Schwäche des Begriffs der regulären Bestätigungsmethode und der ihm zugrundeliegenden Begriffe des regulären Maßes und des regulären Maßes für Zustandsbeschreibungen liegt in seiner Allgemeinheit: Sie legen nicht *einen*, sondern e ne nicht überschaubare Gesamtheit von Wahrscheinlichkeitsbegriffen fest, je nachdem, welches Aprioriqewicht man den Zustandsbeschreibungen beimißt. Von diesen Bewertungen wird nur verlangt, daß sie stets positiv sind und ihre Summe 1 ergibt. Solange jedoch nicht festgelegt ist, wie die Verteilung der Zahlenwerte vorzunehmen ist, kann, abgesehen von einigen wenigen trivialen Fällen, kein Wahrscheinlichkeitswert ermittelt werden. Diese Schwäche teilt der Begriff der regulären Bestätigungsmethode auch mit den noch zu charakterisierenden Begriffen der regulären Wahrscheinlichkeitsfunktion und der strengen Kohärenz.
D 1 kann dahingehend abgeändert werden, daß „c (Σ, T) > 0" durch „c (Σ, T) ⩾ 0" ersetzt wird. Es ist dann in allen drei Definitionen der Begriff „regulär" zu streichen und in D 3 an Stelle der Erfüllbarkeit von Φ zu fordern, daß c (Φ, T) ≠ 0. Denn falls für wenigstens ein Σ aus Zb gilt, daß c (Σ, T) = 0, so kann der Wahrscheinlichkeitsbegriff nicht mehr als ein Gradmesser der Folgerung bezeichnet werden, da diese Zustandsbeschreibung Σ ja eine logisch mögliche Welt beschreibt. Einige Wahrscheinlichkeitstheoretiker, wie etwa Reichenbach oder Waismann, haben zwar versucht, einen nicht regulären Begriff der partiellen Implikation zu charakterisieren, doch wird dieser im folgenden nicht verwendet; die Möglichkeit seiner Einführung sei deshalb nur am Rande erwähnt.
Nicht immer wird der Fall gegeben sein, daß man über die Entitäten, die den Eigenschafts- und Beziehungsausdrücken durch die Interpretation zugeordnet sind, und über ihr wechselseitiges Verhältnis nichts weiß. Wenn der Gegenstandsbereich etwa aus einer endlichen Anzahl n von Würfen mit einem (regelmäßigen oder unregelmäßigen) Würfel besteht, so wird man den Ausdrücken $\Pi_1, \ldots, \Pi_6$ die Eigenschaften *Ziffer* „1" *liegt oben*, ..., *Ziffer* „6" *liegt oben* (bzw. die entsprechenden Begriffsumfänge) zuordnen. Dann ist jedoch bereits bekannt, daß jeder Wurf wenigstens und höchstens, also genau eine der angegebenen Eigenschaften hat (bzw. Element genau einer der

entsprechenden Klassen ist). Bei der Berechnung eines Wahrscheinlichkeitswertes c (Ψ, Φ) kann dann einer der drei folgenden Wege beschritten werden:
(1) Alle Zustandsbeschreibungen (und damit alle Interpretationen), die besagen, daß ein Wurf keine dieser Eigenschaften oder aber mehrere von ihnen hat, erhalten das Maß (die Aprioriwahrscheinlichkeit) 0, d. h. für solche Σ wird festgesetzt: c (Σ, T) = 0. Die c-Werte der übrigen Zustandsbeschreibungen werden so geändert, daß das Verhältnis der Zahlenwerte gleichbleibt und ihre Summe wiederum 1 ergibt.
(2) Es wird ein Satz Γ der Sprache, der besagt, daß jedes Ding wenigstens eine der angegebenen Eigenschaften hat und daß kein Ding zwei dieser Eigenschaften besitzt, als Bedeutungspostulat aufgestellt (als Postulat also, das die Intension der fraglichen Begriffe teilweise oder ganz beschreibt). Zustandsbeschreibungen, die diesem Axiom Γ widersprechen, erhalten das Maß 0; die übrigen Maße von Zustandsbeschreibungen werden wie in (1) geändert.
(3) Die Zustandsbeschreibungen behalten zwar durchgehend ihre positiven Werte (auch jene, die Γ widersprechen), es wird jedoch Γ konjunktiv jeder Prämisse Φ eines induktiven Argumentes hinzugefügt, d. h. es wird stets der Wert c (Ψ, $\Phi \wedge \Gamma$) berechnet.
Die Gleichwertigkeit der Methoden (1) und (2) liegt auf der Hand. Daß beide auch mit (3) gleichwertig sind, sieht man sofort, wenn man bedenkt, daß durch D 3 die bedingte Wahrscheinlichkeit c (Ψ, $\Phi \wedge \Gamma$) auf die Aprioriwahrscheinlichkeiten c ($\Psi \wedge \Phi \wedge \Gamma$, T) und c ($\Phi \wedge \Gamma$, T) zurückgeführt wird; zur Berechnung beider werden nach D 2 jedoch nur die Aprioriwahrscheinlichkeiten von Zustandsbeschreibungen herangezogen, die Γ nicht widersprechen, und daher ist das Verhältnis der Zahlenwerte von c ($\Psi \wedge \Phi \wedge \Gamma$, T) und c ($\Phi \wedge \Gamma$, T) hier das gleiche wie bei (1) und (2).

2. *Reguläre Wahrscheinlichkeitsfunktionen*

Ein Axiomensystem ist eine Klasse von Sätzen, durch die den Entitäten, von denen diese Sätze handeln, gewisse Eigenschaften zuge-

schrieben werden.[6] Damit die in den Axiomen vorkommenden Grundausdrücke solche Eigenschaften und Beziehungen bezeichnen können, muß die Sprache, in der diese Sätze formuliert worden sind, über einem gewissen Bereich von Gegenständen interpretiert worden sein. Eine derartige Interpretation, unter der alle Axiome wahr sind, nennt man ein Modell des Systems. Nun hat ein Axiomensystem, wie etwa jenes für das Rechnen mit Wahrscheinlichkeiten, im allgemeinen viele voneinander versch edene Modelle, die lediglich gewisse strukturelle Merkmale gemeinsam haben. Diese Struktur wird also durch das Axiomensystem definitorisch charakterisiert, nicht jedoch die Grundbegriffe und die Entitäten, die durch sie bezeichnet werden: Sie sind variabel je nach Interpretation; die Ausdrücke, die dabei als Grundbegriffe benützt werden, können daher als Variablen aufgefaßt werden.[7]

Es wird durch die Wahrscheinlichkeitsaxiome einschließlich des sogenannten Regularitätsaxioms (der Bedingung (6) in der folgenden Definition D 4) also nicht ein Funktionsausdruck „c", sondern der Begriff „c ist eine reguläre Wahrscheinlichkeitsfunktion" definiert; da durch sie nicht *eine* Wahrscheinlichkeitsfunktion, sondern, wie noch gezeigt wird, eine nicht überschaubare Menge solcher Funktionen charakterisiert wird, darf diese Formulierung nicht durch „c ist *die* reguläre Wahrscheinlichkeitsfunktion" ersetzt werden. Die Regularität, die im folgenden durch Bedingung (6) zum Ausdruck kommt, garantiert, daß in endlichen Bereichen der Wahrscheinlichkeitswert 1 stets als logische Implikation und der Wahrscheinlichkeitswert 0 als logische Unverträglichkeit gedeutet werden kann. c wird dabei als eine Funktion angesehen, die für Satzpaare (mit erfüllbarem Zweitglied) und nicht, wie üblich, für Paare von Sachverhalten bzw. Ereignissen definiert ist. Denn da bei einer philosophischen Grundlegung der Ereignisbegriff definiert werden muß und dies nicht ohne Verwendung des Ausdrucks „Satz" möglich ist, definiert man den Wahrscheinlichkeitsbegriff zweckmäßigerweise gleich für Sätze statt

[6] Einen kurzen Abriß der Axiomatik findet man in Essler [1], S. 180—191.

[7] David Hilbert hat erstmals die Behauptung aufgestellt, daß die Grundbegriffe eines Axiomensystems durch dieses System implizit definiert werden. Dies ist irreführend ausgedrückt: Definiert wird durch ein Axiomensystem ein Begriff, der eine gewisse Struktur bezeichnet, während die Grundbegriffe, wie schon Hilbert selbst gesehen hat, Variablen sind.

für Sachverhalte oder Ereignisse.[8] Die Definition des Begriffs der regulären Wahrscheinlichkeitsfunktion lautet dann:

D 4 c ist eine reguläre Wahrscheinlichkeitsfunktion genau dann, wenn

(1) $c(\Psi, \Phi) \geqslant 0$ für jeden Satz Ψ und jeden erfüllbaren Satz Φ,

(2) $c(\Phi, \Phi) = 1$ für jeden erfüllbaren Satz Φ,

(3) $c(\Psi, \Phi) = c(\Delta, \Gamma)$ für jeden Satz Ψ und jeden mit Ψ logisch äquivalenten Satz Δ, für jeden erfüllbaren Satz Φ und jeden mit Φ logisch äquivalenten Satz Γ,

(4) $c(\Psi \vee \Delta, \Phi) = c(\Psi, \Phi) + c(\Delta, \Phi)$ für alle Sätze Ψ und Δ und jeden erfüllbaren Satz Φ, wobei Ψ, Δ und Φ nicht simultan erfüllbar sind,

(5) $c(\Psi \wedge \Delta, \Phi) = c(\Psi, \Phi) \cdot c(\Delta, \Psi \wedge \Phi)$ für alle Sätze Ψ, Δ und Φ, wobei $\Psi \wedge \Phi$ erfüllbar ist, und

(6) $c(\Psi, \Phi) \neq 1$ für jede nicht logisch aus dem Satz Φ folgende Aussage Ψ.

Diese sechs Bedingungen entsprechen den üblichen Axiomen der Wahrscheinlichkeitsrechnung. Lediglich (3) ist hier neu: Der Wahrscheinlichkeitswert ändert sich nicht, wenn die Prämisse durch eine andere, mit ihr logisch äquivalente, ersetzt wird, und er bleibt gleich, wenn an Stelle der Hypothese Ψ ein mit ihr logisch äquivalenter Satz genommen wird. Da logisch äquivalente Sätze den gleichen Sachverhalt beschreiben und der Wahrscheinlichkeitsbegriff üblicherweise für Sachverhalte (Ereignisse) definiert wird, ist das Axiom (3) bei dieser Deutung eine logische Trivialität und braucht nicht erwähnt zu werden.

Die Bedingung (1) schließt aus, daß eine Wahrscheinlichkeitsfunktion negative Zahlen als Werte haben kann. Durch (2) wird ausgesagt, daß c den Wert 1 annimmt, wenn die Hypothese durch das Erfahrungsdatum im höchsten Grad bestätigt wird, nämlich, wenn sie gerade dieses Datum ist. Daraus ergibt sich mit (3) und (5), daß die Wahrscheinlichkeit von Ψ bezüglich Φ 1 ist, wenn Ψ aus dem erfüll-

[8] Für Sprachen, die über unendlichen Gegenstandsbereichen interpretiert sind, kann es sich als zweckmäßig erweisen, die Wahrscheinlichkeit als Funktion von Satzmengen oder von Sätzen mit (potentiell) unendlicher Länge einzuführen.

baren Satz Φ logisch folgt. Bedingung (6) fügt hinzu, daß c auch *nur* in diesem Fall den Wert 1 annimmt, daß der Wert also ungleich 1 ist, wenn Ψ nicht logisch aus Φ folgt (mit (4): daß er dann kleiner als 1 ist). (6) gilt ohne Einschränkungen übrigens nur in endlichen Bereichen. Denn man wird sinnvollerweise die Funktion c für unendliche Bereiche als Limes der Wahrscheinlichkeitsfunktionen für alle endlichen Bereiche charakterisieren. Nun kann jedoch der Fall eintreten, daß Ψ nicht logisch aus Φ folgt, und der c-Wert für alle endlichen Bereiche B kleiner als 1 ist, jedoch mit wachsendem B immer größer wird und sich immer mehr der 1 nähert; man erhält dann als Grenzwert 1, obwohl dieser Wert in keinem endlichen Fall auftritt. Wenn also nicht vorausgesetzt wird, wie dies hier geschehen ist, daß B endlich ist, und wenn die Sprache außerdem noch die Quantoren „alle" und „es gibt" enthält, dann muß vorausgesetzt werden, daß Ψ nur von endlich vielen atomaren Ereignissen spricht, also keine Quantoren enthält.

Die Bedingung (4), genannt das spezielle Additionsprinzip, kann folgendermaßen verdeutlicht werden: Angenommen, der erfüllbare Satz Φ beschreibt mein Wissen von der gegenwärtigen innenpolitischen Situation der USA; Ψ sei die Behauptung, daß der Politiker a von der Partei F zur nächsten Präsidentschaftswahl nominiert wird, und Δ sei die Aussage, daß a für diese Wahl von der Partei G als Präsidentschaftskandidat aufgestellt wird. Aus meinem Wissen möge hervorgehen, daß er nicht gleichzeitig von beiden Parteien nominiert werden kann, so daß also die Wahrheit von Φ die Falschheit von wenigstens einem der Sätze Ψ und Δ nach sich zieht; dies ist gleichbedeutend damit, daß Φ, Ψ und Δ nicht simultan erfüllbar sind. Dann ist auf der Grundlage meines Wissens die Wahrscheinlichkeit dafür, daß a von F oder von G nominiert wird, gleich der Wahrscheinlichkeit dafür, daß er von F aufgestellt wird, plus der Wahrscheinlichkeit dafür, daß er von G aufgestellt wird.

Das folgende Beispiel verdeutlicht (5), das Allgemeine Multiplikationsprinzip der Wahrscheinlichkeitslehre. Φ sei eine Beschreibung meines Wissens von der innenpolitischen Situation der USA des vergangenen Jahrhunderts. Ψ möge besagen, daß der Politiker b von der Partei F für eine bestimmte Präsidentschaftswahl nominiert worden ist, und Δ, daß b diese Wahl gewonnen hat. Ψ und Φ seien simultan

erfüllbar; es sei also auf Grund meines Wissens Φ nicht unmöglich, daß damals b von F nominiert worden ist. Dann ist die Wahrscheinlichkeit dafür, daß b nominiert und vom Volk gewählt worden ist, bezüglich meines Wissens Φ, identisch mit dem Produkt der Wahrscheinlichkeit bezüglich Φ dafür, daß b von F nominiert worden ist, multipliziert mit der Wahrscheinlichkeit auf Grund von Φ und des zusätzlichen Wissens, daß er von F nominiert worden ist, dafür, daß er diese Wahl gewonnen hat.

In (5) darf die Gleichung „$c(\Psi \wedge \Delta, \Phi) = c(\Psi, \Phi) \cdot c(\Delta, \Psi \wedge \Phi)$" nicht durch „$c(\Psi \wedge \Delta, \Phi) = c(\Psi, \Phi) \cdot c(\Delta, \Phi)$" ersetzt werden. Denn die Voraussetzung von (5) schließt z. B. nicht aus, daß die Wahrscheinlichkeit von Ψ auf Grund von Φ und die Wahrscheinlichkeit von Δ auf Grund von Φ positiv ist, die Sätze Φ, Ψ und Δ jedoch nicht simultan erfüllbar sind und deshalb wegen (1), (2) und (4) die Wahrscheinlichkeit von $\Psi \wedge \Delta$ auf Grund von Φ mit 0 identisch ist. Würde man (5) also im obigen Sinn abändern, so wäre der Begriff der regulären Wahrscheinlichkeitsfunktion nicht adäquat charakterisiert. Ist $\Psi \wedge \Phi$ erfüllbar, so ist $c(\Psi \wedge \Delta, \Phi) = c(\Psi, \Phi) \cdot c(\Delta, \Phi)$ genau dann, wenn $c(\Delta, \Phi) = c(\Delta, \Psi \wedge \Phi)$, d. h. wenn Ψ für Δ bezüglich Φ irrelevant ist. Dieser aus (5) unmittelbar folgende Zusammenhang wird „Spezielles Multiplikationsprinzip" genannt.

Reguläre Wahrscheinlichkeitsfunktionen haben eine Reihe von bemerkenswerten Eigenschaften, die durch den folgenden Lehrsatz zum Ausdruck kommen:

Theorem 1

Für jede reguläre Wahrscheinlichkeitsfunktion c gilt:

(a) $c(\Psi, \Phi) = 1$ für jeden erfüllbaren Satz Φ und jeden aus Φ logisch folgenden Satz Ψ,

(b) $c(T, \Phi) = 1$ für jeden erfüllbaren Satz Φ,

(c) $c(\Psi, \Phi) + c(\neg \Psi, \Phi) = 1$ für jeden Satz Ψ und jeden erfüllbaren Satz Φ,

(d) $c(\neg \Psi, \Phi) = 1 - c(\Psi, \Phi)$ für jeden Satz Ψ und jeden erfüllbaren Satz Φ,

(e) $c(\Psi, \Phi) \leqslant 1$ für jeden Satz Ψ und jeden erfüllbaren Satz Φ,

(f) $c(\Psi, \Phi) = 0$ für jeden erfüllbaren Satz Φ und jeden nicht mit Φ simultan erfüllbaren Satz Ψ,

(g) $c(\neg T, \Phi) = 0$ für jeden erfüllbaren Satz Φ,
(h) $c(\neg \Phi, \Phi) = 0$ für jeden erfüllbaren Satz Φ,
(i) $c(\Psi, \Phi) = c(\Delta \wedge \Psi, \Phi) + c(\neg \Delta \wedge \Psi, \Phi)$ für alle Sätze Ψ und Δ und jeden erfüllbaren Satz Φ, und
(j) $c(\Psi \vee \Delta, \Phi) = c(\Psi, \Phi) + c(\Delta, \Phi) - c(\Psi \wedge \Delta, \Phi)$ für alle Sätze Ψ und Δ und jeden erfüllbaren Satz Φ.
(k) $c(\Psi, \Phi) \leqslant c(\Delta, \Phi)$ für beliebige Sätze Φ und Δ, so daß $\Delta \wedge \Phi$ erfüllbar ist, und für jeden Satz Ψ, aus dem Δ logisch folgt.

Beweis

c sei eine reguläre Wahrscheinlichkeitsfunktion.
(a) Φ sei ein erfüllbarer Satz und Ψ eine logisch aus Φ folgende Aussage. Dann ist Φ logisch äquivalent mit $\Phi \wedge \Psi$. Wegen (5) gilt: $c(\Phi \wedge \Psi, \Phi) = c(\Phi, \Phi) \cdot c(\Psi, \Phi \wedge \Phi)$, und wegen (3) außerdem: $c(\Phi \wedge \Psi, \Phi) = c(\Phi, \Phi)$, sowie: $c(\Psi, \Phi \wedge \Phi) = c(\Psi, \Phi)$. Aus diesen drei Gleichungen folgt mit (2): $1 = 1 \cdot c(\Psi, \Phi)$ und damit: $c(\Psi, \Phi) = 1$.
(b) T ist logisch wahr und folgt also logisch aus dem erfüllbaren Satz Φ. Wegen (a) gilt deshalb: $c(T, \Phi) = 1$.
(c) T ist logisch äquivalent mit $\Psi \vee \neg \Psi$; wegen (3) gilt daher: $c(T, \Phi) = c(\Psi \vee \neg \Psi, \Phi)$. Nun sind Φ, Ψ und $\neg \Psi$ nicht simultan erfüllbar; wegen (4) gilt daher: $c(\Psi \vee \neg \Psi, \Phi) = c(\Psi, \Phi) + c(\neg \Psi, \Phi)$. Mit (b) folgt daraus: $c(\Psi, \Phi) + c(\neg \Psi, \Phi) = 1$.
(d) folgt aus (c) durch mathematische Umformung.
(e) Bedingung (1) impliziert: $c(\neg \Psi, \Phi) \geqslant 0$; also gilt wegen (d): $c(\Psi, \Phi) \leqslant 1$.
(f) Φ sei ein erfüllbarer Satz und Ψ ein Satz, wobei Φ und Ψ nicht simultan erfüllbar sein mögen; es folgt also $\neg \Psi$ logisch aus Φ. Wegen (a) gilt dann: $c(\neg \Psi, \Phi) = 1$, und wegen (d): $c(\Psi, \Phi) = 0$.
(g) $\neg T$ ist nicht erfüllbar, also sind $\neg T$ und der erfüllbare Satz Φ nicht simultan erfüllbar. Aus (f) folgt dann: $c(\neg T, \Phi) = 0$.
(h) Φ und $\neg \Phi$ sind nicht simultan erfüllbar; daher gilt wegen (f): $c(\neg \Phi, \Phi) = 0$.
(i) $\Delta \wedge \Psi$, $\neg \Delta \wedge \Psi$ und der erfüllbare Satz Φ sind nicht simultan erfüllbar. Aus (4) ergibt sich dann:
$c((\Delta \wedge \Psi) \vee (\neg \Delta \wedge \Psi), \Phi) = c(\Delta \wedge \Psi, \Phi) + c(\neg \Delta \wedge \Psi, \Phi)$.

Da Ψ logisch äquivalent mit $(\Delta \wedge \Psi) \vee (\neg \Delta \wedge \Psi)$ ist, gilt wegen (3):
$c(\Psi, \Phi) = c((\Delta \wedge \Psi) \vee (\neg \Delta \wedge \Psi), \Phi)$, und daher auch:
$c(\Psi, \Phi) = c(\Delta \wedge \Psi, \Phi) + c(\neg \Delta \wedge \Psi, \Phi)$.
(j) Aus (i) folgt durch Spezialisierung:
$c(\Psi \vee \Delta, \Phi) = c(\Delta \wedge (\Psi \vee \Delta), \Phi) + c(\neg \Delta \wedge (\Psi \vee \Delta), \Phi)$.
Nun ist Δ logisch äquivalent mit $\Delta \wedge (\Psi \vee \Delta)$; also gilt mit (3):
$c(\Delta \wedge (\Psi \vee \Delta), \Phi) = c(\Delta, \Phi)$. Da ferner $\neg \Delta \wedge \Psi$ logisch äquivalent mit $\neg \Delta \wedge (\Psi \vee \Delta)$ ist, gilt wegen (3) auch:
$c(\neg \Delta \wedge (\Psi \vee \Delta), \Phi) = c(\neg \Delta \wedge \Psi, \Phi)$.
Aus diesen Gleichungen folgt:
$c(\Psi \vee \Delta, \Phi) = c(\Delta, \Phi) + c(\neg \Delta \wedge \Psi, \Phi)$.
Außerdem folgt aus (i):
$c(\neg \Delta \wedge \Psi, \Phi) = c(\Psi, \Phi) - c(\Delta \wedge \Psi, \Phi)$.
Nun gilt wegen (3): $c(\Delta \wedge \Psi, \Phi) = c(\Psi \wedge \Delta, \Phi)$;
deshalb erhält man schließlich:
$c(\Psi \vee \Delta, \Phi) = c(\Psi, \Phi) + c(\Delta, \Phi) - c(\Psi \wedge \Delta, \Phi)$.
(Allgemeines Additionsprinzip.)
(k) $\Delta \wedge \Phi$ sei erfüllbar, und Δ folge logisch aus Ψ; dann ist $\Delta \wedge \Psi$ logisch äquivalent mit Ψ, und es gilt mit (3): $c(\Psi, \Phi) = c(\Delta \wedge \Psi, \Phi)$.
Mit (5) erhält man: $c(\Delta \wedge \Psi, \Phi) = c(\Delta, \Phi) \cdot c(\Psi, \Delta \wedge \Phi)$. Beide Gleichungen implizieren: $c(\Psi, \Phi) = c(\Delta, \Phi) \cdot c(\Psi, \Delta \wedge \Phi)$.
Nach (e) ist nun $c(\Psi, \Delta \wedge \Phi) \leqslant 1$, also gilt: $c(\Psi, \Phi) \leqslant c(\Delta, \Phi)$.

Alle Teilbehauptungen dieses Lehrsatzes sind ohne Benützung der Bedingung (6) bewiesen worden. Diese ist jedoch notwendig zum Nachweis der Gültigkeit der Teilsätze (a) und (b) des folgenden Theorems:

Theorem 2

Für jede reguläre Wahrscheinlichkeitsfunktion c gilt:
(a) Für alle Sätze Φ und Ψ:
- $c(\Psi, \Phi) = 1$ genau dann, wenn Φ erfüllbar ist und Ψ logisch aus Φ folgt,
- $c(\Psi, \Phi) < 1$ genau dann, wenn $\neg \Psi$ und Φ simultan erfüllbar sind,

$c(\Psi, \Phi) = 0$ genau dann, wenn Φ erfüllbar ist und $\neg \Psi$ aus Φ logisch folgt,

$c(\Psi, \Phi) > 0$ genau dann, wenn Ψ und Φ simultan erfüllbar sind.

(b) Für jeden Satz Ψ:

$c(\Psi, T) = 1$ genau dann, wenn Ψ logisch wahr ist,

$c(\Psi, T) < 1$ genau dann, wenn $\neg \Psi$ erfüllbar ist,

$c(\Psi, T) = 0$ genau dann, wenn $\neg \Psi$ logisch wahr ist, und

$c(\Psi, T) > 0$ genau dann, wenn Ψ erfüllbar ist.

(c) Für alle Sätze Ψ, Δ und Φ:

$c(\Psi \wedge \Delta, \Phi) = c(\Psi, \Phi) \cdot c(\Delta, \Psi \wedge \Phi)$, falls Ψ und Φ simultan erfüllbar sind,

$c(\Psi \wedge \Delta, \Phi) = c(\Delta, \Phi) \cdot c(\Psi, \Delta \wedge \Phi)$, falls Δ und Φ simultan erfüllbar sind, und

$c(\Psi, \Phi) \cdot c(\Delta, \Psi \wedge \Phi) = c(\Delta, \Phi) \cdot c(\Psi, \Delta \wedge \Phi)$, falls Δ, Ψ und Φ simultan erfüllbar sind.

(d) Für alle Sätze Ψ und Φ:

$c(\Psi \wedge \Phi, T) = c(\Phi, T) \cdot c(\Psi, \Phi)$, falls Φ erfüllbar ist,

$c(\Psi, \Phi) = \dfrac{c(\Psi \wedge \Phi, T)}{c(\Phi, T)}$, falls Φ erfüllbar ist.

Die einzelnen Teilsätze dieses Lehrsatzes sind leicht zu beweisen. Wichtig ist der Teil (d), der bedingte Wahrscheinlichkeit und Aprioriwahrscheinlichkeit in ein Verhältnis bringt und zeigt, wie man jene für erfüllbare Φ definieren kann, wenn man diese kennt.
Im folgenden Lehrsatz ist vor allem die Teilbehauptung (b) wichtig; sie wird üblicherweise „Theorem von Bayes" genannt (der von Bayes 1763 bewiesene Lehrsatz ist ein Spezialfall davon).

Theorem 3

c sei eine reguläre Wahrscheinlichkeitsfunktion. Ψ und Φ seien simultan erfüllbare Sätze, und $\Delta_1, \ldots, \Delta_m$ seien Aussagen (mit $m > 1$), wobei $\Phi \wedge \Psi \models \Delta_1 \vee \ldots \vee \Delta_m$ und die Sätze $\Phi \wedge \Psi \wedge \Delta_1, \ldots, \Phi \wedge \Psi \wedge \Delta_m$ paarweise unverträglich sind (d. h. keine zwei von ihnen simultan erfüllbar sind); Δ_i sei einer der Sätze $\Delta_1, \ldots, \Delta_m$. Dann gilt:

$$\text{(a)} \qquad c(\Psi, \Phi) = \sum_{j=1}^{m} c(\Psi \wedge \Delta_j, \Phi)$$

$$= \sum_{j=1}^{m} [c(\Psi, \Delta_j \wedge \Phi) \cdot c(\Delta_j, \Phi)]$$

$$\text{(b)} \quad c(\Delta_i, \Phi \wedge \Gamma) = \frac{c(\Gamma \wedge \Delta_i, \Phi)}{\sum_{j=1}^{m} c(\Gamma \wedge \Delta_j, \Phi)}$$

$$= \frac{c(\Gamma, \Delta_i \wedge \Phi) \cdot c(\Delta_i, \Phi)}{\sum_{j=1}^{m} [c(\Gamma, \Delta_j \wedge \Phi) \cdot c(\Delta_j, \Phi)]}$$

Beweis

Die Voraussetzungen seien erfüllt.
(a) Dann gilt: $\Phi \models \Psi \leftrightarrow \Psi \wedge (\Delta_1 \vee \ldots \vee \Delta_m)$, und $\models \Psi \wedge \Phi \leftrightarrow \Psi \wedge (\Delta_1 \vee \ldots \vee \Delta_m) \wedge \Phi$, und somit wegen Bedingung (3): $c(\Psi \wedge \Phi, T) = c(\Psi \wedge (\Delta_1 \vee \ldots \vee \Delta_m) \wedge \Phi, T)$. Wegen: $c(\Phi, T) = c(\Phi, T)$ folgt daraus mit Theorem 2 (d): $c(\Psi, \Phi) = c(\Psi \wedge (\Delta_1 \vee \ldots \vee \Delta_m), \Phi)$. Da $\Psi \wedge (\Delta_1 \vee \ldots \vee \Delta_m)$ logisch äquivalent ist mit $(\Psi \wedge \Delta_1) \vee \ldots \vee (\Psi \wedge \Delta_m)$, gilt mit (3) weiterhin:
$c(\Psi \wedge (\Delta_1 \vee \ldots \vee \Delta_m), \Phi) = c((\Psi \wedge \Delta_1) \vee \ldots \vee (\Psi \wedge \Delta_m), \Phi)$.
Auf Grund der Voraussetzung gilt von keinen zwei dieser Adjunktionsglieder, daß sie zusammen mit Φ simultan erfüllbar sind; folglich erhält man mit m—1maliger Anwendung der Bedingung (4):
$c((\Psi \wedge \Delta_1) \vee \ldots \vee (\Psi \wedge \Delta_m), \Phi) = c(\Psi \wedge \Delta_1, \Phi) + \ldots + c(\Psi \wedge \Delta_m, \Phi)$
$= \sum_{j=1}^{m} c(\Psi \wedge \Delta_j, \Phi)$ und daher insgesamt: $c(\Psi, \Phi) = \sum_{j=1}^{m} c(\Psi \wedge \Delta_j, \Phi)$.
Mit Theorem 2 (c) gilt für jeden dieser Summanden:
$c(\Psi \wedge \Delta_j, \Phi) = c(\Psi, \Delta_j \wedge \Phi) \cdot c(\Delta_j, \Phi)$ und damit insgesamt:
$c(\Psi, \Phi) = \sum_{j=1}^{m} [c(\Psi, \Delta_j \wedge \Phi) \cdot c(\Delta_j, \Phi)]$.
(b) Da nach Voraussetzung Ψ und Φ simultan erfüllbar sind und $\Phi \wedge \Psi \models \Delta_1 \vee \ldots \vee \Delta_m$, sind auch Ψ, Φ und $\Delta_1 \vee \ldots \vee \Delta_m$ simultan erfüllbar. Mit Theorem 2 (c) gilt daher:

$$c(\Delta_i, \Phi \wedge \Gamma) = \frac{c(\Gamma, \Delta_i \wedge \Phi) \cdot c(\Delta_i, \Phi)}{c(\Gamma, \Phi)}$$

Daraus erhält man einerseits durch Ersetzung des Zählers nach Theorem 2 (c) und des Nenners nach Theorem 3 (a):

$$c(\Delta_i, \Phi \wedge \Gamma) = \frac{c(\Gamma \wedge \Delta_i, \Phi)}{\sum_{j=1}^{m} c(\Gamma \wedge \Delta_j, \Phi)}$$

und andererseits allein durch Ersetzung des Nenners nach der anderen Gleichung von Theorem 3 (a):

$$c(\Delta_i, \Phi \wedge \Gamma) = \frac{c(\Gamma, \Delta_i \wedge \Phi) \cdot c(\Delta_i, \Phi)}{\sum_{j=1}^{m} [c(\Gamma, \Delta_j \wedge \Phi) \cdot c(\Delta_j, \Phi)]}.$$

Dieses Theorem von Bayes wird (mit der Spezialisierung, daß Φ logisch wahr ist) der Ausgangspunkt der Ableitung der Sukzessionsregel von Laplace sein. An entscheidender Stelle beim Nachweis, daß das Problem der Dispositionsprädikate durch eine adäquate induktive Logik lösbar ist, wird der folgende Lehrsatz benützt, der das induktive Analogon zum Deduktionstheorem (zur Regel der Annahmebeseitigung) der deduktiven Logik ist:

Theorem 4

c sei eine reguläre Wahrscheinlichkeitsfunktion; Φ, Ψ und Σ seien simultan erfüllbar. Dann gilt: $\Phi \wedge \Sigma \models \Psi$ genau dann, wenn $c(\Psi, \Phi \wedge \Sigma) = c(\Phi \rightarrow \Psi, \Sigma)$ sowie $c(\Psi, \Phi \wedge \Sigma) = 1$ genau dann, wenn $c(\Psi, \Phi \wedge \Sigma) = c(\Phi \rightarrow \Psi, \Sigma)$.

Beweis

Die Voraussetzungen seien erfüllt.
(I) Es sei $\Phi \wedge \Sigma \models \Psi$, und damit auch $\Sigma \models \Phi \rightarrow \Psi$; dann gilt:
$c(\Psi, \Phi \wedge \Sigma) = 1$
$= c(\Phi \rightarrow \Psi, \Sigma)$.
(II) Es sei $c(\Psi, \Phi \wedge \Sigma) = c(\Phi \rightarrow \Psi, \Sigma)$. Nun ist $\Phi \rightarrow \Psi$ logisch äquivalent mit $\neg \Phi \vee (\Psi \wedge \Phi)$, und die Sätze $\neg \Phi$, $\Psi \wedge \Phi$ und Σ

sind nicht simultan erfüllbar. Daher gilt mit den Bedingungen (3) und (4):

$$c(\Phi \to \Psi, \Sigma) = c(\neg \Phi \vee (\Psi \wedge \Phi), \Sigma)$$
$$= c(\neg \Phi, \Sigma) + c(\Psi \wedge \Phi, \Sigma).$$

Unter Verwendung der Annahme folgt daraus:

$$c(\Psi, \Phi \wedge \Sigma) = c(\neg \Phi, \Sigma) + c(\Psi \wedge \Phi, \Sigma).$$

Dividiert man beide Seiten der Gleichung durch $c(\Psi, \Phi \wedge \Sigma)$, so erhält man mit Bedingung (5):

$$1 = \frac{c(\neg \Phi, \Sigma)}{c(\Psi, \Phi \wedge \Sigma)} + \frac{c(\Psi \wedge \Phi, \Sigma)}{c(\Psi, \Phi \wedge \Sigma)}$$
$$= \frac{c(\neg \Phi, \Sigma)}{c(\Psi, \Phi \wedge \Sigma)} + c(\Phi, \Sigma)$$

Subtrahiert man von beiden Seiten der Gleichung $c(\Phi, \Sigma)$, so erhält man mit Theorem 1 (d) daraus:

$$\frac{c(\neg \Phi, \Sigma)}{c(\Psi, \Phi \wedge \Sigma)} = 1 - c(\Phi, \Sigma)$$
$$= c(\neg \Phi, \Sigma).$$

Daraus folgt durch arithmetische Umformulierungen:
$c(\Psi, \Phi \wedge \Sigma) = 1$, und mit Bedingung (6): $\Phi \wedge \Sigma \Vdash \Psi$.

Das folgende Theorem wird zur Deduktion der Folgesätze 8 und 9 benötigt, die ihrerseits wesentlich in den Beweis des Lehrsatzes 20 eingehen.[9]

Theorem 5

c sei eine reguläre Wahrscheinlichkeitsfunktion, Φ und Γ seien beide erfüllbar, und es sei $c(\Psi, \Phi) = c(\Psi, \Gamma)$; dann gilt:
(a) $c(\Psi, \Phi) = c(\Psi, \Phi \vee \Gamma)$, falls Φ und Γ nicht simultan erfüllbar sind;

[9] Hier ist zu beachten, daß aus der Erfüllbarkeit eines Satzes Φ und der Erfüllbarkeit einer Aussage Γ noch nicht die *simultane* Erfüllbarkeit von Φ und Γ folgt; Φ und $\neg \Phi$ können beide erfüllbar sein, ohne daß sie simultan erfüllbar sind.

(b) $c(\Psi, \Phi) = c(\Psi, \Phi \vee \Gamma)$ genau dann, wenn Ψ für Γ bezüglich Φ irrelevant ist, d. h. wenn $c(\Gamma, \Psi \wedge \Phi) = c(\Gamma, \Phi)$.

Beweis

Die Voraussetzungen seien gegeben.
(a) Φ und Γ seien nicht simultan erfüllbar. Da $c(\Psi, \Phi) = c(\Psi, \Gamma)$, gilt mit Theorem 2 (d):

$$\frac{c(\Psi \wedge \Phi, T)}{c(\Phi, T)} = \frac{c(\Psi \wedge \Gamma, T)}{c(\Gamma, T)}, \text{ und damit:}$$

$c(\Psi \wedge \Phi, T) \cdot c(\Gamma, T) = c(\Psi \wedge \Gamma, T) \cdot c(\Phi, T)$. Addiert man auf beiden Seiten dieser Gleichung $c(\Psi \wedge \Phi, T) \cdot c(\Phi, T)$, so erhält man:
$c(\Psi \wedge \Phi, T) \cdot c(\Phi, T) + c(\Psi \wedge \Phi, T) \cdot c(\Gamma, T) =$
$c(\Psi \wedge \Phi, T) \cdot c(\Phi, T) + c(\Psi \wedge \Gamma, T) \cdot c(\Phi, T)$.
Durch geeignete Zusammenfassung erhält man daraus:
$c(\Psi \wedge \Phi, T) \cdot [c(\Phi, T) + c(\Gamma, T)] =$
$c(\Phi, T) \cdot [c(\Psi \wedge \Phi, T) + c(\Psi \wedge \Gamma, T)]$.
Da Φ und Γ nicht simultan erfüllbar sind, sind es auch $\Psi \wedge \Phi$ und $\Psi \wedge \Gamma$ nicht; wegen (4) gilt dann unter Verwendung von (3):

$$\begin{aligned} c(\Psi \wedge \Phi, T) \cdot c(\Phi \vee \Gamma, T) &= c(\Phi, T) \cdot c((\Psi \wedge \Phi) \vee (\Psi \wedge \Gamma), T) \\ &= c(\Phi, T) \cdot c(\Psi \wedge (\Phi \vee \Gamma), T). \end{aligned}$$

Daraus folgt durch arithmetische Umformungen:

$$\frac{c(\Psi \wedge \Phi, T)}{c(\Phi, T)} = \frac{c(\Psi \wedge (\Phi \vee \Gamma), T)}{c(\Phi \vee \Gamma, T)}, \text{ sowie mit Theorem 2 (d):}$$

$$c(\Psi, \Phi) = c(\Psi, \Phi \vee \Gamma).$$

(b) Der Beweis verläuft ganz analog, nur daß hier statt der Bedingung (4) das Theorem 1 ( ) angewandt werden muß; dabei ist zu berücksicht gen, daß aus der Voraussetzung $c(\Gamma, \Phi) = c(\Gamma, \Psi \wedge \Phi)$ sich nach Theorem 2 (d) die Gleichung
$c(\Psi \wedge \Phi, T) \cdot [-c(\Phi \wedge \Gamma, T)] = c(\Phi, T) \cdot [-c(\Psi \wedge \Phi \wedge \Gamma, T)]$
ergibt. Offensichtlich ist hier die Argumentation auch umkehrbar.

Gegen den Teil der Bedingung (3), der besagt, daß $c(\Psi, \Phi) = c(\Delta, \Phi)$, falls bei erfüllbarem Φ der Satz Ψ logisch äquivalent mit Δ ist, sind

von Philosophen gelegentlich Einwände erhoben worden. So ist gesagt worden, daß auf Grund einer Beobachtung von 20 roten Kühen die Hypothesen, daß alle Raben schwarz sind, und daß alles, was nicht schwarz ist, kein Rabe ist, verschiedene Wahrscheinlichkeitswerte erhalten müßten. Dieser Einwand kann dadurch zurückgewiesen werden, daß man jenen Teil der Bedingung (3) aus den übrigen ableitet, wobei (2) in folgendem Sinn zu verstärken ist: Nicht nur Φ möge auf Grund von Φ den Wahrscheinlichkeitswert 1 erhalten, sondern jeder aus Φ logisch folgende Satz Ψ (gegen diese Verstärkung dürften wohl keine Einwände möglich sein).[10]

Theorem 6

c ist eine reguläre Wahrscheinlichkeitsfunktion genau dann, wenn die Bedingungen (1), (4), (5) und (6) von D 4 erfüllt sind, sowie wenn:

(2′) $c(\Psi, \Phi) = 1$ für jeden erfüllbaren Satz Φ und jeden aus Φ logisch folgenden Satz Ψ, und

(3′) $c(\Psi, \Phi) = c(\Psi, \Gamma)$ für jeden Satz Ψ, jeden erfüllbaren Satz Φ und jeden mit Φ logisch äquivalenten Satz Γ.

Beweis

(I) c sei eine reguläre Wahrscheinlichkeitsfunktion. Dann ergibt sich (3′) unmittelbar aus (3), während (2′) nach Theorem 1 (a) aus (1), (3) und (5) folgt.

(II) c erfülle die Bedingungen (1), (2′), (3′), (4), (5) und (6). Dann folgt (2) unmittelbar aus (2′). (3′) ergibt sich wie folgt: Φ sei erfüllbar, und Δ sei logisch äquivalent mit Ψ. Also ist $\Psi \vee \neg \Delta$ logisch wahr und folgt daher aus Φ, und es sind Ψ, $\neg \Delta$ und Φ nicht simultan erfüllbar. Aus (2′) und (4) ergibt sich dann:

$1 = c(\Psi \vee \neg \Delta, \Phi) = c(\Psi, \Phi) + c(\neg \Delta, \Phi)$.

Ferner ist $\Delta \vee \neg \Delta$ logisch wahr, und es sind Δ, $\neg \Delta$ und Φ nicht simultan erfüllbar; wiederum gilt mit (2′) und (4):

$1 = c(\Delta \vee \neg \Delta, \Phi) = c(\Delta, \Phi) + c(\neg \Delta, \Phi)$.

[10] Diese Form der Charakterisierung des Wahrscheinlichkeitsbegriffs wird von Lehmann [1], S. 254, benützt.

Aus beiden Gleichungen folgt:
$c(\Psi, \Phi) + c(\neg \Delta, \Phi) = c(\Delta, \Phi) + c(\neg \Delta, \Phi)$, und damit:
$c(\Psi, \Phi) = c(\Delta, \Phi)$. Daraus und aus (3') folgt jedoch (3).

Der folgende Lehrsatz zeigt, daß das Spezielle Additionsprinzip durch eine intuitiv einleuchtendere Bedingung ersetzt werden kann:[11]

Theorem 7

c ist eine reguläre Wahrscheinlichkeitsfunktion genau dann, wenn die Bedingungen (1), (2), (3), (5) und (6) von D 4 erfüllt sind und wenn

(4') $c(\Psi, \Phi) + c(\neg \Psi, \Phi) = 1$ für jeden Satz Ψ und jeden erfüllbaren Satz Φ.

Beweis

(I) c sei eine reguläre Wahrscheinlichkeitsfunktion. Dann erfüllt c nach Theorem 1 (c) auch die Bedingung (4').
(II) Angenommen, c erfüllt die Bedingungen (1), (2), (3), (5) und (6) sowie (4'). Es wird zunächst Theorem 1 (i) bewiesen. Φ sei erfüllbar.
(α) Angenommen, $\Psi \wedge \Phi$ ist erfüllbar. Nach (4') gilt dann:
$c(\Delta, \Psi \wedge \Phi) + c(\neg \Delta, \Psi \wedge \Phi) = 1$.
Multipliziert man beide Seiten mit $c(\Psi, \Phi)$, so erhält man daraus:
$c(\Psi, \Phi) \cdot c(\Delta, \Psi \wedge \Phi) + c(\Psi, \Phi) \cdot c(\neg \Delta, \Psi \wedge \Phi) = c(\Psi, \Phi)$.
Durch zweimalige Anwendung von (5) erhält man daraus:
$c(\Psi \wedge \Delta, \Phi) + c(\Psi \wedge \neg \Delta, \Phi) = c(\Psi, \Phi)$.
(β) Angenommen, $\Psi \wedge \Phi$ ist nicht erfüllbar; also sind Ψ und Φ nicht simultan erfüllbar, und es gilt mit Theorem 1 (f)[12]: $c(\Psi, \Phi) = 0$. Ist $\Psi \wedge \Phi$ nicht erfüllbar, so sind auch $\Psi \wedge \Delta$ und Φ nicht simultan erfüllbar, und es gilt darüberhinaus: $c(\Psi \wedge \Delta, \Phi) = 0$; da ferner dann auch $\Psi \wedge \neg \Delta$ und Φ nicht simultan erfüllbar sind, gilt schließlich noch: $c(\Psi \wedge \neg \Delta, \Phi) = 0$. Also hat man auch hier:

---

11 Diese Form der Charakterisierung des Wahrscheinlichkeitsbegriffs stammt von Wright [2], S. 93, [1], S. 176 und 187ff.
12 Theorem 1 (a) und (b) sind ohne (4) beweisbar, Theorem 1 (c) ist gerade (4'), und Theorem 1 (d), (e) und (f) verwenden zum Beweis nur Theorem 1 (c), nicht jedoch (4).

$c(\Psi, \Phi) = c(\Psi \wedge \Delta, \Phi) + c(\Psi \wedge \neg \Delta, \Phi)$, und damit Theorem 1 (i). Daraus folgt jedoch unter Verwendung von (3) das allgemeine Additionsprinzip, und daraus schließlich mit Theorem 1 (f) die Bedingung (4').

Der Beweis des Theorems 1 (a) zeigt, daß die Bedingung (2) in D 4 folgendermaßen abgeschwächt werden kann, da dabei nur (2*) benötigt wird:

Theorem 8

c ist eine reguläre Wahrscheinlichkeitsfunktion genau dann, wenn die Bedingungen (1) und (3) bis (6) von D 4 erfüllt sind, und wenn (2*) $c(\Phi, \Phi) \neq 0$ für jeden erfüllbaren Satz Φ.

Beweis

(a) c sei eine reguläre Wahrscheinlichkeitsfunktion, so daß wegen (2) für einen erfüllbaren Satz Φ gilt: $c(\Phi, \Phi) = 1$; dann gilt jedoch auch: $c(\Phi, \Phi) \neq 0$.
(b) Die Bedingungen (1) und (3) bis (6) sowie (2*) seien erfüllt. Φ sei ein erfüllbarer Satz. Wegen (5) gilt dann:
$c(\Phi \wedge \Phi, \Phi) = c(\Phi, \Phi) \cdot c(\Phi, \Phi \wedge \Phi)$.
Da wegen (3) gilt: $c(\Phi \wedge \Phi, \Phi) = c(\Phi, \Phi)$ und:
$c(\Phi, \Phi \wedge \Phi) = c(\Phi, \Phi)$, folgt daraus mit (2*):
$c(\Phi, \Phi) = c(\Phi, \Phi)/c(\Phi, \Phi) = 1$.

Die Theorie der regulären Wahrscheinlichkeitsfunktionen ist in der Theorie der regulären Bestätigungsmethoden enthalten: Jeder Satz, der in jener Theorie gilt, gilt auch in dieser. Dies zeigt man am einfachsten dadurch, daß man aus dem Begriff der regulären Bestätigungsmethode den der regulären Wahrscheinlichkeitsfunktion, wie er durch die Bedingungen (1) bis (6) ausgedrückt wird, folgert:

Theorem 9

Jede reguläre Bestätigungsmethode ist eine reguläre Wahrscheinlichkeitsfunktion.

Beweis

c sei eine reguläre Bestätigungsmethode.

(1) Nach D 2 gilt für jeden Satz Δ: $c(\Delta, T) \geqslant 0$ und für jeden erfüllbaren Satz Φ: $c(\Phi, T) > 0$, nach D 3 daher für jeden Satz Ψ und jeden erfüllbaren Satz Φ: $c(\Psi, \Phi) \geqslant 0$.

(2) Φ sei ein erfüllbarer Satz, so daß nach D 1 und D 2: $c(\Phi, T) > 0$. Nun ist die Klasse Zb^{Φ} der Zustandsbeschreibungen, die Φ logisch implizieren, identisch mit der Klasse $Zb^{\Phi \wedge \Phi}$ der Zustandsbeschreibungen, die $\Phi \wedge \Phi$ logisch implizieren; daher ist
$c(\Phi, T) = c(\Phi \wedge \Phi, T)$ und somit

$$c(\Phi, \Phi) = \frac{c(\Phi \wedge \Phi, T)}{c(\Phi, T)} = 1.$$

(3) Φ sei ein erfüllbarer Satz; also ist $c(\Phi, T) > 0$. Φ sei mit Γ, Ψ mit Δ und daher $\Psi \wedge \Phi$ mit $\Delta \wedge \Gamma$ logisch äquivalent. Dann ist $Zb^{\Phi} = Zb^{\Gamma}$ und $Zb^{\Psi \wedge \Phi} = Zb^{\Delta \wedge \Gamma}$, sowie $c(\Phi, T) = c(\Gamma, T)$ und $c(\Psi \wedge \Phi, T) = c(\Delta \wedge \Gamma, T)$. Also gilt:

$$c(\Phi, T) = \frac{c(\Psi \wedge \Phi, T)}{c(\Phi, T)} = \frac{c(\Delta \wedge \Gamma, T)}{c(\Gamma, T)} = c(\Delta, \Gamma).$$

(4) Φ sei ein erfüllbarer Satz und Ψ sowie Δ seien Sätze, so daß Ψ, Δ und Φ nicht simultan erfüllbar sind. Es ist also $c(\Phi, T) > 0$, und es sind die Klassen $Zb^{\Psi \wedge \Phi}$ und $Zb^{\Delta \wedge \Phi}$ elementfremd. Nun ist $Zb^{(\Psi \wedge \Phi) \vee (\Delta \wedge \Phi)}$ identisch mit $Zb^{(\Psi \vee \Delta) \wedge \Phi}$. Da $\Psi \wedge \Phi$ logisch äquivalent mit der Standardkonjunktion von $Zb^{\Psi \wedge \Delta}$, $\Delta \wedge \Phi$ logisch äquivalent mit der von $Zb^{\Delta \wedge \Phi}$ und $(\Psi \vee \Delta) \wedge \Phi$ logisch äquivalent mit der von $Zb^{(\Psi \vee \Delta) \wedge \Phi}$ ist, gilt:

$$c(\Psi \vee \Delta, \Phi) = \frac{c((\Psi \vee \Delta) \wedge \Phi, T)}{c(\Phi, T)}$$

$$= \frac{\sum\limits_{\Sigma \in Zb^{(\Psi \vee \Delta) \wedge \Phi}} c(\Sigma, T)}{\sum\limits_{\Sigma \in Zb^{\Phi}} c(\Sigma, T)}$$

$$= \frac{\sum_{\Sigma \in Zb^{\Psi \wedge \Phi}} c(\Sigma, T)}{\sum_{\Sigma \in Zb^{\Phi}} c(\Sigma, T)} + \frac{\sum_{\Sigma \in Zb^{\Delta \wedge \Phi}} c(\Sigma, T)}{\sum_{\Sigma \in Zb^{\Phi}} c(\Sigma, T)}$$

$$= \frac{c(\Psi \wedge \Phi, T)}{c(\Phi, T)} + \frac{c(\Delta \wedge \Phi, T)}{c(\Phi, T)}$$

$$= c(\Psi, \Phi) + c(\Delta, \Phi).$$

(5) Es sei $\Psi \wedge \Phi$ und somit auch Φ erfüllbar, so daß $c(\Psi \wedge \Phi, T) > 0$ und $c(\Phi, T) > 0$. Wegen D 3 gilt dann:

$$c(\Delta \wedge \Psi, \Phi) = \frac{c(\Delta \wedge \Psi \wedge \Phi, T)}{c(\Phi, T)}$$

$$= \frac{c(\Psi \wedge \Phi, T)}{c(\Phi, T)} \cdot \frac{c(\Delta \wedge \Psi \wedge \Phi, T)}{c(\Psi \wedge \Phi, T)}$$

$$= c(\Psi, \Phi) \cdot c(\Delta, \Psi \wedge \Phi)$$

und wegen (3) schließlich: $c(\Psi \wedge \Delta, \Phi) = c(\Psi, \Phi) \cdot c(\Delta, \Psi \wedge \Phi)$.
(6) Φ sei ein erfüllbarer Satz und Ψ ein nicht logisch aus Φ folgender Satz; also ist $c(\Phi, T) > 0$. Angenommen, $c(\Psi, \Phi) = 1$, dann gilt mit D 3: $c(\Psi \wedge \Phi, T) = c(\Phi, T)$, und mit D 2:

$$\sum_{\Sigma \in Zb^{\Psi \wedge \Phi}} c(\Sigma, T) = \sum_{\Sigma \in Zb^{\Phi}} c(\Sigma, T).$$

Nun ist $Zb^{\Psi \wedge \Phi}$ eine Teilklasse von Zb^{Φ}, da jedes Σ aus $Zb^{\Psi \wedge \Phi}$ logisch Φ impliziert und also in Zb^{Φ} ist. Wäre andererseits Zb^{Φ} keine Teilklasse von $Zb^{\Psi \wedge \Phi}$, so wäre wegen D 1:

$$\sum_{\Sigma \in Zb^{\Psi \wedge \Phi}} c(\Sigma, T) < \sum_{\Sigma \in Zb^{\Phi}} c(\Sigma, T),$$

was der obigen Gleichung widerspricht. Also ist $Zb^{\Psi \wedge \Phi}$ mit Zb^{Φ} identisch. Daraus folgt jedoch, daß $\Psi \wedge \Phi$ mit Φ logisch äquivalent ist und daß also Ψ aus Φ logisch folgt, was der Voraussetzung widerspricht. Also gilt: $c(\Psi, \Phi) \neq 1$.

Andererseits sieht man sofort, daß dieser Lehrsatz umkehrbar ist:

Theorem 10

Jede reguläre Wahrscheinlichkeitsfunktion ist eine reguläre Bestätigungsmethode.

Beweis

c sei eine reguläre Wahrscheinlichkeitsfunktion. Es ist zu zeigen, daß c die Bedingungen von D 1, D 2 und D 3 erfüllt.

(α) Eine Zustandsbeschreibung Σ ist ein erfüllbarer Satz; also ist wegen Theorem 2 (b): $c(\Sigma, T) > 0$. Andererseits sind die Zustandsbeschreibungen in ihrer Gesamtheit erschöpfend, d. h. bei jeder Interpretation ist wenigstens eine von ihnen wahr, und ihre Adjunktion Ξ (die Verknüpfung der einzelnen Sätze durch „v" zu einem einzigen Satz) ist logisch wahr. Also wird Ξ von T logisch impliziert, und es gilt mit Theorem 1 (a): $c(\Xi, T) = 1$. Da außerdem keine zwei Zustandsbeschreibungen Σ_1 und Σ_2 simultan erfüllbar sind, kann für die p Zustandsbeschreibungen p—1mal das Spezielle Additionsprinzip (4) angewandt werden, so daß man erhält:

$$\sum_{\Sigma \in Zb} c(\Sigma, T) = c(\Xi, T) = 1.$$

(β) Δ ist logisch äquivalent mit der Adjunktion jener Zustandsbeschreibungen, die Δ logisch implizieren. Also haben Δ und diese Adjunktion wegen (3) die gleiche Aprioriwahrscheinlichkeit. Da ferner keine zwei Zustandsbeschreibungen miteinander simultan erfüllbar sind, kann (4) angewandt werden, und es ist die Aprioriwahrscheinlichkeit jener Adjunktion identisch mit der Summe der Aprioriwahrscheinlichkeiten der einzelnen Zustandsbeschreibungen aus Zb^{Δ}. Daher gilt:

$$\sum_{\Sigma \in Zb^{\Delta}} c(\Sigma, T) = c(\Delta, T).$$

(γ) Φ sei erfüllbar; dann gilt mit Theorem 2 (d):

$$c(\Psi, \Phi) = \frac{c(\Psi \wedge \Phi, T)}{c(\Phi, T)}.$$

Die Theoreme 9 und 10 implizieren unmittelbar das

Korollar 1

c ist eine reguläre Bestätigungsmethode genau dann, wenn c eine reguläre Wahrscheinlichkeitsfunktion ist.

Es zeigt sich also, daß alle und nur die regulären Bestätigungsmethoden reguläre Wahrscheinlichkeitsfunktionen sind, daß die Begriffe der regulären Bestätigungsmethode und der regulären Wahrscheinlichkeitsfunktion somit umfangsgleich und folglich miteinander identisch sind. Charakterisiert man den Begriff der Wahrscheinlichkeitsfunktion in Anlehnung an den der Bestätigungsmethode durch die Bedingungen (1) bis (5), wobei die Sätze „Φ ist erfüllbar" bzw. „$\Psi \wedge \Phi$ ist erfüllbar" durch „$c(\Phi, T) \neq 0$" bzw. „$c(\Psi \wedge \Phi, T) \neq 0$" ersetzt werden, so sieht man sofort, daß auch diese Begriffe miteinander identisch sind.

3. Strenge Kohärenz

Auf eine dritte Weise ist der Wahrscheinlichkeitsbegriff im Hinblick auf seine Anwendung bei Wetten charakterisierbar. Wer genügend Geld besitzt, kann ja auf alle möglichen Sachverhalte Wetten abschließen, und seine Gewinnchancen sind auf längere Sicht mitbestimmt durch die Rationalität seiner Einsätze. Daher darf gesagt werden, daß Wetten ein Hauptanwendungsgebiet der Wahrscheinlichkeitstheorie sind.

Φ möge das gemeinsame Wissen zweier Personen x_1 und x_2 sein. Eine Wette des x_1 gegen x_2 auf die Hypothese Ψ bezüglich dieses Wissens Φ ist dann eine Vereinbarung von der Art, daß x_1 dem x_2 den Betrag t_1 bezahlt, wenn sich Ψ als falsch herausstellt, und daß er von x_2 den Betrag t_2 erhält, falls sich Ψ als wahr erweist; für den Fall, daß sich Φ wider Erwarten als falsch herausstellen sollte und beide also genötigt wären, ihr Wissen zu korrigieren, sei darüberhinaus vereinbart, daß die Wette als aufgehoben gilt, daß dann also keiner von beiden etwas bezahlt, gleichgültig, wie es um die Wahrheit von Ψ bestellt ist. Es wird dabei nicht vorausgesetzt, daß die Beträge t_1 und t_2 gleich sind. Dies ist zwar im Alltag häufig, jedoch durchaus nicht immer der Fall; beim Lotteriespiel etwa wettet man darauf, daß eine bestimmte Zahl gezogen wird, und der Einsatz des Wettenden beträgt z. B. 100,— DM, der der Gesellschaft jedoch vielleicht 100000,— DM.

Eine solche Wette nennt man fair, wenn sie keinen der beiden Partner

begünstigt, so daß eine Person y es also einer Person z überlassen kann, die Rolle des x_1 oder des x_2 zu übernehmen, ohne dabei Nachteile befürchten zu müssen. Nennt man das Verhältnis des Einsatzes des auf Ψ wettenden x_1 zum Gesamteinsatz (zur Summe der beiden Einsätze t_1 und t_2) den Wettquotienten, so wird y versuchen, einen fairen Wettquotienten und damit eine faire Wette zu ermitteln, wenn z entscheiden darf, ob y für oder gegen Ψ zu wetten hat und y zugesagt hat, diese Wahl zu akzeptieren.

Es muß nun noch bestimmt werden, wann y es dem z überlassen kann, die Rolle des x_1 oder des x_2 zu übernehmen, ohne Nachteile befürchten zu müssen. Naheliegend ist es, zu antworten: wenn y sich nicht entscheiden kann, die eine Rolle der anderen vorzuziehen. Dieser Fall wird jedoch genau so selten eintreten, wie der des Esels, der zwischen zwei gleich großen Heuhaufen verhungert, weil er sich für keinen von beiden entscheiden kann. Ein Subjekt wird immer irrationale Gründe dafür haben, die eine Rolle der anderen vorzuziehen. Es müßte daher geantwortet werden: Jener Fall tritt ein, wenn y sich auf Grund rationaler Gründe allein nicht entscheiden kann, ob er *für* oder *gegen* wetten soll. Die Angabe aller rationalen Gründe ist nun äquivalent mit der Formulierung des adäquaten Wahrscheinlichkeitsbegriffs, der jedoch gesucht ist und bei dieser theoretischen Überlegung (im Gegensatz zur praktischen Anwendung) nicht als vorgegeben angesehen werden darf.

Da eine unmittelbare Explikation des Begriffs der Fairneß zu Schwierigkeiten führt, wird auf seine Definition zunächst verzichtet und an ihrer Stelle der der strengen Kohärenz analysiert. Eine Anzahl von Wetten, die Gesamtheit der damit gegebenen Wettquotienten sowie eine Funktion c, die diese Wettquotienten liefert, werden dabei streng kohärent genannt, wenn bei ihnen, falls ein Gewinn, so auch ein Verlust möglich ist. Denn ein Wettsystem (eine Klasse von Wetten), durch das zwar ein Gewinn, nicht aber ein Verlust ausgeschlossen wird, ist offensichtlich nicht fair; ob die Umkehrung ebenfalls behauptet werden darf, sei vorerst dahingestellt.

Der Zusammenhang, der zwischen gewissen Wettsystemen und dem Wahrscheinlichkeitsbegriff besteht, sei nun an einigen Beispielen verdeutlicht. Der Gesamteinsatz u sei dabei die Summe der beiden Einsätze t_1 und t_2. Für den Wettquotienten r gilt dann: $r = t_1/u$, bzw.

$r = t_1/(t_1 + t_2)$. Der Wettquotient einer fairen Wette auf Ψ mit dem Wissen Φ ist ja gerade der Wahrscheinlichkeitswert von Ψ bezüglich Φ; daher ist für das folgende dieses Verhältnis des Einsatzes des auf die Hypothese Wettenden zum Gesamteinsatz wichtig. Sind der Gesamteinsatz und der Wettquotient bekannt, so können die beiden Einsätze t_1 und t_2 folgendermaßen berechnet werden: $t_1 = r \cdot u$, $t_2 = (1 - r) \cdot u$.

Gegeben sei eine nichtleere Klasse G mit n Individuen. Das zur Verfügung stehende Erfahrungsdatum Φ möge für kein Ding aus G aussagen, ob es in der Klasse F ist oder nicht. Angenommen, x_1 hat gegen x_2 n-mal für jedes Ding aus G mit dem Gesamteinsatz u und dem Wettquotienten r gewettet, daß es in der Klasse F ist; angenommen ferner, $s \cdot n$ der n betrachteten Objekte sind in F. Dann gewinnt x_1 $s \cdot n$ Wetten und erhält den Betrag $s \cdot n \cdot t_2$. Da er $(1 - s) \cdot n$ Wetten verliert, bezahlt er $(1 - s) \cdot n \cdot t_1$. Seine Gesamtbilanz ist demnach $s \cdot n \cdot t_2 - (1 - s) \cdot n \cdot t_1$, was mit $n \cdot (t_1 + t_2) \cdot (s - r)$ identisch ist.

Es wird vorausgesetzt, daß x_1 tatsächlich auf die einzelnen n Hypothesen gewettet hat, so daß der Gesamteinsatz u jeweils positiv ist (ist der Gesamteinsatz 0, so liegt keine Wette vor, ist er negativ, so liegt, wie später gezeigt wird, eine Wette gegen die betreffende Hypothese vor). Dann erzielt also x_1 einen Gewinn, wenn $s - r > 0$, d. h. wenn $r < s$, und er verliert, wenn $s - r < 0$ bzw. wenn $r > s$. Er gewinnt und verliert nichts, wenn $s - r = 0$, so daß $r = s$.

Angenommen, aus Φ folgt, daß s die relative Häufigkeit von F in G ist. Da weiterhin vorausgesetzt wird, daß x_1 und x_2 vernünftige Leute sind, die nichts unüberlegt tun, wobei das Wissen eines jeden von ihnen durch Φ beschrieben wird, kommt zwischen beiden eine Wette nur dann zustande, wenn der Wettquotient r der einzelnen Wetten mit s identisch ist. Denn ist $r < s$, so wird x_2 die Wetten ablehnen, und x_1 wird die Wetten nicht akzeptieren, wenn $r > s$. Es hat für beide dann natürlich keinen Reiz, für alle n Gegenstände aus G Wetten mit den gleichen Einsätzen abzuschließen. Sie werden dies vielmehr für eine nichtleere und von G verschiedene Teilklasse H von G tun, oder aber für alle Dinge aus G, aber mit jeweils verschieden großen Gesamteinsätzen, und sie werden als Schätzwert der relativen Häufigkeit von F in H die bekannte und durch Φ beschriebene rela-

tive Häufigkeit s von F in G nehmen. Diesen Schätzwert s werden sie dann für jede der einzelnen Wetten darauf, daß ein Ding aus H in der Klasse F ist, als fairen Wettquotienten betrachten; denn die Schätzung des Gesamtverlustes ist in diesem Fall für beide 0.

Angenommen, H enthält nur ein einziges Ding a, dessen Name α ist; Π bezeichne die Klasse F. Dann ist die Schätzung der relativen Häufigkeit von F in H auf Grund des Wissens Φ identisch mit der Wahrscheinlichkeit von $\alpha \in \Pi$ auf Grund des Wissens Φ. Damit ist der Zusammenhang zwischen den Begriffen der Schätzung der relativen Häufigkeit und der logischen Wahrscheinlichkeit hergestellt.

Der Begriff des Wettsystems soll im folgenden in seiner allgemeinsten Verwendung analysiert werden. Es darf dabei nicht allzu großes Gewicht auf den Ausdruck „System" gelegt werden. Ein Wettsystem soll vielmehr nur eine Klasse von gleichzeitig abgeschlossenen Wetten sein, wobei für jede der einzelnen Wetten festgelegt ist, welches geordnete Paar $\langle \Psi, \Phi \rangle$ von Sätzen benützt wird, wie groß die einzelnen Einsätze und Wettquotienten sind und ob *auf* die Hypothese Ψ oder *gegen* sie gewettet wird. Die beiden Einsätze können durch die oben angeführten Gleichungen aus Wettquotient und Gesamteinsatz errechnet werden. Es bleibt die Frage, wodurch ausgedrückt werden kann, daß die Person y für oder gegen die Hypothese Ψ wettet. Aufschluß hierüber gibt uns die folgende Tabelle:

	Gewinn von x_1	Verlust von x_1 = Gewinn von x_2	
1. Fall: Φ ist wahr			
Fall a: Ψ ist wahr	$(1-r)\cdot u$	$-(1-r)\cdot u$	$[=(1-r)(-u)]$
Fall b: Ψ ist falsch	$-r\cdot u$	$r\cdot u$	$[=-r\cdot(-u)]$
2. Fall: Φ ist falsch	0	0	$[=-r\cdot 0]$

Eine Wette *auf* Ψ ist also gleichbedeutend mit einem *positiven* Einsatz u, eine Wette *gegen* Ψ mit einem *negativen* Einsatz u und eine nicht abgeschlossene oder annullierte Wette mit dem Einsatz 0. Die natürliche Erweiterung des Wertbereiches der Variablen „u" auf die Gesamtheit der reellen Zahlen hat also zur Folge, daß durch den Gesamteinsatz allein charakterisiert werden kann, ob auf die Hypothese,

gegen sie oder überhaupt nicht gewettet wird. Die tatsächliche Größe von u ist jedoch im folgenden ohne Bedeutung; relevant ist nur die Frage, ob u positiv, 0 oder negativ ist. Der Begriff „Wette" kann demnach folgendermaßen definiert werden:[13]

D 5 d ist eine Wette auf Ψ bezüglich Φ mit dem Wettquotienten r und dem Gesamteinsatz u genau dann, wenn $d = \langle \Psi, \Phi, r, u \rangle$, wobei Ψ ein Satz und Φ ein erfüllbarer Satz ist, sowie r und u reelle Zahlen sind.

Durch diese Definition wird nicht ausgeschlossen, daß $r > 1$ oder daß $r < 0$. Ist z. B. $r = 3$, so liegt bei positivem u und wahrem Φ folgende Situation vor: Wenn sich herausstellt, daß Ψ wahr ist, so erhält x_1 den Betrag $(1 - 3) \cdot u$, d. h. er bezahlt dann $2 \cdot u$; wenn sich hingegen ergibt, daß Ψ falsch ist, dann bekommt er $-3 \cdot u$, d. h. er bezahlt dann $3 \cdot u$. Wer mit diesem Wettquotienten auf eine Hypothese wettet, hat also, falls die Prämisse Φ wahr is', in jedem Fall einen Verlust, gleichgültig, ob die Hypothese, auf die gewettet wird, wahr oder falsch ist; der zu bezahlende Betrag hängt allerdings davon ab, ob Ψ wahr ist oder nicht. Das Umgekehrte gilt von einer Wette auf Ψ (so daß u positiv ist) mit $r = -3$ und wahrem Φ: Wenn sich herausstellt, daß auch Ψ wahr ist, dann erhält er den Betrag $(1 - (-3)) \cdot u$, d. h. $4 \cdot u$, und wenn sich erweist, daß Ψ falsch ist, so bekommt er $-(-3) \cdot u$, d. h. $3 \cdot u$. Es erweist sich daher als zweckmäßig, auch bezüglich der Variablen „r" jene Erweiterung vorzunehmen, da man in einem allgemeinen Sinn hier ebenfalls noch von Wetten sprechen kann und da derartige Wettquotienten durch den Begriff der strengen Kohärenz ausgeschlossen werden.

Nach Voraussetzung ist c eine zweistellige Funktion; die Variablen der Argumentstellen stehen für Sätze, und als Funktionswerte kommen reelle Zahlen in Frage. Die in D 5 formulierte Bedingung, daß Φ ein *erfüllbarer* Satz sein muß, wird für die Funktion c, die die Wettquotienten liefert, die gleichen Konsequenzen haben wie für reguläre Bestätigungsmethoden in D 3 und für reguläre Wahrscheinlichkeitsfunktionen in D 4: Sie wird den Definitionsbereich von c auf erfüllbare Prämissen Φ beschränken.

[13] Vgl. für das folgende Bruno de Finetti [1], [2], Ramsey [1], Lehmann [1] und Kemeny [6].

Ein Wettsystem ist eine Klasse von Wetten. Von den Wettquotienten dieser Wetten kann angenommen werden, daß sie durch eine Methode, durch eine Funktion, geliefert worden sind; denn selbst dann, wenn diese Wettquotienten scheinbar ganz willkürlich gewählt worden sind, kann eine Funktion angegeben werden, die sie als Werte liefert. Es wird daher der Begriff des Wettsystems im Hinblick auf eine Funktion c definiert, die diese Wettquotienten vorgibt. Da wir als Menschen nur endlich viele Wetten abschließen können, bedeutet es für die prakti che Anwendung keine Einschränkung, sich nur auf endliche (und selbstverständlich nichtleere) Wettsysteme zu beziehen: es wird sich erweisen, daß dies zur Ableitung der Wahrscheinlichkeitsaxiome genügt. Der Begriff des Wettsystems wird also als Relation zwischen einer Klasse D von Wetten und einer Funktion c durch die folgende Definition eingeführt:

D 6 D ist ein mit c übereinstimmendes endliches Wettsystem genau dann, wenn D eine nichtleere endliche Klasse ist und wenn es zu jedem d aus D Sätze Ψ und Φ sowie reelle Zahlen r und u gibt, so daß d eine Wette auf Ψ bezüglich Φ mit dem Wettquotienten r und dem Gesamteinsatz u ist, wobei $c(\Psi, \Phi) = r$.

D ist demnach eine Klasse von geordneten Quadrupeln, also eine vierstellige Beziehung.

Ein Ausgang eines Wettsystems D ist ein Satz, der für jede einzelne Wette aus D beschreibt, ob die Prämisse zutrifft, und wenn ja, ob die Hypothese, auf die gewettet wird, gilt. Er kann also als Standardkonjunktion einer Gesamtheit von Sätzen formuliert werden, die für jede einzelne Wette d aus D entweder das Konjunktionsglied $\neg \Phi$ enthält (so daß die Wette als nicht abgeschlossen gilt), oder aber $\Psi \wedge \Phi$, oder schließlich $\neg \Psi \wedge \Phi$. Ein normaler Ausgang eines Wettsystems D ist ein Ausgang, bei dem wenigstens eine Wette aus D verwirklicht ist, so daß nicht alle Wetten annulliert werden; ein solcher Ausgang ist demnach verschieden von der Standardkonjunktion der Negationen der einzelnen Prämissen, d. h. der Negationen der Sätze des 2-Bereichs[14] von D.

14 Der 2-Bereich von D ist die Gesamtheit der Sätze Φ, so daß es Ψ, r und u gibt, wobei $\langle \Psi, \Phi, r, u \rangle$ Element des Wettsystems D ist (Abkürzung: wobei $\langle \Psi, \Phi, r, u \rangle \in D$).

D 7 D sei ein mit c übereinstimmendes endliches Wettsystem. Dann ist Σ ein Ausgang bezüglich D [kurz: Σ e A_D] genau dann, wenn Σ erfüllbar und die Standardkonjunktion der Klasse der Sätze Θ ist, so daß für alle Ψ, Φ, r und u gilt: Falls $\langle\Psi,\Phi,r,u\rangle$ e D, dann ist Θ der Satz $\Psi \wedge \Phi$, der Satz $\neg\, \Psi \wedge \Phi$ oder der Satz $\neg\, \Phi$.

D 8 D sei ein mit c übereinstimmendes endliches Wettsystem. Dann ist Σ ein normaler Ausgang bezüglich D [Σ e A^n_D] genau dann, wenn Σ ein Ausgang bezüglich D ist, jedoch verschieden ist von der Standardkonjunktion der Negationen der Sätze des 2-Bereichs von D.

Der Begriff der Wette ist sehr allgemein eingeführt worden. Es wird durch ihn nicht festgelegt, welchen Gewinn die Person x_1, die ja auf Ψ wettet, für die einzelnen Fälle erwartet. Es müssen nämlich die folgenden Fälle unterschieden werden:

(1) Φ ist wahr, mit den Unterfällen:
   (a) Ψ ist wahr,
   (b) Ψ ist falsch, sowie
(2) Φ ist falsch.

Objektsprachlich werden diese Fälle durch die Sätze $\Psi \wedge \Phi$, $\neg\, \Psi \wedge \Phi$ und $\neg\, \Phi$ charakterisiert. Von ihnen ist bei jeder Interpretation wenigstens und höchstens, also genau einer wahr. Die obigen Beispiele legen es nahe, den Begriff „Gewinn bei einer Wette d bezüglich eines Ausgangs Σ" [Abkürzung: „gn (d, Σ)"] durch folgende bedingte Definition einzuführen:[15]

D 9 D sei ein mit c übereinstimmendes endliches Wettsystem, das die Wette d auf Ψ bezüglich Φ mit dem Wettquotienten r und dem Gesamteinsatz u enthält, und Σ sei ein Ausgang bezüglich D. Dann ist

$$\text{gn}(d, \Sigma) = \begin{cases} (1 - r) \cdot u, & \text{wenn aus } \Sigma \text{ logisch } \Psi \wedge \Phi \text{ folgt,} \\ -r \cdot u, & \text{wenn aus } \Sigma \text{ logisch } \neg\, \Psi \wedge \Phi \text{ folgt, und} \\ 0, & \text{wenn aus } \Sigma \text{ logisch } \neg\, \Phi \text{ folgt.} \end{cases}$$

---

[15] Diese bedingte Definition erfüllt nicht die Bedingungen, die Patrick Suppes in „Introduction to Logic", Princeton, N.J. ^{6}1963, S. 165, aufgestellt hat. Ich folgere daraus, daß diese Bedingungen den Begriff der bedingten Definition nicht adäquat charakterisieren.

Der Ausgang Σ ist erfüllbar. Da andererseits die Sätze $\Psi \wedge \Phi$, $\neg \Psi \wedge \Phi$ und $\neg \Phi$ miteinander logisch unverträglich sind und Σ einen von ihnen als Konjunktionsglied enthält, impliziert Σ nur diesen logisch. Daher ist der Funktionswert von gn bei vorgegebenen d und Σ eindeutig festgelegt. Die Ziffer „1" garantiert, daß die zu charakterisierende Funktion c normiert ist, also keine Werte annimmt, die größer als 1 sind; dies geschieht in Übereinstimmung mit D 1, D 2, D 3 und D 4.
Der Gesamtgewinn beim Ausgang Σ des Wettsystems D [gt (D, Σ)] ist die Summe der Gewinne bei den einzelnen Wetten aus D:

D 10 D sei ein mit c übereinstimmendes endliches Wettsystem; Σ sei ein Ausgang bezügl ch D. Dann ist
$\mathrm{gt}(D, \Sigma) = \sum_{d \in D} \mathrm{gn}(d, \Sigma)$.

Eine Funktion c wird streng kohärent genannt, wenn es kein mit ihr übereinstimmendes endliches Wettsystem gibt, das zwar einen Gewinn ausschließt, nicht jedoch einen Verlust, wenn also jedes mit c übe einstimmende endliche Wettsystem, bei dem ein Verlust möglich ist, auch einen Gewinn nicht ausschließt:

D 11 c ist streng kohärent genau dann, wenn es kein mit c übereinstimmendes endliches Wettsystem D gibt, so daß für jeden Ausgang Σ bezüglich D gilt: $\mathrm{gt}(D, \Sigma) \leqslant 0$, wobei es einen Ausgang Σ bezüglich D gibt, so daß: $\mathrm{gt}(D, \Sigma) < 0$.

Formal eleganter, wenn auch intuitiv weniger durchsichtig, kann der Begriff der strengen Kohärenz auf folgendem Weg eingeführt werden: In den Definitionen D_a, D_b, D_c und D_d werden die Begriffe „Wette", „Wettsystem", „Ausgang" und „normaler Ausgang" so charakterisier wie in D 5, D 6, D 7 und D 8. Sodann wird der Begriff der Gültigkeit bezüglich eines Ausgangs eingeführt; dabei ist „ag $(\Sigma, \Delta) = 1$" zu lesen als: „Bei dem Ausgang Σ gilt Δ (hat Δ den Gültigkeitswert 1)" und „ag $(\Sigma, \Delta) = 0$" als: „Bei dem Ausgang Σ gilt Δ nicht (hat Δ den Gültigkeitswert 0)".

D_e D sei ein mit c übereinstimmendes endliches Wettsystem; Σ sei ein Ausgang bezüglich D. Dann ist

$$\text{ag}(\Sigma, \Delta) = \begin{cases} 1, \text{ falls } \Delta \text{ aus } \Sigma \text{ logisch folgt, und} \\ 0, \text{ falls } \Delta \text{ aus } \Sigma \text{ nicht logisch folgt.} \end{cases}$$

Die Definition des Begriffs des Gewinnes bei einer Wette bezüglich eines Ausgangs erhält dann die folgende einfache Form:

D_f D sei ein mit c übereinstimmendes endliches Wettsystem, das die Wette d auf Ψ bezüglich Φ mit dem Wettquotienten r und dem Gesamteinsatz u enthält, und Σ sei ein Ausgang bezüglich D. Dann ist $\text{gn}(d, \Sigma) = \text{ag}(\Sigma, \Phi) \cdot [\text{ag}(\Sigma, \Psi) - r] \cdot u$.

Gilt also z. B. Φ beim Ausgang Σ nicht, so ist $\text{ag}(\Sigma, \Phi) = 0$ und damit $\text{gn}(d, \Sigma) = 0$. Gilt jedoch Φ bei Σ, so ist $\text{ag}(\Sigma, \Phi) = 1$ und daher $\text{gn}(d, \Sigma) = [\text{ag}(\Sigma, \Psi) - r] \cdot u$. Gilt Ψ bei Σ nicht, so sind $\text{ag}(\Sigma, \Psi) = 0$ und $\text{gn}(d, \Sigma) = -r \cdot u$, und gilt Ψ bei Σ, so sind $\text{ag}(\Sigma, \Psi) = 1$ und $\text{gn}(d, \Sigma) = [1 - r] \cdot u$. Die durch D_f charakterisierte Funktion ist also identisch mit der durch D 9 beschriebenen. Durch D_g und D_h werden sodann die Begriffe des Gesamtgewinns und der strengen Kohärenz wie in D 10 und D 11 eingeführt.
Unter Verwendung dieser Definitionen würden die Beweise der folgenden Lehrsätze mathematisch einfacher und eleganter werden. Die Charakterisierung durch D 5 bis D 11 zeichnet sich jedoch durch größere Verständlichkeit aus; da der Philosoph bei gleicher Exaktheit der größeren Anschaulichkeit den Vorzug gibt, wird diese im folgenden verwendet.
Auf diesem Weg kann der logische Begriff der strengen Kohärenz unter Verwendung des Ansatzes der subjektiven Wahrscheinlichkeitstheorie eingeführt werden. Er ist jedoch nicht von Personen und Zeitpunkten abhängig. Natürlich werden derartige streng kohärente Wahrscheinlichkeitsfunktionen nur von Personen benützt, die sich bei der Errechnung des Wahrscheinlichkeitswertes nicht von irrationalen Gefühlen beeinflussen lassen. Das gleiche gilt jedoch auch für die deduktive Logik: Deren Regeln werden ebenfalls nicht von allen Menschen beachtet, und oft genug äußern Personen widerspruchsvolle Ansichten. Dies ist jedoch kein Grund dafür, die deduktive

Logik auf Personen und Zeitpunkte zu relativieren, und dasselbe gilt für die induktive Logik.
Unter Verwendung der so definierten Begriffe kann folgendes bewiesen werden:

Lemma 1

Wenn c streng kohärent ist, dann gibt es kein mit c übereinstimmendes endliches Wettsystem D, so daß für jeden normalen Ausgang Σ bezüglich D gilt: $gt(D, \Sigma) < 0$.

Beweis

c sei streng kohärent. Angenommen, es gibt ein mit c übereinstimmendes endliches Wettsystem D, so daß für jeden normalen Ausgang Σ bezüglich D gilt: $gt(D, \Sigma) < 0$. Für ein solches Wettsystem D gilt dann auch, daß für alle Ausgänge[16] Σ bezügl ch D: $gt(D, \Sigma) \leqslant 0$, und für wenigstens einen Ausgang Σ bezüglich D: $gt(D, \Sigma) < 0$. Dies ist wegen D 11 jedoch gleichbedeutend damit, daß c nicht streng kohärent ist, was der Voraussetzung widerspricht; damit ist der Hilfssatz bewiesen.
Die Umkehrung dieser Behauptung gilt jedoch nicht: im Teil (6) des Beweises des Theorems 11 wird ein mit c übereinstimmendes Wettsystem angegeben, das nicht bei jedem normalen Ausgang einen Verlust garantiert und das trotzdem nicht streng kohärent ist. Die im Lemma 1 formulierte notwend ge Bedingung für d e strenge Kohärenz könnte mit der Kohärenz gleichgesetzt werden, die somit eine Abschwächung der strengen Kohärenz wäre. Damit dieser Kohärenzbegriff jedoch mit denen der Bestätigungsmethode und der Wahrscheinlichkeitsfunktion zusammenfällt, erweist es sich als notwendig, die Definition des Begriffs der Wette in D 5 abzuändern und an Stelle der Erfüllbarkeit von Φ zu fordern, daß $c(\Phi, T) \neq 0$.
Es kann nun gezeigt werden, daß strenge Kohärenz reguläre Wahr-

[16] Normale Ausgänge sind ja nach D 8 Ausgänge. Für den nicht normalen Ausgang Σ_0, der die Standardkonjunktion der Negationen aller Prämissen ist, gilt außerdem: $gt(D, \Sigma_0) = 0$.

scheinlichkeit nach sich zieht und umgekehrt. Also sind auch diese Begriffe umfangsgleich und daher identisch.[17]

Theorem 11

Wenn c streng kohärent ist, dann ist c eine reguläre Wahrscheinlichkeitsfunktion.

Beweis

c sei streng kohärent.
(1) Ψ sei ein Satz und Φ ein erfüllbarer Satz. Angenommen, $c(\Psi, \Phi) = r$, wobei $r < 0$; dann kann auf folgende Weise ein Wettsystem charakterisiert werden, das einen Verlust garantiert: u sei eine reelle Zahl mit $u < 0$. Dann sei D das Wettsystem, das nur die Wette d auf Ψ bezüglich Φ mit dem Wettquotienten r und dem Gesamteinsatz u enthält. Nach D 8 und D 7 gibt es zwei normale Ausgänge bezüglich D, nämlich $\Psi \wedge \Phi$ und $\neg \Psi \wedge \Phi$.
(α) Für den Ausgang $\Psi \wedge \Phi$ gilt nach D 10 und D 9:
$gt(D, \Psi \wedge \Phi) = \sum_{d \in D} gn(d, \Psi \wedge \Phi) = gn(d, \Psi \wedge \Phi) = (1 - r) \cdot u.$
Da $(1 - r) \cdot u < 0$, gilt: $gt(D, \Psi \wedge \Phi) < 0$.
(β) Für den Ausgang $\neg \Psi \wedge \Phi$ gilt:
$gt(D, \neg \Psi \wedge \Phi) = gn(d, \neg \Psi \wedge \Phi) = -r \cdot u.$
Da $-r \cdot u < 0$, gilt auch hier: $gt(D, \neg \Psi \wedge \Phi) < 0$.
D ist also ein mit c übereinstimmendes endliches Wettsystem, so daß für jeden normalen Ausgang Σ bezüglich D der Gesamtgewinn negativ ist. Nach Lemma 1 ist deshalb c nicht streng kohärent, was der Voraussetzung widerspricht. Also gilt: $c(\Psi, \Phi) \geqslant 0$.
(2) Φ sei ein erfüllbarer Satz, wobei $c(\Phi, \Phi) = r$ und $r \neq 1$.
(a) Angenommen, $r < 1$; dann möge D die einzige Wette d enthalten, wobei $d = \langle \Phi, \Phi, r, -1 \rangle$. Es gibt nur den normalen Ausgang $\Phi \wedge \Phi$; daher gilt:

[17] Die folgenden Lehrsätze sind erstmals von Bruno de Finetti formuliert und skizzenhaft bewiesen worden; vgl. die in Fußnote 13 dieses Kapitels genannten Aufsätze.

$gt(D, \Phi \wedge \Phi) = \sum_{d \in D} gn(d, \Phi \wedge \Phi) = gn(d, \Phi \wedge \Phi) = (1 - r)(-1)$
$= r - 1.$
Da $r < 1$, gilt: $r - 1 < 0$ und damit: $gt(D, \Phi \wedge \Phi) < 0$.
(b) Angenommen, $r > 1$; dann möge D nur die Wette d mit $d = \langle \Phi, \Phi, r, 1 \rangle$ enthalten. Da $\Phi \wedge \Phi$ der einzige normale Ausgang bezüglich D ist, gilt:
$gt(D, \Phi \wedge \Phi) = gn(d, \Phi \wedge \Phi) = (1 - r)\,1 = 1 - r.$
Nun ist $r > 1$ und daher $1 - r < 0$, also $gt(D, \Phi \wedge \Phi) < 0$.
Ist also $r \neq 1$, so gibt es ein Wettsystem, das bei jedem normalen Ausgang einen Verlust garantiert, was mit Lemma 1 der Voraussetzung über die strenge Kohärenz von c widerspricht. Also gilt: $c(\Phi, \Phi) = 1$.
(3) Ψ sei ein mit Δ logisch äquivalenter Satz und Φ ein mit Γ logisch äquivalenter erfüllbarer Satz. Angenommen, $c(\Psi, \Phi) = r$ und $c(\Delta, \Gamma) = s$, wobei $r \neq s$.
(a) Es sei $r < s$. Dann möge D das Wettsystem sein, das aus den zwei Wetten d_1 und d_2 besteht, wobei $d_1 = \langle \Psi, \Phi, r, -1 \rangle$ und $d_2 = \langle \Delta, \Gamma, s, 1 \rangle$.
Nach D 8 und D 7 gibt es also nur zwei normale Ausgänge Σ_α und Σ_β, wobei Σ_α den Satz $(\Delta \wedge \Gamma) \wedge (\Psi \wedge \Phi)$ und Σ_β den Satz $(\neg \Delta \wedge \Gamma) \wedge (\neg \Psi \wedge \Phi)$ abkürzen soll.
(α) Für den Ausgang Σ_α gilt nach D 10 und D 9:
$gt(D, \Sigma_\alpha) = \sum_{d \in D} gn(d, \Sigma_\alpha) = gn(d_1, \Sigma_\alpha) + gn(d_2, \Sigma_\alpha)$
$= (1 - r)(-1) + (1 - s)\,1 = r - s.$
Da $r < s$, gilt $r - s < 0$ und damit $gt(D, \Sigma_\alpha) < 0$.
(β) Für Σ_β gilt: $gt(D, \Sigma_\beta) = (-r)(-1) + (-s) \cdot 1 = r - s$.
Daher gilt auch hier: $gt(D, \Sigma_\beta) < 0$.

(b) Es sei $r > s$. D sei das Wettsystem, bestehend aus den zwei Wetten d_1 und d_2 mit $d_1 = \langle \Psi, \Phi, r, 1 \rangle$ und $d_2 = \langle \Delta, \Gamma, s, -1 \rangle$. Dann kann analog gezeigt werden, daß auch in diesem Fall für jeden normalen Ausgang der Gesamtgewinn negativ ist, nämlich $s - r$. Ist also $r \neq s$, dann ist c nach Lemma 1 nicht streng kohärent. Da dies der Voraussetzung widerspricht, gilt: $c(\Psi, \Phi) = c(\Delta, \Gamma)$.
(4) Φ sei ein erfüllbarer Satz und Ψ wie Δ seien Sätze, so daß Ψ, Δ und Φ nicht simultan erfüllbar sind. Angenommen, $c(\Psi \vee \Delta, \Phi) = r$, $c(\Psi, \Phi) = s$ und $c(\Delta, \Phi) = t$, wobei $r \neq s + t$.

(a) Es sei $r < s + t$. D sei das aus den Wetten d_1, d_2 und d_3 bestehende Wettsystem, wobei $d_1 = \langle \Psi \vee \Delta, \Phi, r, -1\rangle$, $d_2 = \langle \Psi, \Phi, s, 1\rangle$ und $d_3 = \langle \Delta, \Phi, t, 1\rangle$. Es gibt drei normale Ausgänge bezüglich D, nämlich:

$(\neg \Delta \wedge \Phi) \wedge ((\Psi \vee \Delta) \wedge \Phi) \wedge (\Psi \wedge \Phi)$ [Abkürzung: Σ_α]

$(\Delta \wedge \Phi) \wedge ((\Psi \vee \Delta) \wedge \Phi) \wedge (\neg \Psi \wedge \Phi)$ [Abkürzung: Σ_β]

$(\neg \Delta \wedge \Phi) \wedge (\neg (\Psi \vee \Delta) \wedge \Phi) \wedge (\neg \Psi \wedge \Phi)$ [Abkürzung: Σ_γ]

(α) Für Σ_α gilt:

$$gt(D, \Sigma_\alpha) = \sum_{d \in D} gn(d, \Sigma_\alpha) = gn(d_1, \Sigma_\alpha) + gn(d_2, \Sigma_\alpha) + gn(d_3, \Sigma_\alpha)$$
$$= (1 - r)(-1) + (1 - s)1 + (-t)1 = r - s - t.$$

Da $r < s + t$, gilt $r - s - t < 0$ und daher $gt(D, \Sigma_\alpha) < 0$.

(β) Für Σ_β gilt:

$gt(D, \Sigma_\beta) = (1 - r)(-1) + (-s)1 + (1 - t)1 = r - s - t.$

Auch hier ist also $gt(D, \Sigma_\beta) < 0$.

(γ) Für Σ_γ gilt:

$gt(D, \Sigma_\gamma) = (-r)(-1) + (-s) \cdot 1 + (-t) \cdot 1 = r - s - t.$

Somit ist auch hier $gt(D, \Sigma_\gamma) < 0$.

(b) Es sei $r > s + t$. D sei das aus den Wetten d_1, d_2 und d_3 bestehende Wettsystem, wobei $d_1 = \langle \Psi \vee \Delta, \Phi, r, 1\rangle$, $d_2 = \langle \Psi, \Phi, r, -1\rangle$ und $d_3 = \langle \Delta, \Phi, t, -1\rangle$. Analog kann gezeigt werden, daß auch in diesem Fall für jeden normalen Ausgang Σ bezüglich D der Gesamtgewinn negativ ist. Ist deshalb $r \neq s + t$, so gibt es ein mit c übereinstimmendes endliches Wettsystem, das einen Verlust garantiert. Da dies mit Lemma 1 der Annahme widerspricht, gilt:

$c(\Psi \vee \Delta, \Phi) = c(\Psi, \Phi) + c(\Delta, \Phi)$.

(5) Ψ, Δ und Φ seien Sätze, so daß $\Psi \wedge \Phi$ erfüllbar ist. Angenommen, $c(\Psi \wedge \Delta, \Phi) = r$, $c(\Psi, \Phi) = s$ und $c(\Delta, \Psi \wedge \Phi) = t$, wobei $r \neq s \cdot t$.

(a) Es sei $r < s \cdot t$. D sei das aus den Wetten d_1, d_2 und d_3 bestehende Wettsystem, wobei $d_1 = \langle \Psi \wedge \Delta, \Phi, r, -1\rangle$, $d_2 = \langle \Psi, \Phi, s, t\rangle$ und $d_3 = \langle \Delta, \Psi \wedge \Phi, t, 1\rangle$. Es gibt drei normale Ausgänge bezüglich D, nämlich:

$(\Delta \wedge (\Psi \wedge \Phi)) \wedge ((\Psi \wedge \Delta) \wedge \Phi) \wedge (\Psi \wedge \Phi)$ [Abkürzung: Σ_α]

$(\neg \Delta \wedge (\Psi \wedge \Phi)) \wedge (\neg (\Psi \wedge \Delta) \wedge \Phi) \wedge (\Psi \wedge \Phi)$ [Abkürzung: Σ_β]

$(\neg (\Psi \wedge \Phi)) \wedge (\neg (\Psi \wedge \Delta) \wedge \Phi) \wedge (\neg \Psi \wedge \Phi)$ [Abkürzung: Σ_γ]

(α) Für Σ_α gilt:

$gt(D, \Sigma_\alpha) = (1-r)(-1) + (1-s) \cdot t + (1-t) \cdot 1 = r - s \cdot t$.
Da $r < s \cdot t$, ist $r - s \cdot t < 0$ und daher $gt(D, \Sigma_\alpha) < 0$.
(β) Für Σ_β gilt:
$gt(D, \Sigma_\beta) = (-r)(-1) + (1-s) \cdot t + (-t)\,1 = r - s \cdot t$.
Also ist auch hier $gt(D, \Sigma_\beta) < 0$.
(γ) Für Σ_γ gilt:
$gt(D, \Sigma_\gamma) = (-r)(-1) + (-s) \cdot t + 0 = r - s \cdot t$.
Auch hier ist demnach $gt(D, \Sigma_\gamma) < 0$.
(b) Es sei $r > s \cdot t$. D möge dann aus den folgenden drei Wetten d_1, d_2 und d_3 bestehen: $d_1 = \langle \Psi \wedge \Delta, \Phi, r, 1\rangle$, $d_2 = \langle \Psi, \Phi, s, -t\rangle$, sowie $d_3 = \langle \Delta, \Psi \wedge \Phi, t, -1\rangle$. Auch hier ist der Gesamtgewinn bei jedem normalen Ausgang bezüglich D negativ.
Ist also $r \neq s \cdot t$, so ist c nach Lemma 1 nicht streng kohärent. Da dies der Voraussetzung widerspricht, gilt:
$c(\Psi \wedge \Delta, \Phi) = c(\Psi, \Phi) \cdot c(\Delta, \Psi \wedge \Phi)$.
(6) Angenommen, Ψ ist ein nicht logisch aus Φ folgender Satz; es sei $c(\Psi, \Phi) = 1$. D sei das Wettsystem, das nur die Wette d mit $d = \langle \Psi, \Phi, 1, 1\rangle$ enthält. Es gibt nach D 7 drei Ausgänge bezüglich D, nämlich $\neg \Phi$, $\Psi \wedge \Phi$ und $\neg \Psi \wedge \Phi$.
(α) Für $\neg \Phi$ gilt: $gt(D, \neg \Phi) = gn(d, \neg \Phi) = 0$.
(β) Für $\Psi \wedge \Phi$ gilt: $gt(D, \Psi \wedge \Phi) = (1-1)\,1 = 0$.
(γ) Für $\neg \Psi \wedge \Phi$ gilt: $gt(D, \neg \Psi \wedge \Phi) = -1 \cdot 1 = -1 < 0$.
Der Gesamtgewinn ist also bei keinem Ausgang bezüglich D positiv, aber bei wenigstens einem negativ. Nach D 11 ist c also nicht streng kohärent; da dies der Voraussetzung widerspricht, gilt: $c(\Psi, \Phi) \neq 1$. Damit ist Theorem 11 bewiesen.

D sei ein mit c übereinstimmendes endliches Wettsystem; c liefert also die Wettquotienten für die einzelnen Wetten aus D. Nun kann mittels c auf folgende Weise der voraussichtliche Gewinn aus D geschätzt werden: Es wird bei jedem vorgegebenen Ausgang Σ der Gesamtgewinn von D mit der Wahrscheinlichkeit dieses Ausgangs relativ auf das tautologische Wissen multipliziert, wobei diese Aprioriwahrscheinlichkeit mit eben jener Funktion c gemessen wird, die die Wettquotienten liefert; sodann wird die Summe dieser Produkte gebildet. Diese Summe ist die c-Mittel-Schätzung des Gewinns des mit c übereinstimmenden endlichen Wettsystems D [kurz: die $sg_c(D)$].

Auf ähnliche Weise kann etwa ein Gastwirt seinen voraussichtlichen Gewinn schätzen, der eine gewisse Anzahl von Mittagessen vorbereitet hat. Ein Ausgang ist bei ihm ein Satz, der beschreibt, welche Gerichte er verkauft und welche übrigbleiben. Er ermittelt seinen voraussichtlichen Gewinn, indem er zunächst den Gesamtgewinn bei den einzelnen Ausgängen mit der Wahrscheinlichkeit dieser Ausgänge multipliziert und sodann die Summe dieser Produkte bildet.

D 12 D sei ein mit c übereinstimmendes endliches Wettsystem. Dann ist $sg_c(D) = \sum_{\Sigma \in A_D} c(\Sigma, T) \cdot gt(D, \Sigma)$.

Wenn c die Funktion ist, die für die Wetten aus D die vernünftigen Wettquotienten liefern soll, so ist zu vermuten, daß der durch c geschätzte mittlere Gewinn aus diesem Wettsystem 0 sein wird, da es ja dann, unter diesem Wahrscheinlichkeitsbegriff, betrachtet, keinen der beiden Partner begünstigt. Dies kommt durch den folgenden Hilfssatz zum Ausdruck:

Lemma 2

Wenn c eine reguläre Wahrscheinlichkeitsfunktion ist, dann gilt für jedes mit c übereinstimmende endliche Wettsystem D: $sg_c(D) = 0$.

Beweis

c sei eine reguläre Wahrscheinlichkeitsfunktion. Wegen D 10 und D 9 gilt für ein beliebiges mit c übereinstimmendes endliches Wettsystem D:

$$\begin{aligned} sg_c(D) &= \sum_{\Sigma \in A_D} c(\Sigma, T) \cdot \sum_{d \in D} gn(d, \Sigma) \\ &= \sum_{\Sigma \in A_D} \sum_{d \in D} c(\Sigma, T) \cdot gn(d, \Sigma) \\ &= \sum_{d \in D} \sum_{\Sigma \in A_D} c(\Sigma, T) \cdot gn(d, \Sigma). \end{aligned}$$

Wird also definiert: $tg_c(d) = \sum_{\Sigma \in A_D} c(\Sigma, T) \cdot gn(d, \Sigma)$, so gilt: $sg_c(D) = \sum_{d \in D} tg_c(d)$.

d sei eine beliebige, aber für das folgende fest vorgegebene Wette aus D. Wegen D 5 ist $d = \langle \Psi, \Phi, r, u \rangle$ für gewisse Sätze Ψ und Φ

und gewisse reelle Zahlen r und u. Relativ auf dieses feste d kann die Klasse A_D der Ausgänge (bezüglich D) vollständig in die drei sich ausschließenden Mengen A_D^+, A_D^- und A_D^0 aufgeteilt werden, wobei A_D^+ die Klasse derjenigen Elemente aus A_D ist, die $\Psi \wedge \Phi$ logisch implizieren, A_D^- hingegen die Gesamtheit jener Ausgänge bezüglich D, die $\neg \Psi \wedge \Phi$ logisch implizieren und A_D^0 schließlich die Menge jener Elemente aus A_D, aus denen $\neg \Phi$ logisch folgt. Dann gilt:

$$tg_c(d) = \sum_{\Sigma \in A_D^+} c(\Sigma, T) \cdot gn(d, \Sigma) + \sum_{\Sigma \in A_D^-} c(\Sigma, T) \cdot gn(d, \Sigma) + \sum_{\Sigma \in A_D^0} c(\Sigma, T) \cdot gn(d, \Sigma).$$

Wegen D 9 gilt weiterhin:

$$tg_c(d) = \sum_{\Sigma \in A_D^+} c(\Sigma, T)(1-r) \cdot u + \sum_{\Sigma \in A_D^-} c(\Sigma, T)(-r) \cdot u + \sum_{\Sigma \in A_D^0} c(\Sigma, T) \cdot 0$$

$$= \sum_{\Sigma \in A_D^+} c(\Sigma, T)(1-r) \cdot u + \sum_{\Sigma \in A_D^-} c(\Sigma, T)(-r) \cdot u$$

Da D ein mit c übereinstimmendes endliches Wettsystem und d eine Wette aus D mit dem Wettquotienten r ist, gilt wegen D 6: $c(\Psi, \Phi) = r$, und weiterhin:

$$tg_c(d) = \sum_{\Sigma \in A_D^+} c(\Sigma, T)(1 - c(\Psi, \Phi)) \cdot u + \sum_{\Sigma \in A_D^-} c(\Sigma, T) \cdot (-c(\Psi, \Phi)) \cdot u$$

$$= u \cdot [(1 - c(\Psi, \Phi)) \cdot \sum_{\Sigma \in A_D^+} c(\Sigma, T) - c(\Psi, \Phi) \cdot \sum_{\Sigma \in A_D^-} c(\Sigma, T)].$$

Keine zwei voneinander verschiedenen Ausgänge Σ_1 und Σ_2 bezüglich D sind simultan erfüllbar. Sind daher $\Sigma_{i_1}, \ldots, \Sigma_{i_k}$ die Elemente von A_D^+, so erhält man durch k—1-malige Benützung der Bedingung (4) aus D 4:

$$\sum_{\Sigma \in A_D^+} c(\Sigma, T) = c(\Sigma_{i_1} \vee \ldots \vee \Sigma_{i_k}, T).$$

Sind $\Sigma_{j_1}, \ldots, \Sigma_{j_l}$ die Elemente von A_D^-, so erhält man durch l—1-malige Benützung dieser Bedingung (4):

$$\sum_{\Sigma \in A_D^-} c(\Sigma, T) = c(\Sigma_{j_1} \vee \ldots \vee \Sigma_{j_l}, T).$$

Da die in D 7 definierte Klasse D erschöpfend ist, ist bei jeder Interpretation wenigstens (und damit genau) ein Ausgang bezüglich D wahr. Dann ist die Adjunktion aller Ausgänge, die einen gewissen Satz logisch implizieren, logisch äquivalent mit diesem Satz. Daher

ist $\Sigma_{i_1} \vee \ldots \vee \Sigma_{i_k}$ logisch äquivalent mit $\Psi \wedge \Phi$ und $\Sigma_{j_1} \vee \ldots \vee \Sigma_{j_l}$ mit $\neg \Psi \wedge \Phi$. Wegen Bedingung (3) von D 4 gilt daher:
$c(\Sigma_{i_1} \vee \ldots \vee \Sigma_{i_k}, T) = c(\Psi \wedge \Phi, T)$, und
$c(\Sigma_{j_1} \vee \ldots \vee \Sigma_{j_l}, T) = c(\neg \Psi \wedge \Phi, T)$. Folglich gilt auch:
$\sum_{\Sigma \in A_D^+} c(\Sigma, T) = c(\Psi \wedge \Phi, T)$ und $\sum_{\Sigma \in A_D^-} c(\Sigma, T) = c(\neg \Psi \wedge \Phi, T)$.
Aus diesen Gleichungen erhält man:

$$\begin{aligned} tg_c(d) &= u \cdot [(1 - c(\Psi, \Phi)) \cdot c(\Psi \wedge \Phi, T) \\ &\quad - c(\Psi, \Phi) \cdot c(\neg \Psi \wedge \Phi, T)] \\ &= u \cdot [c(\Psi \wedge \Phi, T) - [c(\Psi \wedge \Phi, T) \\ &\quad + c(\neg \Psi \wedge \Phi, T)] \cdot c(\Psi, \Phi)]. \end{aligned}$$

Mit Theorem 1 (i) folgt daraus:
$tg_c(d) = u \cdot [c(\Psi \wedge \Phi, T) - c(\Phi, T) \cdot c(\Psi, \Phi)]$.
Da Φ von den Ausgängen aus A_D^+ wie aus A_D^- logisch impliziert wird und Ausgänge erfüllbar sind, ist auch Φ erfüllbar.
Wegen Theorem 2 (d) gilt daher: $tg_c(d) = 0$.
Da d eine beliebige Wette aus D ist, gilt: $\sum_{d \in D} tg_c(d) = 0$, und daher: $sg_c(D) = 0$.

Theorem 12

Wenn c eine reguläre Wahrscheinlichkeitsfunktion ist, dann ist c streng kohärent.

Beweis

c sei eine reguläre Wahrscheinlichkeitsfunktion. Angenommen, c ist nicht streng kohärent; dann gibt es nach D 11 ein mit c übereinstimmendes endliches Wettsystem D, so daß für jedes Σ aus A_D: $gt(D, \Sigma) \leqslant 0$, und für wenigstens ein Σ aus A_D: $gt(D, \Sigma) < 0$. Ein solches Σ sei festgehalten. Es ist wegen D 7 erfüllbar, weshalb aus Theorem 2 (b) folgt: $c(\Sigma, T) > 0$. Also gilt für dieses Σ: $c(\Sigma, T) \cdot gt(D, \Sigma) < 0$. Da für jedes andere Σ aus A_D stets $c(\Sigma, T) > 0$ und $gt(D, \Sigma) \leqslant 0$, gilt für jedes derartige Σ: $c(\Sigma, T) \cdot gt(D, \Sigma) \leqslant 0$. Somit ist nach D 12: $sg_c(D) < 0$. Andererseits folgt mit Lemma 2: $sg_c(D) = 0$. Also führt die Annahme, daß c nicht streng kohärent ist, zu einem Widerspruch.

Korollar 2

c ist streng kohärent genau dann, wenn c eine reguläre Wahrscheinlichkeitsfunktion ist.

Korollar 3

c ist streng kohärent genau dann, wenn c eine reguläre Bestätigungsmethode ist.

4. Die logische Stärke der Wahrscheinlichkeitsaxiome

Da die verschiedensten Ansätze zur Explikation des Begriffs der *Berechenbarkeit* miteinander logisch äquivalent sind, sieht man die Vermutung, daß jeder dieser formalen Charakterisierungen den intuitiven Berechenbarkeitsbegriff adäquat wiedergibt, als begründet an. Analog könnte man schließen, daß die verschiedenen, miteinander logisch äquivalenten Ansätze zur Charakterisierung des Begriffs der *Wahrscheinlichkeit* zeigen, daß dieser intuitive Begriff damit formal befriedigend beschrieben worden ist. Daß dieser Schluß unberechtigt ist, sollen die folgenden Überlegungen zeigen.
Die Axiome (1) bis (6), durch die in D 4 der Begriff der regulären Wahrscheinlichkeitsfunktion definiert worden ist, legen zwar c als eine Funktion fest, die nur Werte zwischen 0 und 1 annimmt, und gestatten auch die Berechnung einzelner Wahrscheinlichkeitswerte, nämlich dann, wenn die Prämisse die Hypothese logisch impliziert oder mit ihr logisch unverträglich ist, so daß c dann die Werte 1 bzw. 0 annimmt. Doch abgesehen von diesen trivialen Werten kann unter ihrer Verwendung allein kein einziger Wahrscheinlichkeitswert ermittelt werden. Dies ist um so erstaunlicher, als in einer Reihe von Lehrbüchern der Wahrscheinlichkeitstheorie nur diese Bedingungen als Axiome gefordert werden. Meist werden bei den Berechnungen allerdings stillschweigend das noch zu charakterisierende Indifferenzprinzip oder aber andere Voraussetzungen verwendet, die Aprioriaussagen über das Verhältnis der Begriffsumfänge der einzelnen Prä-

dikate machen. Jene Bedingungen schließen, für sich genommen, jedoch noch nicht einmal den Fall aus, daß mit wachsender Anzahl der für die Hypothese positiv relevanten Beobachtungen der c-Wert immer kleiner wird.

Ähnliches gilt für den Begriff der strengen Kohärenz; er ist wesentlich schwächer als der der Fairneß, so wie dieser intuitiv verstanden wird. Man wird etwa die folgende Wette als unfair empfinden: Darauf, daß bei einem Wurf mit einem normalen Würfel die Ziffer „4" oben liegt, wird mit einem Quotienten von 0,9 gewettet; die Chancen, zu verlieren, sind hier wesentlich größer als die, zu gewinnen. Andererseits schließt ein Wettsystem, das nur diese eine Wette enthält, weder einen Gewinn noch einen Verlust aus: Es gibt Ausgänge, bei denen man (wenn auch relativ wenig) gewinnt, und es gibt Ausgänge, bei denen man verliert. Desgleichen ist etwa eine Wette mit einer Lotteriegesellschaft oder mit einer anderen Körperschaft, die Glücksspiele anbietet, zwar streng kohärent, da nicht nur der Verlust des eigenen Einsatzes, sondern auch ein Gewinn möglich ist. Sie kann jedoch nicht als fair angesehen werden; denn die Schätzung des mittleren Gewinns der Gesellschaft ist positiv und die des eigenen daher negativ, d. h. das Produkt des eigenen Einsatzes mit der Wahrscheinlichkeit dafür, daß das Los eine Niete ist, ist größer als das Produkt aus Reingewinn (Gewinn abzüglich Einsatz) und der Wahrscheinlichkeit dafür, daß das Los ein Treffer ist.

Wie wenig determiniert die Begriffe der regulären Bestätigungsmethode sowie der strengen Kohärenz und der regulären Wahrscheinlichkeitsfunktion sind, zeigt der folgende, erstmals von Carnap bewiesene Lehrsatz:[18]

Theorem 13

Ψ und Φ seien Sätze von der Art, daß sowohl Ψ und Φ als auch $\neg \Psi$ und Φ simultan erfüllbar sind. r sei eine beliebige reelle Zahl, so daß $0 < r < 1$. Dann gibt es (wenigstens) eine reguläre Bestätigungsmethode c (es gibt sogar unendlich viele), so daß $c(\Psi, \Phi) = r$.

---

[18] Vgl. Carnap [9], S. 321ff., sowie Carnap-Stegmüller [1], S. 165f.

Beweis

Die Voraussetzungen des Lehrsatzes seien erfüllt.
Die Sätze $\neg \Phi$, $\Psi \wedge \Phi$ und $\neg \Psi \wedge \Phi$ sind logisch erschöpfend und paarweise miteinander logisch unverträglich. Jedes Σ aus Zb gehört also entweder zu $Zb^{\neg \Phi}$ oder zu $Zb^{\Psi \wedge \Phi}$ oder zu $Zb^{\neg \Psi \wedge \Phi}$. Daher ist wegen D 2 und D 1 für eine reguläre Bestätigungsmethode c:
$c(\neg \Phi, T) + c(\Psi \wedge \Phi, T) + c(\neg \Psi \wedge \Phi, T) = 1$.
Außerdem ist Φ logisch äquivalent mit $(\Psi \wedge \Phi) \vee (\neg \Psi \wedge \Phi)$, so daß:
$c(\Phi, T) = c(\Psi \wedge \Phi, T) + c(\neg \Psi \wedge \Phi, T)$.
Für jedes derartige c gilt dann:

$$
\begin{aligned}
c(\Psi, \Phi) &= \frac{c(\Psi \wedge \Phi, T)}{c(\Phi, T)} \\
&= \frac{c(\Psi \wedge \Phi, T)}{c((\Psi \wedge \Phi) \vee (\neg \Psi \wedge \Phi), T)} \\
&= \frac{c(\Psi \wedge \Phi, T)}{c(\Psi \wedge \Phi, T) + c(\neg \Psi \wedge \Phi, T)} \\
&= \frac{\sum\limits_{\Sigma \in Zb^{\Psi \wedge \Phi}} c(\Sigma, T)}{\sum\limits_{\Sigma \in Zb^{\Psi \wedge \Phi}} c(\Sigma, T) + \sum\limits_{\Sigma \in Zb^{\neg \Psi \wedge \Phi}} c(\Sigma, T)}
\end{aligned}
$$

s sei irgendeine reelle Zahl, so daß $0 < s < 1$. Den Zustandsbeschreibungen aus $Zb^{\Psi \wedge \Phi}$ werden nun auf willkürliche Weise (z. B. in gleichen Beträgen) positive reelle Zahlen als reguläre Maße zugeordnet, so daß $\sum\limits_{\Sigma \in Zb^{\Psi \wedge \Phi}} c(\Sigma, T) = r \cdot (1 - s)$, und den Σ aus $Zb^{\neg \Psi \wedge \Phi}$ willkürlich (z. B. wiederum in gleichen Beträgen) positive reelle Zahlen als reguläre Maße, so daß $\sum\limits_{\Sigma \in Zb^{\neg \Psi \wedge \Phi}} c(\Sigma, T) = (1 - r) \cdot (1 - s)$.
Der Vollständigkeit halber werden auch den Σ aus $Zb^{\neg \Phi}$ willkürlich positive reelle Zahlen als reguläre Maße zugeordnet, so daß
$\sum\limits_{\Sigma \in Zb^{\neg \Phi}} c(\Sigma, T) = s$.
Dann ist c ein reguläres Maß für alle Zustandsbeschreibungen und damit wegen D 2 auch für alle Sätze. Es gilt mit D 2 und D 3:

$$
c(\Psi, \Phi) = \frac{r \cdot (1 - s)}{r \cdot (1 - s) + (1 - r) \cdot (1 - s)} = r .
$$

Dieser Lehrsatz zeigt, wie schwach die Begriffe der regulären Bestätigung methode sowie die der strengen Kohärenz und der regulären Wahrscheinlichkeitsfunktion sind. Mit einer beliebigen Verteilung von positiven reellen Zahlen mit dem Gesamtbetrag 1 auf die Σ aus Zb können wegen D 2 und D 3 eine Reihe von Funktionen als reguläre Bestätigungsmethoden erwiesen werden, die man von einem intuitiven Standpunkt aus nicht als adäquate Explikationen der Begriffe der Bestätigung und der logischen Wahrscheinlichkeit ansehen würde.

Besagt z. B. Φ, daß von 10^6 Dingen eines Bereichs alle eine gewisse Eigenschaft haben, und Ψ, daß ein neues Objekt ebenfalls diese Eigenschaft besitzt, so würde man von einer adäquaten Methode der induktiven Bestätigung mit Recht erwarten, daß der Bestätigungswert von Ψ durch Φ nur wenig kleiner als 1 ist. Der obige Lehrsatz zeigt jedoch, daß es auch eine reguläre Bestätigungsmethode c gibt, die für Ψ bezüglich Φ etwa den Wert $1/10^{10}$ liefert. $\neg\, \Psi$ erhält dann bezüglich Φ den c-Wert $1 - (1/10^{10})$.

Die Gesamtheit der regulären Bestätigungsmethoden enthält also Funktionen, die man nicht als adäquate Explikationen der logischen Wahrscheinlichkeit ansehen kann. Es erhebt sich daher die Frage, welche weiteren Adäquatheitsbedingungen den Wahrscheinlichkeitsbegriff einengen und derartige unerwünschte Funktionen ausschließen.

IV. Das Kontinuum der induktiven Methoden

1. Adäquate induktive Methoden

Es stellt sich daher die Frage, welchen Bedingungen eine adäquate induktive Methode genügen muß. Sicherlich wird es sich dabei um eine reguläre Bestätigungsmethode handeln, doch muß die Klasse dieser Methoden auf die Menge derjenigen Wahrscheinlichkeitsfunktionen eingeschränkt werden, die wir auch intuitiv als adäquat empfinden; dabei soll zunächst die Frage offenbleiben, ob es genau eine oder ob es mehrere von ihnen gibt. Unter einer induktiven Methode wird hier, wie schon erwähnt, nicht ein Verfahren verstanden, das von Erfahrungsdaten mechanisch zu wissenschaftlichen Hypothesen führt; vielmehr handelt es sich dabei um eine Funktion, durch die bewertet wird, in welchem Grad eine bereits vorgegebene Hypothese durch ein derartiges Datum gestützt wird.

Die bekannteste zusätzliche Forderung an den Wahrscheinlichkeitsbegriff bildet das klassische *Indifferenzprinzip*, auch *Satz von unzureichenden Grund* genannt; danach sind zwei Ereignisse (bzw. zwei Aussagen, die diese Ereignisse beschreiben) als gleichwahrscheinlich anzusetzen, wenn sie gleichmöglich sind, wenn also kein Grund dafür vorhanden ist, daß ihre Wahrscheinlichkeitswerte verschieden sind. Diese sicherlich vage Formulierung hat zu vielen Mißverständnissen Anlaß gegeben und auch zu Widersprüchen geführt,[1] so daß dieses Prinzip von vielen Wahrscheinlichkeitstheoretikern verworfen worden ist. Andererseits scheint es klar zu sein, daß es einen widerspruchsfreien und gültigen Kern besitzen muß. Denn für eine Urne mit verschiedenfarbigen Kugeln etwa wird zur Beurteilung der Wahrscheinlichkeit dafür, daß die nächste gezogene Kugel gelb ist, lediglich die *Anzahl* der bisher beobachteten Kugeln mit einer bestimmten Farbe maßgeblich sein, nicht jedoch die Frage, um welche Kugeln es sich dabei gehandelt hat; wäre also die Kugel a_2 an Stelle von a_1 gelb ge-

[1] Vgl. z. B. Kneale [1], S. 147ff., 172f. und 204.

wesen, so hätte dies keinen anderen Wahrscheinlichkeitswert ergeben. Ferner darf man der Hypothese, daß von den 30 Kugeln 10 rot, 5 gelb und die übrigen zu gleichen Teilen grün, blau und violett sind, nicht eine höhere Anfangswahrscheinlichkeit zuordnen als der Hypothese, daß 10 Kugeln gelb, 5 rot und jeweils 5 grün, blau und violett sind, und umgekehrt; eine derartige unterschiedliche Bewertung würde ja besagen, daß man eine Hypothese deshalb a priori für wahrscheinlicher hält als eine andere, weil einem etwa die Farbe rot sympathisch ist, was allen Rationalitätsforderungen widerspräche. Erst Beobachtungen können entscheiden, welche der beiden Hypothesen wahrscheinlicher ist. Dieser widerspruchsfreie Kern des Indifferenzprinzips ist von Carnap „Symmetriebedingung" genannt worden.

Es werde die Hypothese aufgestellt, daß als nächstes eine rote Kugel gezogen wird; dann ist für die Frage, welchen Wahrscheinlichkeitswert sie im Hinblick auf eine beliebige Stichprobe erhält, nur das Verhältnis der beobachteten roten zu den beobachteten nichtroten Kugeln a priori relevant; dies ist gleichbedeutend damit, daß a priori nur das Verhältnis der beobachteten roten Kugeln zu den beobachteten Kugeln insgesamt berücksichtigt wird. A priori irrelevant ist hingegen etwa die Feststellung, daß von den beobachteten nichtroten Kugeln über die Hälfte grün, die restlichen Kugeln jedoch ungefähr zu gleichen Anzahlen gelb, blau und violett sind. Welche Erfahrungsdaten außerdem noch irrelevant sind, kann hingegen nur durch Erfahrung, also a posteriori, entschieden werden.

Eine adäquate induktive Methode soll es uns schließlich ermöglichen, aus der Erfahrung zu lernen. Damit ist gemeint, daß die Wahrscheinlichkeit dafür, daß eine rote Kugel gezogen wird, wächst, falls die vergangene Ziehung *rot* ergeben hat. Eine für die Hypothese positive Ziehung muß also für die Wahrscheinlichkeitsbewertung der Hypothese positiv relevant sein. Diese Bedingung ist teilweise, jedoch nicht ganz, in der der Symmetrie enthalten.

Alle diese Bedingungen sind notwendig dafür, daß eine induktive Methode c adäquat ist. Versuchsweise wird angenommen, daß sie auch hinreichend dafür sind, zumindest für eine induktive Methode, die für eine über einem endlichen Gegenstandsbereich interpretierte Sprache charakterisiert wird. Dann kann definiert werden:

D 13 c ist eine adäquate induktive Methode (für Sprachen mit endlich vielen Gegenstandsausdrücken und endlich vielen einstelligen Grundprädikaten, deren Anzahl größer als 1 ist) genau dann, wenn

(I) c eine reguläre Bestätigungsmethode ist,

(II) c symmetrisch bezüglich der außerlogischen Konstanten ist,

(III) durch c nur der für die Hypothese a priori relevante Teil der Prämisse berücksichtigt wird, und

(IV) c es ermöglicht, aus der Erfahrung zu lernen.

Um diesem Begriff der adäquaten induktiven Methode eine klare Bedeutung zu geben, ist es notwendig, die Bedingungen (II), (III) und (IV) zu analysieren. Es wird dabei vorausgesetzt, daß $\Pi_1, \ldots, \Pi_p$ einstellige Beziehungsausdrücke sind, also unter der Interpretation J Eigenschaften bzw. Klassen bezeichnen. Diese Ausdrücke seien voneinander unabhängig, so daß also keiner unter Verwendung der anderen definierbar ist. (Definierbare Ausdrücke seien also bereits auf die übliche Art eliminiert; wenn Bedeutungsabhängigkeiten bestehen sollten, die nicht durch Definitionen, sondern nur durch Bedeutungspostulate ausdrückbar sind, so ist diesen durch eine der drei am Schluß des 1. Abschnittes des III. Kapitels angedeuteten Verfahren Rechnung zu tragen, am besten wohl durch das dritte.) Die Interpretation J bildet die Gegenstandsausdrücke $\alpha_1, \ldots, \alpha_n$ umkehrbar eindeutig auf den Gegenstandsbereich B ab. Damit in der Sprache überhaupt Sätze gebildet werden können, muß schließlich noch vorausgesetzt werden, daß $n > 0$ und $p > 0$.

In den nächsten beiden Abschnitten wird allerdings zusätzlich angenommen, daß $p > 1$. Diese Voraussetzung ist nur eine *theoretische* Einschränkung; sie ist in so gut wie allen *praktischen* Fällen erfüllt. Wenn ich etwa die relative Häufigkeit der weißen Kugeln in der Gesamtheit der sich in der Urne d befindenden Kugeln schätze, so verwende ich dabei die (voneinander logisch unabhängigen) Eigenschaften bzw. Klassen *weiß* und *Kugel in der Urne d*. Da es sich erweist, daß für $p = 1$ die Bedingung (III) aus (I) logisch folgt und daher ohne faktische Relevanz ist, wird im 3. Abschnitt für diesen Fall eine zusätzliche Adäquatheitsbedingung formuliert, die es gestattet, die glei-

chen Ergebnisse herzuleiten. Wie schon für die Ausführungen des letzten Kapitels wird auch für die der beiden folgenden Abschnitte vorausgesetzt, daß n und p endliche Anzahlen, also natürliche Zahlen, sind. Auch hier werden im 3. Abschnitt Hinweise zur Verallgemeinerung der Resultate für unendliche Bereiche gegeben.

Häufig haben die Dinge des Gegenstandsbereichs B eine und nur eine der Eigenschaften, die durch die Prädikate $\Pi_1, \ldots, \Pi_p$ bezeichnet werden; dies ist z. B. dann der Fall, wenn $p = 6$, Π_1 die Eigenschaft *rot*, Π_2 *orange*, Π_3 *gelb*, Π_4 *grün*, Π_5 *blau* und Π_6 *violett* bezeichnet, sowie wenn kein Ding aus B farblos und jedes einfarbig ist. In diesem Fall nennt man jene Prädikate eine *Division* oder eine *Familie*. Eine andere Division bilden jene Prädikate, die die 6 Seiten eines Würfels oder die zwei Seiten einer Münze bezeichnen. Bezüglich derartiger Familien von Prädikaten lassen sich die Bedingungen (II), (III) und (IV) besonders einfach analysieren.

Dennoch wird im folgenden auf diese Einschränkung verzichtet. In sehr vielen Fällen benützt man nämlich den Wahrscheinlichkeitsbegriff auch bezüglich solcher Prädikate, die sich nicht ausschließen, etwa wenn man die relative Häufigkeit der Brillenträger unter der Landbevölkerung oder die der Akademiker unter den Fabrikbesitzern schätzt; solche Überlegungen werden hinfällig, wenn sich die entsprechenden Eigenschaften logisch ausschließen. Andererseits benötigt man derartige Prädikatfamilien zur Analyse einiger der Bedingungen, insbesondere aber von (II).

Wenn nun die einzelnen Kugeln in jener Urne mehrfarbig sind, so kann es sein, daß bestimmte Farben häufig, andere selten zusammen vorkommen. Dann kann zur Beurteilung der Wahrscheinlichkeit dafür, daß die nächste Kugel (zumindest teilweise) rot ist, relevant sein, ob einige der beobachteten Kugeln grün sind, nämlich dann, wenn grün entweder selten oder aber häufig mit rot zusammen erscheint.

Es erweist sich dann als zweckmäßig, von diesen Grundprädikaten zu gewissen definierbaren Prädikaten überzugehen, die eine Familie bilden; als solche bieten sich Ausdrücke an, die durch Sätze der Art „a ist rot, aber nicht orange, nicht gelb und nicht grün, jedoch blau und violett" charakterisiert werden. Solche Sätze beschreiben vollständig die in der Sprache ausdrückbaren Eigenschaften, die Eigenschaften,

die ein Gegenstand möglicherweise haben kann. Zu jenen Farbprädikaten gibt es 64 ($= 2^6$) derartige Sätze, wenn diejenigen außer acht gelassen werden, die durch Vertauschung der Konjunktionsglieder aus einem vorgegebenen Satz entstehen. Allgemein gilt, daß es bei p Grundprädikaten 2^p ($= q$) Sätze dieser Art gibt, die einen Gegenstand durch die in der Sprache ausdrückbaren Eigenschaften vollständig beschreiben. Ein solcher Satz wird im folgenden ein V-Satz [vollständig (beschreibender) Satz] und das durch ihn definierbare Prädikat ein V-Prädikat genannt.[2] Da nach Voraussetzung alle Grundprädikate Eigenschaften bezeichnen und jeder Basissatz deshalb genau einen Gegenstandsausdruck enthält, können diese Begriffe folgendermaßen definiert werden:

D 14 Φ ist ein V-Satz in α genau dann, wenn Φ die Standardkonjunktion einer maximalkonsistenten Klasse von Basissätzen ist, die alle als Objektausdruck α enthalten.

D 15 Π^* ist ein V-Prädikat genau dann, wenn es einen V-Satz Φ in einem Objektausdruck α gibt, so daß $\alpha \in \Pi^*$ eine Abkürzung für Φ ist.

Ist $p = 3$ (und daher $q = 8$), so erhält man demnach die folgenden V-Sätze in α und die folgenden durch sie definierbaren V-Prädikate:

$\alpha \in \Pi_1 \wedge \alpha \in \Pi_2 \wedge \alpha \in \Pi_3$	[Abkürzung: $\alpha \in \Pi^*_1$]
$\alpha \in \Pi_1 \wedge \alpha \in \Pi_2 \wedge \neg (\alpha \in \Pi_3)$	[Abkürzung: $\alpha \in \Pi^*_2$]
$\alpha \in \Pi_1 \wedge \neg (\alpha \in \Pi_2) \wedge \alpha \in \Pi_3$	[Abkürzung: $\alpha \in \Pi^*_3$]
$\alpha \in \Pi_1 \wedge \neg (\alpha \in \Pi_2) \wedge \neg (\alpha \in \Pi_3)$	[Abkürzung: $\alpha \in \Pi^*_4$]
$\neg (\alpha \in \Pi_1) \wedge \alpha \in \Pi_2 \wedge \alpha \in \Pi_3$	[Abkürzung: $\alpha \in \Pi^*_5$]
$\neg (\alpha \in \Pi_1) \wedge \alpha \in \Pi_2 \wedge \neg (\alpha \in \Pi_3)$	[Abkürzung: $\alpha \in \Pi^*_6$]
$\neg (\alpha \in \Pi_1) \wedge \neg (\alpha \in \Pi_2) \wedge \alpha \in \Pi_3$	[Abkürzung: $\alpha \in \Pi^*_7$]
$\neg (\alpha \in \Pi_1) \wedge \neg (\alpha \in \Pi_2) \wedge \neg (\alpha \in \Pi_3)$	[Abkürzung: $\alpha \in \Pi^*_8$]

Δ_1 und Δ_2 seien zwei beliebige voneinander verschiedene V-Sätze in α; dann sind Δ_1 und Δ_2 nicht simultan erfüllbar. Eine Adjunktion

[2] Dies geschieht in enger Anlehnung an die Terminologie von Carnap; vgl. Carnap [9], S. 123ff. Carnap nennt solche Prädikate „Q-Prädikate".

aller V-Sätze in α ist darüber hinaus logisch wahr. Also ist eine Adjunktion der Klasse jener V-Sätze in α, die einen Satz Δ logisch implizieren, der an Gegenstandsausdrücken nur α enthält, logisch äquivalent mit Δ. Dies legt die folgenden Definitionen nahe:

D 16 (a) Die Weite eines Satzes Δ, der an Gegenstandsausdrücken nur α enthält (und damit die Weite eines durch diesen Satz definierbaren Prädikates Π^+), ist die Anzahl der V-Sätze in α, deren Adjunktion logisch äquivalent mit Δ ist.

(b) Die relative Weite eines derartigen Satzes Δ (und damit auch die relative Weite von Π^+) ist m/q, wobei m die Weite von Δ und q die Anzahl aller V-Sätze in α ist.

Die Forderung der Symmetrie erschöpft sich für die vorgegebene einfache Sprache, die an außerlogischen Zeichen nur Gegenstands- und Eigenschaftsausdrücke enthält, in den beiden folgenden Bedingungen: Wenn sich zwei Prämissen und zwei Hypothesen nur dadurch voneinander unterscheiden, daß sie das gleiche von verschiedenen Gegenständen aussagen, so haben beide Hypothesen auf Grund der jeweiligen Prämissen die gleiche Wahrscheinlichkeit; ein Gegenstand wird also einem anderen gegenüber erst bevorzugt, wenn man mehr von ihm weiß als von diesem (wenn die eine Prämisse von ihm mehr aussagt als die andere von diesem) und nicht schon, weil er von diesem verschieden ist. Nur die Art und Weise, wie die Hypothese mit der Prämisse logisch verknüpft ist, ist relevant. Bilden die Grundprädikate eine Familie, so ändert sich die Aprioriwahrscheinlichkeit eines Satzes Φ nicht, wenn darin zwei Prädikate an allen Stellen, an denen sie darin vorkommen, durch einander ersetzt werden. Bilden sie keine Familie, so ist Φ mit einem Satz Φ^* logisch äquivalent, der nur V-Prädikate enthält, die ihrerseits eine Division bilden; dann ändert sich die Wahrscheinlichkeit nicht, wenn in Φ^* zwei V-Prädikate an allen Stellen, an denen sie darin vorkommen, durch einander ersetzt werden. Der Begriff der Symmetrie kann demnach wie folgt analysiert werden:

D 17 c ist symmetrisch bezüglich der außerlogischen Konstanten genau dann, wenn

(a) für alle Sätze Ψ und Δ gilt: Wenn Δ aus Ψ dadurch entsteht, daß zwei Gegenstandsausdrücke darin an allen Stellen durch einander ersetzt werden, so ist $c(\Psi, T) = c(\Delta, T)$, und

(b) für alle Sätze Ψ und Δ gilt: Wenn Ψ an Prädikaten nur V-Prädikate enthält, und zwei V-Prädikate darin durch einander ersetzt werden, so daß der Satz Δ entsteht, so ist $c(\Psi, T) = c(\Delta, T)$.

Wegen D 3 überträgt sich die Symmetrie der Maßfunktion c auch auf die bedingte Wahrscheinlichkeitsfunktion c. Φ sei z. B. ein Satz, in dem die Gegenstandsausdrücke α und β nicht vorkommen. Dann gilt für eine symmetrische reguläre Bestätigungsmethode c: $c(\alpha \in \Pi^* \wedge \Phi, T) = c(\beta \in \Pi^* \wedge \Phi, T)$, sowie: $c(\Phi, T) = c(\Phi, T)$, und daher: $c(\alpha \in \Pi^*, \Phi) = c(\beta \in \Pi^*, \Phi)$. Dieses Verfahren des Austauschens von Gegenstandsausdrücken kann iteriert werden; daher ändert sich für ein symmetrisches c der Wert nicht, wenn eine beliebige Anzahl solcher Ausdrücke permutiert wird. Das Analoge gilt für V-Prädikate.

Auf den ersten Blick mag es erscheinen, daß die Bedingung (III) erst dann expliziert werden kann, wenn der Begriff der Relevanz zur Verfügung steht, der seinerseits zu seiner Definition den Begriff der adäquaten induktiven Methode c voraussetzt [indem z. B. gesagt wird, daß Γ (positiv oder negativ) relevant für Ψ bezüglich Φ genau dann ist, wenn $c(\Psi, \Phi \wedge \Gamma) \neq c(\Psi, \Phi)$]. Andererseits ist bereits deutlich gemacht worden, daß mit dem Begriff der Apriorirelevanz nicht dieser allgemeine Relevanzbegriff gemeint sein kann, sondern der folgende Sachverhalt: Wenn einige Prädikate auf Grund ihrer Definition elementfremd sind (wie etwa die V-Prädikate) und die Hypothese behauptet, daß ein Ding eine durch ein solches Prädikat bezeichnete Eigenschaft habe, so ist bezüglich dieser Hypothese a priori nur relevant, ob die in der Prämisse erwähnten Gegenstände diese Eigenschaft haben oder nicht haben; a priori irrelevant ist jedoch, welche anderen Eigenschaften sie haben, falls sie die erwähnte nicht besitzen. Unter Verwendung von V-Prädikaten kann daher folgende Definition jenes Begriffs angegeben werden:

D 18 Durch c wird nur der für die Hypothese a priori relevante Teil der Prämisse berücksichtigt genau dann, wenn für alle Sätze Ψ, Φ und Γ, für jedes V-Prädikat Π^* und für jeden Gegenstandsausdruck α gilt: Falls Ψ ein Satz von der Form $\alpha \in \Pi^*$, Φ eine erfüllbare Konjunktion von V-Sätzen und Γ ein Satz ist, der aus Φ dadurch entsteht, daß jeder V-Satz $\beta \in \Pi^+$ in Φ (für jeden Gegenstandsausdruck β und jedes von Π^* verschiedene V-Prädikat Π^+) durch $\neg (\beta \in \Pi^*)$ ersetzt wird, so ist $c(\Psi, \Phi) = c(\Psi, \Gamma)$.

Ist $p = 1$ und damit $q = 2$, so sind die Sätze Φ und Γ identisch; denn das eine V-Prädikat Π^*_1 ist gerade das Grundprädikat Π_1, so daß $\alpha \in \Pi^*_1$ und $\alpha \in \Pi_1$ dieselben Sätze sind, und der Satz $\alpha \in \Pi^*_2$ mit dem anderen V-Prädikat Π^*_2 ist eine abkürzende Schreibweise für $\neg (\alpha \in \Psi_1)$, so daß bei der Ersetzung von $\alpha \in \Pi^*_2$ durch $\neg (\alpha \in \Pi_1)$ der gleiche Satz entsteht. Wenn Φ und Γ miteinander identisch sind, so gilt jedoch aus rein logischen Gründen: $c(\Psi, \Phi) = c(\Psi, \Gamma)$, so daß die Bedingung (III) in diesem Fall keine echte Aussage über den Wahrscheinlichkeitsbegriff macht.

c ermöglicht es uns, aus der Erfahrung zu lernen, wenn wir bei der Anwendung von c zur Ermittlung von Wahrscheinlichkeiten nicht irregeführt werden, wenn sich also (a) der c-Wert nicht vermindert, falls zum bisherigen Wissen Φ ein für die Hypothese im intuitiven Sinn positiv relevantes Zusatzwissen Γ hinzukommt, und wenn (b) der c-Wert in diesem Fall auch nicht gleich bleibt, wenn er sich also erhöht. Die Teilaussage (a) hat ohne Zweifel eine höhere Plausibilität: Es würde allem induktiven Argumentieren widersprechen, wenn eine in Frage kommende Funktion c bei einem positiven Zusatzdatum einen geringeren c-Wert liefern würde. Daß (b) nicht in allen Fällen erfüllt ist, wird man auf den ersten Blick vielleicht zuzugeben bereit sein; denn wenn etwa zwei Münzwürfe mit dem Ergebnis *Wappen* geendet haben, so wird man bei einem zusätzlichen Wurf mit dem gleichen Ergebnis den Wahrscheinlichkeitswert dafür, daß auch der nächste Wurf wiederum *Wappen* ergibt, nicht wesentlich ändern. Andererseits folgt aus (I), (II) und (III), daß der Wahrscheinlichkeitswert dafür, daß der nächste Wurf *Wappen* ergibt, sich niemals (auch nicht z. B. nach einer Million Würfen) ändert, wenn er sich nicht

schon nach dem ersten Wurf, wenn auch vielleicht unmerklich, von der Aprioriwahrscheinlichkeit unterscheidet.[3] Die Bedingung (IV) braucht also nicht in voller Allgemeinheit analysiert zu werden; es genügt vielmehr, von dem Anfangszuwachs zu fordern, daß er bei einer positiven Instanz (a) nicht negativ und (b) von 0 verschieden ist, daß eine positive Instanz also nicht negativ relevant und nicht irrelevant ist.

Da nicht nur die V-Prädikate durch die Grundprädikate, sondern auch umgekehrt die Grundprädikate durch die V-Prädikate definierbar sind und somit alle in der Sprache definierbaren Prädikate auf V-Prädikate zurückführbar sind, genügt es, den Begriff des Lernens aus der Erfahrung für V-Prädikate einzuführen:

D 19 c ermöglicht es, aus der Erfahrung zu lernen genau dann, wenn für jedes V-Prädikat Π^* und für zwei beliebige Gegenstandsausdrücke α und β gilt:
(a) $c(\beta \in \Pi^*, \alpha \in \Pi^*) \lessdot c(\alpha \in \Pi^*, T)$, sowie
(b) $c(\beta \in \Pi^*, \alpha \in \Pi^*) \neq c(\alpha \in \Pi^*, T)$.

Es kann gezeigt werden, daß aus (IV) (b) unter Zugrundelegung von (1) bis (5), (II) und (III) die Bedingung (6) von D 4 logisch folgt und somit überflüssig wird. Will man andererseits an dieser Bedingung festhalten, und ist man darauf bedacht, daß die einzelnen Forderungen voneinander unabhängig sind, so kann man (b) dahingehend abschwächen, daß man ihr die Voraussetzung „falls $c(\alpha \in \Pi^*, T) \neq 1$" hinzufügt.

Die Bedingung (IV) (a) folgt teilweise und bei unendlichem Gegenstandsbereich ganz aus (I), (II) und (III),[4] so daß es für unendl che Bereiche genügen würde, (IV) (b) mit der angegebenen Abschwächung zu fordern. Für endliche Bereiche gibt es hingegen Funktionen c, die (I), (II) und (III) erfüllen, jedoch (IV) (a) verletzen. Es sei z. B. $n = 2$ und $p = 1$, so daß an V-Prädikaten nur Π_1 und das durch $\neg(\alpha \in \Pi_1)$ gegebene Prädikat definierbar sind. Diese Sprache enthält die folgenden 4 Zustandsbeschreibungen:

---

[3] Vgl. den Folgesatz 4 auf S. 163.
[4] Vgl. Theorem 20 auf S. 218 dieser Arbeit.

$\alpha_1 \varepsilon \Pi_1 \wedge \alpha_2 \varepsilon \Pi_1$ $\quad$ $\alpha_1 \varepsilon \Pi_1 \wedge \neg (\alpha_2 \varepsilon \Pi_1)$

$\neg (\alpha_1 \varepsilon \Pi_1) \wedge \alpha_2 \varepsilon \Pi_1$ $\quad$ $\neg (\alpha_1 \varepsilon \Pi_1) \wedge \neg (\alpha_2 \varepsilon \Pi_1)$

Die c-Maße dieser Zustandsbeschreibungen seien durch die folgenden Gleichungen gegeben:[5]

$c (\alpha_1 \varepsilon \Pi_1 \wedge \alpha_2 \varepsilon \Pi_1, T) = 1/8$

$c (\neg (\alpha \varepsilon \Pi_1) \wedge \alpha_2 \varepsilon \Pi_1, T) = 3/8$

$c (\alpha_1 \varepsilon \Pi_1 \wedge \neg (\alpha_2 \varepsilon \Pi_1), T) = 3/8$

$c (\neg (\alpha_1 \varepsilon \Pi_1) \wedge \neg (\alpha_2 \varepsilon \Pi_1), T) = 1/8$

Die übrigen Werte der Funktion c seien durch D 2 und D 3 gegeben. c erfüllt also die Bedingung (I) und darüber hinaus auch (II), sowie (III), wie man sofort sieht. Andererseits ist (IV)(a) verletzt, denn es gilt: $1/4 = c (\alpha_2 \varepsilon \Pi_1, \alpha_1 \varepsilon \Pi_1) < c (\alpha_1 \varepsilon \Pi_1, T) = 1/2$.

Wer dieses c benützt, denkt *anti-induktiv*; denn es gilt weiterhin: $3/4 = c (\neg (\alpha_2 \varepsilon \Pi_1), \alpha_1 \varepsilon \Pi_1) > c (\alpha_1 \varepsilon \Pi_1, T) = 1/2$.

Es ist dies das Verhalten von sogenannten Systemspielern bei Glücksspielen; wenn etwa beim Münzwurf ein Wurf (oder mehrere Würfe hintereinander) *Wappen* ergeben haben, so ist nach ihrer Ansicht die Wahrscheinlichkeit dafür, daß der nächste Wurf nicht mehr *Wappen* ergibt, sehr groß.

Die Bedingung (IV) (b) ist verletzt, wenn den obigen Zustandsbeschreibungen durchweg die Aprioriwahrscheinlichkeiten $1/4$ zugeordnet werden, was zunächst naheliegen würde. Denn dann ändert sich die Ausgangswahrscheinlichkeit bei zusätzlicher Erfahrung nicht, sondern es gilt:

$1/2 = c (\alpha_2 \varepsilon \Pi_1, \alpha_1 \varepsilon \Pi_1) = c (\alpha_2 \varepsilon \Pi_1, T) = c (\alpha_1 \varepsilon \Pi_1, T)$

$= c (\neg (\alpha_2 \varepsilon \Pi_1), \alpha_1 \varepsilon \Pi_1) = c (\neg (\alpha_2 \varepsilon \Pi_1), T) = c (\neg (\alpha_1 \varepsilon \Pi_1), T)$.

2. *Die Normalform der adäquaten induktiven Methoden*

Eine adäquate induktive Methode ist also eine reguläre Bestätigungsmethode mit gewissen zusätzlichen Eigenschaften. Der Wahrscheinlichkeitswert ist demnach für jedes geordnete Paar, bestehend aus einem Satz und einem erfüllbaren Satz, eindeutig berechenbar, wenn die Aprioriwahrscheinlichkeiten für alle Zustandsbeschreibungen ge-

[5] Wegen Theorem 15 gilt, daß dieses c durch $L = -3$ bestimmt ist.

geben sind. Eine Zustandsbeschreibung Σ kann formuliert werden als eine Konjunktion von n V-Sätzen in den Gegenstandsausdrücken $\alpha_1, \ldots, \alpha_n$, z. B. als $\Psi_{i_1} \wedge \Psi_{i_2} \wedge \ldots \wedge \Psi_{i_n}$, wobei für jede natürliche Zahl k mit $0 \leqslant k < n$ der Satz $\Psi_{i_{k+1}}$ die Form $\alpha_{k+1} \varepsilon \Pi^*_{i_{k+1}}$ hat. Φ_0 sei logisch äquivalent mit T, Φ_1 sei der gleiche Satz wie Ψ_{i_1} und Φ_k sei eine Abkürzung für $\Psi_{i_1} \wedge \ldots \wedge \Psi_{i_k}$. Da Σ eine Zustandsbeschreibung und also erfüllbar ist, ist auch jedes Konjunktionsglied von Σ erfüllbar. Es gilt daher:

$$\begin{aligned} c(\Sigma, T) &= c(\Psi_{i_1} \wedge \Psi_{i_2} \wedge \ldots \wedge \Psi_{i_n}, T) \\ &= c(\Psi_{i_1}, T) \cdot c(\Psi_{i_2}, \Psi_{i_1}) \cdot \ldots \cdot c(\Psi_{i_n}, \Psi_{i_1} \wedge \ldots \wedge \Psi_{i_{n-1}}) \\ &= c(\Psi_{i_1}, \Phi_0) \cdot c(\Psi_{i_2}, \Phi_1) \cdot \ldots \cdot c(\Psi_{i_n}, \Phi_{n-1}). \end{aligned}$$

Φ_k sagt also von den ersten k Gegenständen aus B aus, daß von ihnen k_i die durch $\Pi^*_{i_{k+1}}$ bezeichnete Eigenschaft haben.[6] Da q die Anzahl der V-Prädikate ist, kann folgendes Theorem bewiesen werden:[7]

Theorem 14

c sei eine reguläre Bestätigungsmethode.
Dann ist c eine adäquate induktive Methode (für Sprachen mit endlich vielen Gegenstandsausdrücken und endlich vielen einstelligen Grundprädikaten, deren Anzahl größer als 1 ist) genau dann, wenn es eine positive reelle Zahl L gibt, so daß für alle natürlichen Zahlen k mit $0 \leqslant k < n$ und für alle natürlichen Zahlen i mit $1 \leqslant i \leqslant q$ (so daß k_i eine natürliche Zahl mit $0 \leqslant k_i \leqslant k$ ist und Φ_k sowie $\Psi_{i_{k+1}}$ Sätze der beschriebenen Art sind) gilt:

$$c(\Psi_{i_{k+1}}, \Phi_k) = \frac{k_i + (1/q) \cdot L}{k + L}.$$

Beweis

c sei eine reguläre Bestätigungsmethode.
(a) c sei eine Funktion der angegebenen Form für eine gewisse positive reelle Zahl L.

---

[6] k_i ist also eine Funktion von k und i_{k+1}, bzw. kurz von k und i.

[7] Ein dem folgenden Lehrsatz ähnlicher Sachverhalt wurde erstmals von Rudolf Carnap in [10] bewiesen. Der hier gegebene Beweis verwendet eine Argumentation von John G. Kemeny in [8]. Vgl. auch Carnap-Stegmüller [1], S. 242 bis 249.

Dann ist c auch symmetrisch bezüglich der außerlogischen Konstanten. Denn angenommen, Δ entsteht aus Ψ dadurch, daß zwei Gegenstandsausdrücke α_i und α_j durch einander ersetzt werden; dann gilt: impliziert eine Zustandsbeschreibung Σ den Satz Ψ und entsteht Σ' aus Σ durch den gleichen Austausch der Gegenstandsausdrücke α_i und α_j, dann folgt aus Σ' logisch Δ, und umgekehrt. Von jedem derartigen Paar Σ, Σ' von Zustandsbeschreibungen gilt jedoch: $c(\Sigma, T) = c(\Sigma', T)$, denn die einzelnen Brüche und ihr Produkt hängen nicht davon ab, *welche* Gegenstände die einzelnen, durch V-Prädikate ausgedrückten Eigenschaften haben, sondern nur, *wievielen* Gegenständen diese Eigenschaften zukommen. Das gleiche gilt, wenn Δ aus Ψ und analog Σ' aus Σ durch wechselseitige Ersetzung zweier V-Prädikate entstehen; die Brüche und das Produkt dieser Brüche hängen nicht von dem Ausdruck $\Pi^*_{i_{k+1}}$ ab, sondern nur von der Anzahl k_i der Dinge, die die damit bezeichnete Eigenschaft besitzen. Sowohl die Gegenstandsausdrücke als auch die V-Prädikate können also neu numeriert werden, ohne daß damit eine Änderung des Wahrscheinlichkeitswertes verbunden ist.

Ferner wird durch c nur der für die Hypothese a priori relevante Teil der Prämisse berücksichtigt. Denn angenommen, Π^* ist ein V-Prädikat, Ψ ein Satz von der Form $\alpha \varepsilon \Pi^*$ und Φ eine erfüllbare Konjunktion von V-Sätzen; wegen (II) (a) können die Gegenstandsausdrücke so umnumeriert werden, daß $\Psi \wedge \Phi$ ein Anfangsstück einer Zustandsbeschreibung ist und Ψ sowie Φ demnach Sätze der Form $\Psi_{i_{k+1}}$ und Φ_k sind. Der dadurch festgelegte Bruch hängt jedoch (bei fest vorgegebenem L) nur von den Zahlen k_i und k ab; die übrigen Anzahlen k_j sind für den Funktionswert irrelevant.

Schließlich ermöglicht es c, aus der Erfahrung zu lernen. Denn ist $\Pi^*_{i_1} = \Pi^*_{i_2}$, so ist Ψ_{i_2} der Satz $\alpha_2 \varepsilon \Pi^*_1$, und es gilt:

$$c(\Psi_{i_2}, \Psi_{i_1}) = \frac{1 + L/q}{1 + L} > \frac{1}{q} = c(\Psi_{i_1}, T).$$

(b) c sei eine adäquate induktive Methode; es ist zu zeigen, daß der Wert von $c(\Psi_{i_{k+1}}, \Phi_k)$ für eine reelle Zahl L durch jenen Bruch gegeben ist. $\Psi_{i_{k+1}}$ hat die Form $\alpha_{k+1} \varepsilon \Pi^*_{i_{k+1}}$. Ein solches i_{k+1} wird für das folgende festgehalten; es wird angenommen, daß es mit i identisch ist (im gegenteiligen Fall wird diese Identität durch Umnumerierung

der V-Prädikate, die ja wegen (II) (b) möglich ist, hergestellt). Daher wird im folgenden „Π^*_i" statt „$\Pi^*_{i_{k+1}}$" geschrieben. Nach (II) (a) hängt der Wert von $c(\Phi_k, T)$ nicht von den in Φ_k vorkommenden Gegenstandsausdrücken und den dadurch bezeichneten Dingen ab; entscheidend ist vielmehr nur die Anzahl der Gegenstände, die die jeweiligen, durch $\Pi^*_1, \ldots, \Pi^*_q$ bezeichneten Eigenschaften haben.[8] c ist also eine Funktion h dieser Anzahlen $k_1, \ldots, k_q$, d. h. es gilt:
$c(\Phi_k, T) = h(k_1, \ldots, k_q) = h(k_1, \ldots, k_i, \ldots, k_q)$.
Das Analoge gilt dann bezüglich $\Psi_{i_{k+1}} \wedge \Phi_k$:
$c(\Psi_{i_{k+1}} \wedge \Phi_k, T) = h(k_1, \ldots, k_i + 1, \ldots, k_q)$.
Der Wert von $c(\Psi_{i_{k+1}}, \Phi_k)$ ist also durch eine Funktion g bestimmt, die nur von diesen Anzahlen abhängt:

$$c(\Psi_{i_{k+1}}, \Phi_k) = \frac{c(\Psi_{i_{k+1}} \wedge \Phi_k, T)}{c(\Phi_k, T)}$$

$$= \frac{h(k_1, \ldots, k_i + 1, \ldots, k_q)}{h(k_1, \ldots, k_i, \ldots, k_q)} = g(k_1, \ldots, k_i, \ldots, k_q).$$

Es ist $k = k_1 + \ldots + k_i + \ldots + k_q$. Entsteht Γ_k aus Φ_k dadurch, daß in Φ_k jeder V-Satz $\alpha_j \in \Pi^*_{i_j}$ in α_j, der verschieden von $\alpha_j \in \Pi^*_{i_{k+1}}$ ist, durch $\neg(\alpha_j \in \Pi^*_{i_{k+1}})$ ersetzt wird, so ändert sich nach (III) der c-Wert nicht. c hängt also nicht von den von q verschiedenen Anzahlen k_j ab, sondern nur von k_i und $k - k_i$ bzw., was rechnerisch auf das gleiche hinausläuft, von k und k_i. c ist also eine Funktion f, die nur von diesen beiden Anzahlen abhängt: $c(\Psi_{i_{k+1}}, \Phi_k) = f(k, k_i)$.
Das Theorem wird nun unter Verwendung dreier weiterer Hilfssätze bewiesen.

Lemma 3

f sei die angegebene Funktion; es sei $k = \sum_{j=1}^{q} k_j$.
Dann ist $\sum_{j=1}^{q} f(k, k_j) = 1$.

---

[8] Φ'_k entstehe aus Φ_k durch Vertauschung von Gegenstandsausdrücken; dann gilt wegen (II) (a): $c(\Phi_k, T) = h(k_1, \ldots, k_q) = c(\Phi'_k, T)$.

Beweis

Die Bedingungen seien gegeben. Es ist $f(k, k_j) = c(\Psi_{j_{k+1}}, \Phi_k)$. Also:

$$\sum_{j=1}^{q} f(k, k_j) = f(k, k_1) + \ldots + f(k, k_q)$$
$$= c(\alpha_{k+1} \varepsilon \Pi^*_1, \Phi_k) + \ldots + c(\alpha_{k+1} \varepsilon \Pi^*_q, \Phi_k).$$

Zwei verschiedene Sätze $\alpha_{k+1} \varepsilon \Pi^*_j$ und $\alpha_{k+1} \varepsilon \Pi^*_i$ sind miteinander logisch unverträglich; es gilt also:

$c(\alpha_{k+1} \varepsilon \Pi^*_1, \Phi_k) + \ldots + c(\alpha_{k+1} \varepsilon \Pi^*_q, \Phi_k) =$
$c(\alpha_{k+1} \varepsilon \Pi^*_1 \vee \ldots \vee \alpha_{k+1} \varepsilon \Pi^*_q, \Phi_k)$.

Die Gesamtheit aller $\alpha_{k+1} \varepsilon \Pi^*_j$ ist logisch erschöpfend,
$\alpha_{k+1} \varepsilon \Pi^*_1 \vee \ldots \vee \alpha_{k+1} \varepsilon \Pi^*_q$ also logisch wahr, und daher:
$c(\alpha_{k+1} \varepsilon \Pi^*_1 \vee \ldots \vee \alpha_{k+1} \varepsilon \Pi^*_q, \Phi_k) = c(T, \Phi_k) = 1$,
womit Lemma 3 bewiesen ist.

Es sei $k = 0$ und daher $k_j = 0$ für alle j $(1 \leqslant j \leqslant q)$. Dann folgt aus Lemma 3: $q \cdot f(0,0) = 1$, und daher:
(A) $f(0,0) = 1/q$.
Steht also keine faktische Information zur Verfügung, so haben alle Sätze Ψ_{j_1} die gleiche Aprioriwahrscheinlichkeit $1/q$.
Es sei $k = 1$, $k_i = 1$ und $k_j = 0$ für alle übrigen j, so daß
$f(1,1) = c(\alpha_2 \varepsilon \Pi^*_{i_1}, \alpha_1 \varepsilon \Pi^*_{i_1})$ und $f(1,0) = c(\alpha_2 \varepsilon \Pi^*_{j_1}, \alpha_1 \varepsilon \Pi^*_{i_1})$.
Dann ist wegen Lemma 3: $f(1,1) + (q-1) \cdot f(1,0) = 1$. Unter Benützung der Abkürzung $l = f(1,0)$ erhält man:
(B) $f(1,1) = 1 - (q-1) \cdot l$.
Der folgende Hilfssatz liefert die Rekursionsschritte für eine rekursive Definition von f:

Lemma 4

f sei die angegebene Funktion; es sei $0 \leqslant k < n$ und $0 \leqslant k_i \leqslant k$. Dann gilt: $f(k+1, k_i) \cdot f(k, 0) = f(k+1, 0) \cdot f(k, k_i)$.

Beweis

Die Bedingungen seien gegeben. h ist, wie früher erwähnt, die Funktion, die die Aprioriwahrscheinlichkeit bestimmt. Φ_k und Φ_{k+1} sind erfüllbar, so daß $c(\Phi_k, T) \neq 0$ sowie $c(\Phi_{k+1}, T) \neq 0$ und somit

auch die in der folgenden allgemeingültigen Gleichung vorkommenden Nenner verschieden von 0 sind:

$$\frac{h(\ldots, k_i+1, \ldots, k_j+1, \ldots)}{h(\ldots, k_i, \ldots, k_j, \ldots)} = \frac{h(\ldots, k_i+1, \ldots, k_j+1, \ldots)}{h(\ldots, k_i, \ldots, k_j, \ldots)}$$

Multipliziert man die linke Seite der Gleichung mit

$$\frac{h(\ldots, k_i, \ldots, k_j+1, \ldots)}{h(\ldots, k_i, \ldots, k_j+1, \ldots)} \text{ und die rechte mit } \frac{h(\ldots, k_i+1, \ldots, k_j, \ldots)}{h(\ldots, k_i+1, \ldots, k_j, \ldots)},$$

so erhält man daraus durch geeignete Kommutation:

$$\frac{h(\ldots, k_i+1, \ldots, k_j+1, \ldots)}{h(\ldots, k_i, \ldots, k_j+1, \ldots)} \cdot \frac{h(\ldots, k_i, \ldots, k_j+1, \ldots)}{h(\ldots, k_i, \ldots, k_j, \ldots)} =$$

$$= \frac{h(\ldots, k_i+1, \ldots, k_j+1, \ldots)}{h(\ldots, k_i+1, \ldots, k_j, \ldots)} \cdot \frac{h(\ldots, k_i+1, \ldots, k_j, \ldots)}{h(\ldots, k_i, \ldots, k_j, \ldots)}$$

Auf Grund der Definition von g gilt nun:

$$\frac{h(\ldots, k_i+1, \ldots, k_j+1, \ldots)}{h(\ldots, k_i, \ldots, k_j+1, \ldots)} = g(\ldots, k_i, \ldots, k_j+1, \ldots)$$

$$\frac{h(\ldots, k_i+1, \ldots, k_j, \ldots)}{h(\ldots, k_i, \ldots, k_j, \ldots)} = g(\ldots, k_i, \ldots, k_j, \ldots)$$

Wegen (II) (b) ändern sich die Werte von c (Φ_k, T), c (Φ_{k+1}, T) und c (Φ_{k+2}, T) und damit auch die entsprechenden h-Werte nicht, wenn die V-Prädikate Π^*_i und Π^*_j in diesen Sätzen durch einander ersetzt werden:

$$\frac{h(\ldots, k_i, \ldots, k_j+1, \ldots)}{h(\ldots, k_i, \ldots, k_j, \ldots)} = \frac{h(\ldots, k_j+1, \ldots, k_i, \ldots)}{h(\ldots, k_j, \ldots, k_i, \ldots)}$$

$$= g(\ldots, k_j, \ldots, k_i, \ldots)$$

$$\frac{h(\ldots, k_i+1, \ldots, k_j+1, \ldots)}{h(\ldots, k_i+1, \ldots, k_j, \ldots)} = \frac{h(\ldots, k_j+1, \ldots, k_i+1, \ldots)}{h(\ldots, k_j, \ldots, k_i+1, \ldots)}$$

$$= g(\ldots, k_j, \ldots, k_i+1, \ldots)$$

Auf Grund dieser Gleichungen gilt:

$g(\ldots, k_i, \ldots, k_j+1, \ldots) \cdot g(\ldots, k_j, \ldots, k_i, \ldots) =$
$= g(\ldots, k_j, \ldots, k_i+1, \ldots) \cdot g(\ldots, k_i, \ldots, k_j, \ldots).$

Wegen (III) sind zur Berechnung der g-Werte neben den Anzahlen k und $k+1$ nur die an den jeweils ersten der beiden hervorgehobenen Stellen bezeichneten Anzahlen relevant, so daß g durch f ersetzt werden kann:
$f(k+1, k_i) \cdot f(k, k_j) = f(k+1, k_j) \cdot f(k, k_i)$.
Ist $k_j = 0$,[9] so folgt daraus:
$f(k+1, k_i) \cdot f(k, 0) = f(k+1,0) \cdot f(k, k_i)$.
Damit ist das Lemma 4 bewiesen.

Da Φ_k erfüllbar ist, womit $c(\Phi_k, T) \neq 0$, impliziert Lemma 4 unmittelbar:

$$(C) \quad f(k+1, k_i) = \frac{f(k, k_i)}{f(k, 0)} \cdot f(k+1,0).$$

Aus Lemma 3 folgt, wenn man „k" durch „$k+1$" ersetzt, für den Fall, daß alle $k+1$ Dinge die durch Π^*_i bezeichnete Eigenschaft haben: $f(k+1, k+1) + (q-1) \cdot f(k+1,0) = 1$, und damit:

$$(D) \quad f(k+1, k+1) = 1 - (q-1) \cdot f(k+1,0).$$

Haben hingegen nur k Dinge diese Eigenschaft und der restliche Gegenstand eine davon verschiedene (die durch ein anderes Π^*_j bezeichnet wird), so ist:
$f(k+1, k) + f(k+1,1) + (q-2) \cdot f(k+1,0) = 1$.
Unter Verwendung von (C) erhält man daraus:

$$f(k+1,0) \cdot \left[\frac{f(k,k)}{f(k,0)} + \frac{f(k,1)}{f(k,0)} + (q-2)\right] = 1, \text{ sowie}$$

$f(k+1,0) \cdot [f(k,k) + f(k,1) + (q-2) \cdot f(k,0)] = f(k,0)$.
Da Zustandsbeschreibungen erfüllbar sind, folgt daraus:

$$(E) \quad f(k+1,0) = \frac{f(k,0)}{f(k,k) + f(k,1) + (q-2) \cdot f(k,0)}$$

Die Gleichungen (C), (D) und (E) definieren rekursiv die f-Werte für $k > 1$ in f-Werten für $k = 0$ und $k = 1$. Für diese Argumente

---

[9] Wenn $p = 1$ und damit $q = 2$, darf diese Annahme nicht gemacht werden; denn auf Grund der Ausgangsgleichung ist dann k_i kein von k unabhängiger Wert mehr, sondern mit k identisch. (Diesen Hinweis verdanke ich Herrn Dipl.-Math. Jürgen Humburg.)

wiederum definieren (A) und (B) die f-Werte unter Bezugnahme auf l. Die so rekursiv gewonnene Funktion f kann auch explizit charakterisiert werden, wie der folgende Hilfssatz zeigt:

Lemma 5

f sei die angegebene Funktion; es sei $0 \leqslant k < n$ und $0 \leqslant k_1 \leqslant k$. Dann gilt:

$$f(k, k_1) = \frac{k_1 + \dfrac{l}{1 - q \cdot l}}{k + \dfrac{q \cdot l}{1 - q \cdot l}}$$

Beweis

Die Bedingungen seien gegeben. Die Behauptung wird dann durch Rekursion bewiesen.

$$\text{(A)} \quad f(0,0) = \frac{1}{q} = \frac{0 + \dfrac{}{1 - q}}{0 + \dfrac{ql}{1 - ql}}$$

$$\text{(B)} \quad f(1,1) = 1 - (q-1) \cdot l = \frac{1 - (q-1) \cdot l}{1}$$

$$= \frac{(1 - ql + l) \cdot (1 - ql)}{(1 - ql) \cdot (1 - ql + ql)} = \frac{1 + \dfrac{l}{1 - ql}}{1 + \dfrac{ql}{1 - ql}}$$

$$\text{(E)} \quad f(k+1,0) = \frac{f(k,0)}{f(k,k) + f(k,1) + (q-2) \cdot f(k,0)}$$

$$= \frac{0 + \dfrac{l}{1 - ql}}{k + \dfrac{ql}{1 - ql}} : \left[\frac{k + \dfrac{l}{1 - ql}}{k + \dfrac{ql}{1 - ql}} + \frac{1 + \dfrac{l}{1 - ql}}{k + \dfrac{ql}{1 - ql}} + \frac{\dfrac{(q-2) \cdot l}{1 - ql}}{k + \dfrac{ql}{1 - ql}} \right]$$

$$= \frac{0 + \dfrac{1}{1-ql}}{k + \dfrac{1}{1-ql} + 1 + \dfrac{1}{1-ql} + \dfrac{(q-2)\cdot 1}{1-ql}}$$

$$= \frac{0 + \dfrac{1}{1-ql}}{k + 1 + \dfrac{ql}{1-ql}}$$

(D) $f(k+1, k+1) = 1 - (q-1)\cdot f(k+1, 0)$

$$= 1 - \frac{(q-1)\cdot\left(0 + \dfrac{1}{1-ql}\right)}{k + 1 + \dfrac{ql}{1-ql}}$$

$$= \frac{k + 1 + \dfrac{ql}{1-ql} - \dfrac{(q-1)1}{1-ql}}{k + 1 + \dfrac{ql}{1-ql}}$$

$$= \frac{k + 1 + \dfrac{1}{1-ql}}{k + 1 + \dfrac{ql}{1-ql}}$$

(C) $f(k+1, k_i) = f(k+1, 0) \cdot \dfrac{f(k, k_i)}{f(k, 0)}$

$$= \frac{f(k, k_i)}{f(k, k) + f(k, 1) + (q-2)\cdot f(k, 0)}$$

$$= \frac{k_i + \dfrac{1}{1-ql}}{k + \dfrac{ql}{1-ql}} : \left[\frac{k + \dfrac{1}{1-ql}}{k + \dfrac{ql}{1-ql}} + \frac{1 + \dfrac{1}{1-ql}}{k + \dfrac{ql}{1-ql}} + \frac{\dfrac{(q-2)1}{1-ql}}{k + \dfrac{ql}{1-ql}}\right]$$

$$= \frac{k_i + \frac{1}{1-ql}}{k + 1 + \frac{ql}{1-ql}}$$

Damit ist Lemma 5 bewiesen.

Benützt man die Abkürzung: $L = \frac{ql}{1-ql}$, so folgt aus Lemma 5:

$$f(k, k_i) = \frac{k_i + \frac{1}{q} \cdot L}{k + L}, \text{ wobei } 0 \leqslant k < n \text{ und } 0 \leqslant k_i \leqslant k.$$

Ist also L fest vorgegeben, so hängt der f-Wert (und damit auch der c-Wert) nur von der Anzahl k der beobachteten Dinge und der Anzahl k_i der positiven Fälle ab.

Es bleibt noch zu zeigen, aus welchem Intervall L genommen werden darf. Auf Grund einer früheren Festsetzung gilt: $l = f(1,0)$. l ist also ein Wahrscheinlichkeitswert, so daß $0 \leqslant l \leqslant 1$. Da sowohl Ψ_{i_1} und $\neg \Psi_{i_2}$ als auch Ψ_{i_1} und Ψ_{i_2} simultan erfüllbar sind, gilt wegen Theorem 2 (a) darüber hinaus: $0 < l < 1$. Nun ist: $f(0,0) = c(\alpha_1 \varepsilon \Pi^*_{i_1}, T)$ und wegen (3): $f(1,1) = c(\alpha_2 \varepsilon \Pi^*_{i_1}, \alpha_1 \varepsilon \Pi^*_{i_1})$. Mit Bedingung (IV) folgt daraus: $f(1,1) > f(0,0)$, was wegen (A) und (B) impliziert: $1 - (q-1) \cdot l > 1/q$, $1 - 1/q > (q-1) \cdot l$, $1/q > l$, $ql < 1$ und schließlich $L < \infty$. Da $ql < 1$ und $0 < l < 1$, gilt andererseits: $0 < ql$, sowie: $0 < L$.

Damit ist auch die andere Implikation des Theorems 14 bewiesen.

3. *Weitere Adäquatheitsbedingungen für induktive Methoden*

Es stellt sich nun die Frage, ob man dieses Ergebnis auch erhält, wenn die Sprache nur ein Grundprädikat und damit an V-Prädikaten nur Π_1 und dessen Komplement enthält. Da Bedingung (III) in diesem Fall keinen faktischen Gehalt besitzt, kann man sich leicht klarmachen, daß die Bedingungen (I), (II) und (IV) auch durch eine Funktion c erfüllt werden, die bei festem l und k nicht eine *lineare* Funktion von k_i, sondern z. B. die fünfte Potenz davon ist.

Andererseits wird man mit Recht erwarten, daß der Wert von c (Ψ, Φ) auch in einer Sprache, die zusätzliche Prädikate enthält (die in Ψ und Φ nicht vorkommen), der gleiche wie in der vorgegebenen Sprache ist, daß also insbesondere Sätze einer Sprache mit nur *einem* Grundprädikat Π_1 die gleichen Wahrscheinlichkeitswerte erhalten wie in einer Sprache, die außerdem noch das von Π_1 unabhängige Grundprädikat Π_2 besitzt. Es ist nicht einzusehen, warum ein solches zusätzliches Grundprädikat, das in diesen Sätzen gar nicht vorkommt, auf den Wahrscheinlichkeitswert irgendeinen Einfluß haben soll.[10]
Es muß daher die Bedingung (III) durch eine Bedingung (III′) ersetzt werden, die für $p = 1$ zu dem gleichen Ergebnis führt wie (III) für $p > 1$. Dazu bietet sich sofort die Forderung an, daß bei festen k und l die Funktion c linear von k_i abhängen soll. Der Begriff der adäquaten induktiven Methode wird deshalb für $p = 1$ (und endlichem Gegenstandsbereich) so festgelegt:

D 20 c ist eine adäquate induktive Methode (für Sprachen mit endlich vielen Gegenstandsausdrücken und einem einstelligen Grundprädikat) genau dann, wenn die Bedingungen (I), (II) und (IV) von D 13 durch c erfüllt sind und (III′) c ($\Psi_{i_{k+1}}$, Φ_k) bei vorgegebenem l und festem k eine lineare Funktion von k_i ist, also der Gleichung $c(\Psi_{i_{k+1}}, \Phi_k) = k_i \cdot h^*(k) + h'(k)$ genügt (dabei ist $\Psi_{i_{k+1}}$ der Satz $\alpha_{k+1} \varepsilon \Pi^*_i$ mit $i = 1$ oder $i = 2$ und Φ_k die Konjunktion $\alpha_1 \varepsilon \Pi^*_{i_1} \wedge \ldots \wedge \alpha_k \varepsilon \Pi^*_{i_k}$ mit $i_j = 1$ oder $i_j = 2$).

Dann kann auch für $p = 1$ ein Satz über die Normalform der adäquaten induktiven Methoden bewiesen werden:[11]

Theorem 15

c sei eine reguläre Bestätigungsmethode.
Dann ist c eine adäquate induktive Methode (für Sprachen mit endlich vielen Gegenstandsausdrücken und einem einstelligen Grund-

10 Wenn eine Sprache z. B. nur das Grundprädikat „schön", eine andere zusätzlich auch noch den Begriff „intelligent" enthält, so ist nicht einzusehen, warum Sätze in dieser anderen Sprache, die an Eigenschaftsbegriffen nur „schön" enthält, einen anderen Wahrscheinlichkeitswert haben sollen als in der ersten.
11 Vgl. Carnap-Stegmüller [1], S. 249, und Humburg [1], S. 16.

prädikat) genau dann, wenn es eine positive reelle Zahl L gibt, so daß für alle natürlichen Zahlen k mit $0 \leqslant k < n$ und für alle natürlichen Zahlen i mit $1 \leqslant i \leqslant 2$ (so daß k_i eine natürliche Zahl mit $0 \leqslant k_i \leqslant k$ ist und Φ_k sowie $\Psi_{i_{k+1}}$ Sätze der beschriebenen Art sind) gilt:

$$c(\Psi_{i_{k+1}}, \Phi_k) = \frac{k_i + (1/q) \cdot L}{k + L}$$

Beweis

c sei eine reguläre Bestätigungsmethode.

(a) Wenn c dieser Gleichung genügt, so ist c offensichtlich bei vorgegebenem l (und damit auch eindeutig bestimmtem L) sowie bei festem k eine lineare Funktion von k_i.

(b) Der Beweis des Theorems 14 (b) zeigt, daß bis auf Lemma 4 alles wörtlich übernommen werden kann. Auch Lemma 4 kann bis zum vorletzten Schritt übernommen werden (wobei zu berücksichtigen ist, daß h und g hier nur zwei Argumentstellen besitzen). Es darf dann allerdings nicht $k_j = 0$ gesetzt werden, sondern man muß folgendermaßen fortfahren:

Da bei $p = 1$ nur zwei V-Prädikate vorkommen, gilt: $k_j = k - k_i$ und aus: $f(k+1, k_i) \cdot f(k, k_j) = f(k+1, k_j) \cdot f(k, k_i)$ ergibt sich daher: $f(k+1, k_i) \cdot f(k, k - k_i) = f(k+1, k - k_i) \cdot f(k, k_i)$.

Unter Verwendung von (III') und der Voraussetzungen $c(\Psi_{i_{k+1}}, \Phi_k) = f(k, k_i)$, $c(\Psi_{i_{k+2}}, \Phi_{k+1}) = f(k+1, k_i)$, $c(\Psi_{j_{k+1}}, \Phi_k) = f(k, k_j) = f(k, k - k_i)$ und $c(\Psi_{j_{k+2}}, \Phi_{k+1}) = f(k+1, k_j) = f(k+1, k - k_i)$ ergibt sich daraus:

$$[k_i \cdot h^*(k+1) + h'(k+1)] \cdot [(k - k_i) \cdot h^*(k) + h'(k)] =$$
$$[(k - k_i) \cdot h^*(k+1) + h'(k+1)] \cdot [k_i \cdot h^*(k) + h'(k)].$$

Durch elementare algebraische Umformungen ergeben sich daraus die Gleichungen:

$$k_i \cdot (k - k_i) \cdot h^*(k) \cdot h^*(k+1) + k_i \cdot h^*(k+1) \cdot h'(k)$$
$$+ (k - k_i) \cdot h^*(k) \cdot h'(k+1) + h'(k) \cdot h'(k+1) =$$
$$k_i \cdot (k - k_i) \cdot h^*(k) \cdot h^*(k+1) + (k - k_i) \cdot h^*(k+1) \cdot h^*(k)$$
$$+ k_i \cdot h^*(k) \cdot h'(k+1) + h'(k) \cdot h'(k+1);$$

$$k_i \cdot h^*(k+1) \cdot h'(k) + (k - k_i) \cdot h^*(k) \cdot h'(k+1) =$$
$$(k - k_i) \cdot h^*(k+1) \cdot h'(k) + k_i \cdot h^*(k) \cdot h'(k+1);$$

$$(k_i - k + k_i) \cdot h^*(k+1) \cdot h'(k) = (k_i - k + k_i) \cdot h^*(k) \cdot h'(k+1);$$

$$h^*(k+1) \cdot h'(k) = h^*(k) \cdot h'(k+1);$$

$k_i \cdot h^*(k+1) \cdot h'(k) = k_i \cdot h^*(k) \cdot h'(k+1);$
$k_i \cdot h^*(k+1) \cdot h'(k) + h'(k+1) \cdot h'(k) =$
$\quad h'(k+1) \cdot h'(k) + k_i \cdot h^*(k) \cdot h'(k+1);$
$[k_i \cdot h^*(k+1) + h'(k+1)] \cdot [0 \cdot h^*(k) + h'(k)] =$
$\quad [0 \cdot h^*(k+1) + h'(k+1)] \cdot [k_i \cdot h^*(k) + h'(k)];$
Daraus erhält man durch Einsetzungen nach (III') bzw. nach den Voraussetzungen jedoch:
$f(k+1, k_i) \cdot f(k, 0) = f(k+1, 0) \cdot f(k, k_i),$
womit auch für $p = 1$ das Lemma 4 bewiesen ist. Alle weiteren Teile des Beweises von Theorem 14 können ohne Einschränkungen übernommen werden. Damit ist gezeigt, daß die Normalform der adäquaten induktiven Methoden (unter der zusätzlichen Bedingung) auch für $p = 1$ gilt.

Aus den beiden Theoremen 14 und 15 ergeben sich einige bemerkenswerte Folgesätze. So kann ohne Schwierigkeiten gezeigt werden, daß ein Erfahrungsdatum, auch wenn es noch so viele Beobachtungen beschreibt, auf die Berechnung des c-Wertes ohne Einfluß ist, wenn c zwar die Bedingungen (I) bis (IV) (a) erfüllt, jedoch (IV) (b) verletzt, so daß sich der c-Wert nach der ersten Beobachtungsinstanz nicht einmal um einen kleinen Betrag erhöht. Daher ist es in diesem Zusammenhang korrekt, den Begriff des Lernens aus der Erfahrung in D 19 auf diese einfache Art zu definieren. Zur eleganteren Formulierung des Folgesatzes 4 benötigt man dazu den Begriff der individuellen Verteilung; der Vollständigkeit halber sei auch gleich der der statistischen Verteilung charakterisiert.
Eine *individuelle Verteilung* Φ für die Gegenstandsausdrücke $\beta_1, \ldots, \beta_k$ bezüglich des Prädikates Π^+ sagt von jedem der durch $\beta_1, \ldots, \beta_k$ bezeichneten Dinge $b_1, \ldots, b_k$ aus, ob es die durch Π^+ bezeichnete Eigenschaft F^+ hat oder nicht hat. Es kann daher definiert werden:

D 21 (a) Φ ist eine individuelle Verteilung für die Gegenstandsausdrücke $\beta_1, \ldots, \beta_k$ bezüglich Π^+ genau dann, wenn Φ die Standardkonjunktion einer maximal-konsistenten Klasse von Basissätzen jener Teilsprache ist, die an außerlogischen Ausdrücken nur $\beta_1, \ldots, \beta_k$ und Π^+ enthält (Φ ist also konsistent, wird jedoch inkonsistent, wenn ihm konjunktiv

ein Basissatz in den Ausdrücken $\beta_1, \ldots, \beta_k$ angefügt wird, der darin nicht schon als Konjunktionsglied vorkommt).

(b) Φ ist eine individuelle Verteilung für die Gegenstandsausdrücke $\beta_1, \ldots, \beta_k$ bezüglich der Prädikate $\Pi'_1, \ldots, \Pi'_m$ genau dann, wenn $\Pi'_1, \ldots, \Pi'_m$ eine Division bilden und Φ die Standardkonjunktion $\beta_1 \varepsilon \Pi'_{i_1} \wedge \ldots \wedge \beta_k \varepsilon \Pi'_{i_k}$ ist, wobei die $\Pi'_{i_1}, \ldots, \Pi'_{i_k}$ Prädikate jener Division sind.

(c) Φ ist eine individuelle Verteilung für die Gegenstandsausdrücke $\beta_1, \ldots, \beta_k$ (bezüglich der V-Prädikate $\Pi^*_1, \ldots, \Pi^*_q$)[12] genau dann, wenn Φ die Standardkonjunktion $\beta_1 \varepsilon \Pi^*_{i_1} \wedge \ldots \wedge \beta_k \varepsilon \Pi^*_{i_k}$ ist, wobei die $\Pi^*_{i_1}, \ldots, \Pi^*_{i_k}$ V-Prädikate sind.

(d) Φ ist eine individuelle Verteilung (bezüglich der V-Prädikate $\Pi^*_1, \ldots, \Pi^*_q$)[12] genau dann, wenn Φ eine individuelle Verteilung für alle Gegenstandsausdrücke $\alpha_1, \ldots, \alpha_n$ (bezüglich der V-Prädikate $\Pi^*_1, \ldots, \Pi^*_q$) ist.

Individuelle Verteilungen (ohne Bezugnahme auf Gegenstands- oder Eigenschaftsausdrücke) sind demnach genau die Zustandsbeschreibungen.

Eine statistische Verteilung Γ unterscheidet sich von einer derartigen individuellen Verteilung Φ dadurch, daß sie nur relative Häufigkeiten formuliert und dabei offenläßt, *welche* Gegenstände die betreffenden Eigenschaften besitzen. Γ kann dann als Standardadjunktion von Φ und allen jenen individuellen Verteilungen dargestellt werden, die in folgendem Sinn miteinander *isomorph* sind: Jede von ihnen entsteht aus den anderen durch eine Permutation (Vertauschung) von Gegenstandsausdrücken und durch Neuordnung der Konjunktionsglieder in lexikographischer Reihenfolge.

Ist etwa $\alpha_1 \varepsilon \Pi_1 \wedge \neg (\alpha_3 \varepsilon \Pi_1) \wedge \alpha_3 \varepsilon \Pi_1$ eine derartige individuelle Verteilung Φ, so entsteht durch eine Permutation der Gegenstandsausdrücke daraus der Satz $\alpha_2 \varepsilon \Pi_1 \wedge \neg (\alpha_1 \varepsilon \Pi_1) \wedge \alpha_3 \varepsilon \Pi_1$ und durch Neuordnung der Konjunktionsglieder schließlich die zu Φ isomorphe Aussage $\neg (\alpha_1 \varepsilon \Pi_1) \wedge \alpha_3 \varepsilon \Pi_1 \wedge \alpha_2 \varepsilon \Pi_1$. Eine dritte (und letzte) zu beiden isomorphe Verteilung ist dann der Satz $\alpha_1 \varepsilon \Pi_1 \wedge \alpha_2 \varepsilon \Pi_1 \wedge \neg (\alpha_3 \varepsilon \Pi_1)$. Die Standardadjunktion Γ dieser Sätze ist somit

[12] Dieser Zusatz wird im folgenden meist weggelassen.

$[\neg (\alpha_1 \in \Pi_1) \wedge \alpha_2 \in \Pi_1 \wedge \alpha_3 \in \Pi_1] \vee [\alpha_1 \in \Pi_1 \wedge \neg (\alpha_2 \in \Pi_1) \wedge \alpha_3 \in \Pi_1] \vee [\alpha_1 \in \Pi_1 \wedge \alpha_2 \in \Pi_1 \wedge \neg (\alpha_3 \in \Pi_1)]$. Dieser Satz sagt tatsächlich nicht mehr aus, *welche* Gegenstände die durch Π_1 bezeichnete Eigenschaft F_1 besitzen, sondern nur noch, *wieviele* es sind, nämlich 2 der 3 Dinge; *er besagt also, daß die relative Häufigkeit von F_1 in der aus diesen drei Dingen bestehenden Bezugsklasse $^2/_3$ ist.* Der Begriff der statistischen Verteilung kann daher folgendermaßen definiert werden:

D 22 (a) Γ ist eine statistische Verteilung für die Gegenstandsausdrücke $\beta_1, \ldots, \beta_k$ bezüglich des Prädikates Π^+ genau dann, wenn Γ die Standardadjunktion aller und nur jener individuellen Verteilungen $\Phi_1, \ldots, \Phi_j$ für die $\beta_1, \ldots, \beta_k$ bezüglich Π^+ ist, die miteinander im oben erläuterten Sinn isomorph sind (die aus einander durch Vertauschung der Gegenstandsausdrücke und Neuordnung der Konjunktionsglieder hervorgehen).

(b) Γ ist eine statistische Verteilung für die Gegenstandsausdrücke $\beta_1, \ldots, \beta_k$ bezüglich der Prädikate $\Pi'_1, \ldots, \Pi'_m$ genau dann, wenn Γ die Standardadjunktion aller und nur jener individuellen Verteilungen für die $\beta_1, \ldots, \beta_k$ bezüglich der $\Pi'_1, \ldots, \Pi'_m$ ist, die miteinander isomorph sind.

(c) Γ ist eine statistische Verteilung für die Gegenstandsausdrücke $\beta_1, \ldots, \beta_k$ (bezüglich der V-Prädikate $\Pi^*_1, \ldots, \Pi^*_q$) genau dann, wenn Γ die Standardadjunktion aller und nur jener individuellen Verteilungen für die $\beta_1, \ldots, \beta_k$ (bezüglich der $\Pi^*_1, \ldots, \Pi^*_q$) ist, die miteinander isomorph sind.

(d) Γ ist eine statistische Verteilung (bezüglich der V-Prädikate $\Pi^*_1, \ldots, \Pi^*_q$) genau dann, wenn Γ eine statistische Verteilung für alle Gegenstandsausdrücke $\alpha_1, \ldots, \alpha_n$ (bezüglich der $\Pi^*_1, \ldots, \Pi^*_q$) ist.

Statistische Verteilungen (bezüglich der V-Prädikate) werden von Carnap „Strukturbeschreibungen" genannt. Eine statistische Verteilung Γ entspricht der individuellen Verteilung Φ, wenn sie diese als Adjunktionsglied enthält.
Wenn Bedingung (IV) (b) verletzt ist, so daß es also ein V-Prädikat Π^*

und zwei Gegenstandsausdrücke α und β gibt mit: $c(\beta \varepsilon \Pi^\star, \alpha \varepsilon \Pi^\star) = c(\alpha \varepsilon \Pi^\star, T)$, so ist, die Gültigkeit der übrigen Bedingungen vorausgesetzt, (IV) (b) global verletzt, d. h. der Wahrscheinlichkeitswert ist dann stets mit der Aprioriwahrscheinlichkeit identisch, gleichgültig, welche individuellen Verteilungen Φ bekannt sind:

Korollar 4

(a) c erfülle die Bedingungen (I), (II), (III) und (IV) (a), verletze jedoch (IV) (b) für ein V-Prädikat $\Pi^\star_i$. Dann gilt: Wenn Φ irgendeine individuelle Verteilung für die Gegenstandsausdrücke $\beta_1, \ldots, \beta_k$ bezüglich $\Pi^\star_i$ ist, so ist $c(\alpha \varepsilon \Pi^\star_i, \Phi) = c(\alpha \varepsilon \Pi^\star_i, T)$, d. h. keine Beobachtung hat auf den c-Wert einen Einfluß.
(b) Genügt c den Bedingungen (I) bis (IV) (a), so erfüllt es dann und nur dann (IV) (b), wenn für jede individuelle Verteilung Φ für Gegenstandsausdrücke $\beta_1, \ldots, \beta_k$ bezüglich eines V-Prädikates $\Pi^\star_i$ gilt: Falls $c(\alpha \varepsilon \Pi^\star_i, \Phi) = c(\alpha \varepsilon \Pi_i, T)$, so ist die durch Φ beschriebene relative Häufigkeit k_i/k der durch $\Pi^\star_i$ beschriebenen Eigenschaft in der Klasse der durch $\beta_1, \ldots, \beta_k$ bezeichneten Dinge gleich der relativen Weite von $\Pi^\star_i$, nämlich $1/q$.

Beweis

(a) c erfülle die Bedingungen (I) bis (IV) (a), verletze jedoch (IV) (b) für ein V-Prädikat $\Pi^\star_i$, so daß es also zwei Gegenstandsausdrücke α und β gibt mit: $c(\beta \varepsilon \Pi^\star_i, \alpha \varepsilon \Pi^\star_i) = c(\alpha \varepsilon \Pi^\star_i, T)$. Wegen (II) (a) gilt dies dann und nur dann, wenn $c(\alpha_2 \varepsilon \Pi^\star_i, \alpha_1 \varepsilon \Pi^\star_i) = c(\alpha_1 \varepsilon \Pi^\star_i, T)$. Der Beweis des Theorems 14 zeigt, daß mit (IV) (a) allein nur folgt: $L \leqslant \infty$. Es gilt also:

$$\frac{1 + \frac{1}{q} \cdot L}{1 + L} = \frac{1}{q}, \text{ und somit: } q + L = 1 + L.$$

Ist nun $L < \infty$, so folgt daraus: $q = 1$, was der generellen Voraussetzung widerspricht, daß $p \geqslant 1$ und damit $q > 1$. Also gilt: $L = \infty$. Dann gilt jedoch für ein beliebiges k:

$$\frac{k_i + \frac{1}{q} \cdot L}{k + L} = \frac{1}{q}.$$

Es sei Γ der Satz, der aus der individuellen Verteilung Φ für die Gegenstandsausdrücke $\beta_1, \ldots, \beta_k$ bezüglich Π^*_i dadurch entsteht, daß die darin vorkommenden Gegenstandsausdrücke durch $\alpha_1, \ldots, \alpha_k$ ersetzt werden. Wegen (II) (a) gilt dann:
$c(\alpha \varepsilon \Pi^*_i, \Phi) = c(\alpha_{k+1} \varepsilon \Pi^*_i, \Gamma) = c(\alpha_{k+1} \varepsilon \Pi^*_i, \Phi_k) = 1/q$.
(b) c genüge den Bedingungen (I) bis (IV) (a).
Angenommen, c erfüllt auch die Bedingung (IV) (b), so daß c also eine adäquate induktive Methode ist; ist dann Φ eine beliebige individuelle Verteilung für die Gegenstandsausdrücke $\beta_1, \ldots, \beta_k$ bezüglich eines V-Prädikates Π^*_i, wobei $c(\alpha \varepsilon \Pi^*_i, \Phi) = c(\alpha \varepsilon \Pi^*_i, T)$, so gilt nach dem im Teil (a) Gesagten:

$$\frac{k_i + \frac{1}{q} \cdot L}{k + L} = \frac{1}{q}, \text{ und somit: } k_i q + L = k + L, \text{ sowie: } \frac{k_i}{k} = \frac{1}{q}.$$

Erfüllt c hingegen nicht die Bedingung (IV) (b), so gibt es eine individuelle Verteilung Φ, nämlich die Aussage $\beta \varepsilon \Pi^*_i$, so daß zwar $c(\alpha \varepsilon \Pi^*_i, \Phi) = c(\alpha \varepsilon \Pi^*_i, T)$, aber $k_1/k \neq 1/q$ wegen $q > 1$ und $k = k_1 = 1$.
Damit ist der Folgesatz bewiesen.

Zur Analyse von (IV) (b) und, wie leicht zu zeigen ist, auch von (IV) (a), genügt es für diese Zwecke also, den Anfangszuwachs zu betrachten. Wesentlich für den Beweis dieses Folgesatzes sind die Bedingungen (II) (a) und (b). Ist etwa (II) (a) durch eine Funktion c verletzt, so kann der Fall eintreten, daß eine Person, die dieses c zur Berechnung von Wahrscheinlichkeiten benützt, zwar noch nicht beim ersten Erfahrungsdatum, wohl aber bei späteren lernt. Ist ferner (II) (b) durch ein c nicht erfüllt, so ist es möglich, daß eine dieses c benützende Person zwar bei einigen, nicht jedoch bei allen Prädikaten aus der Erfahrung lernt.
Da der Gegenstandsbereich B endlich ist, können All- bzw. Existenzbehauptungen als endliche Konjunktionen bzw. Adjunktionen dar-

gestellt werden. Es ist daher angenommen worden, daß die Sprache an logischen Zeichen nur die junktorenlogischen Verknüpfungen, nicht jedoch Quantoren enthält. Erweitert man nun B um m Dinge zu dem neuen Bereich B^* und fügt daher zur Sprache die neuen Konstanten $\beta_1, \ldots, \beta_m$ hinzu, so ändert sich der Wahrscheinlichkeitswert nicht für die ursprünglichen Aussagen, in denen an Gegenstandsausdrücken nur $\alpha_1, \ldots, \alpha_n$ vorkommen. Dies kommt durch den folgenden Satz zum Ausdruck:

Korollar 5

c sei eine adäquate induktive Methode. Ψ und Φ seien irgendwelche Sätze, wobei Φ erfüllbar sei. Dann ist $c(\Psi, \Phi)$ nur abhängig von den Gegenständen aus B, die in $\Psi \wedge \Phi$ erwähnt, d. h. durch Gegenstandsausdrücke bezeichnet werden, nicht jedoch von den übrigen Dingen. Der Wert von $c(\Psi, \Phi)$ ändert sich also nicht, wenn Gegenstände, die in $\Psi \wedge \Phi$ nicht erwähnt werden, aus B entfernt werden, oder aber wenn neue Gegenstände zu B hinzugefügt werden und wenn man die Menge der Gegenstandsausdrücke der Sprache entsprechend ändert.

Beweis

c sei eine adäquate induktive Methode, die also durch eine positive reelle Zahl L in der durch Theorem 14 gegebenen Form charakterisiert werden kann. Ψ sei ein Satz und Φ ein erfüllbarer Satz. $\gamma_1, \ldots, \gamma_{n^*}$ seien die in $\Psi \wedge \Phi$ vorkommenden Gegenstandsausdrücke; dabei ist $n^* \leqslant n$. Es wird nun diejenige Teilsprache betrachtet, die neben den logischen Konstanten und den Prädikaten $\Pi_1, \ldots, \Pi_p$ nur diese Gegenstandsausdrücke enthält. B^{n^*} sei der Teilbereich von B, der genau die durch diese Gegenstandsausdrücke bezeichneten Dinge enthält. Dann ist der Wert von $c(\Psi, \Phi)$ durch die Werte von $c(\Sigma^{n^*}, T)$ für jede Zustandsbeschreibung Σ^{n^*} dieser Teilsprache eindeutig bestimmt. Ihr werden nun weitere Gegenstandsausdrücke $\gamma_{n^*+1}, \ldots, \gamma_{n^*+m^*}$ hinzugefügt, die alle verschiedene Gegenstände bezeichnen; sofern $i \leqslant n$, bezeichnet dabei γ_i ein Ding aus B. Es genügt, zu zeigen, daß

in dieser so erweiterten Sprache jeder dieser Sätze Σ^{n*} (die dann keine Zustandsbeschreibungen mehr sind) durch c die gleiche Aprioriwahrscheinlichkeit erhält wie in der ursprünglichen Teilsprache.
Ein beliebiger dieser Sätze Σ^{n*} sei für das folgende fest vorgegeben. Alle und nur die Zustandsbeschreibungen der erweiterten Sprache von der Form $\Sigma^{n*} \wedge \Delta^{m*}$ implizieren logisch Σ^{n*}; es seien dies die Sätze $\Sigma^{n*} \wedge \Delta^{m*}_1, \ldots, \Sigma^{n*} \wedge \Delta^{m*}_{m*q}$. Also ist Σ^{n*} logisch äquivalent mit $(\Sigma^{n*} \wedge \Delta^{m*}_1) \vee \ldots \vee (\Sigma^{n*} \wedge \Delta^{m*}_{m*q})$ sowie mit $\Sigma^{n*} \wedge (\Delta^{m*}_1 \vee \ldots \vee \Delta^{m*}_{m*q})$, wobei diese Adjunktion $\Delta^{m*}_1 \vee \ldots \vee \Delta^{m*}_{m*q}$ logisch wahr ist. Da c unabhängig von der Sprache durch das gleiche L charakterisiert ist, ist c (Σ^{n*}, T) in dieser erweiterten Sprache mit dem gleichen Produkt von Brüchen identisch wie in jener zunächst betrachteten Teilsprache. Ist außerdem $n^* + m^* = n$, so sind die Zustandsbeschreibungen dieser erweiterten Sprache (bis auf die Reihenfolge der Konjunktionsglieder) identisch mit denen der ursprünglichen, über B interpretierten Sprache. Damit ist der Folgesatz bewiesen.

Korollar 6

c sei eine adäquate induktive Methode. Ψ und Φ seien irgendwelche Sätze, die nur Grundprädikate enthalten, wobei Φ erfüllbar ist. Dann ist c (Ψ, Φ) nur von den Eigenschaften abhängig, die in $\Psi \wedge \Phi$ erwähnt, also durch Eigenschaftsausdrücke bezeichnet werden. Der Wert von c (Ψ, Φ) ändert sich also nicht, wenn Eigenschaftsausdrücke, die in $\Psi \wedge \Phi$ nicht vorkommen, aus der Sprache entfernt werden, oder aber, wenn neue Eigenschaftsausdrücke der Sprache hinzugefügt werden.

Beweis

c sei eine adäquate induktive Methode und damit durch eine positive reelle Zahl L charakterisiert. Ψ sei ein Satz und Φ ein erfüllbarer Satz, wobei in beiden nur Grundprädikate vorkommen, nämlich $\Pi_{i_1}, \ldots, \Pi_{i_{p*}}$. Es wird diejenige Teilsprache betrachtet, die neben den logischen Konstanten und den Gegenstandsausdrücken $\alpha_1, \ldots, \alpha_n$ nur diese Prädikate enthält. Der Wert von c (Ψ, Φ) ist dann durch die

Werte von c (Σ^{p^*}, T) für jede Zustandsbeschreibung Σ^{p^*} dieser Teilsprache eindeutig bestimmt. Dieser Teilsprache werden nun die Prädikate $\Pi_{i_{p^*+1}}, \ldots, \Pi_{i_{p^*+p'}}$ hinzugefügt; sofern $j \leqslant p$, sei dabei Π_j eines der ursprünglichen Grundprädikate. Wiederum genügt es, zu zeigen, daß in dieser erweiterten Sprache jeder dieser Sätze Σ^{p^*} (die dann keine Zustandsbeschreibungen mehr sind) durch c die gleiche Aprioriwahrscheinlichkeit erhält wie in jener Teilsprache. Einer dieser Sätze Σ^{p^*} wird für das folgende festgehalten. $\Sigma^{p^*+p'}$ sei eine Zustandsbeschreibung der erweiterten Sprache, die Σ^{p^*} logisch impliziert; dann erhält man aus ihr durch Änderung der Reihenfolge der Konjunktionsglieder einen Satz der Form $\Sigma^{p^*} \wedge \Delta^{p'}$. Die Adjunktion aller dieser Sätze $\Delta^{p'}$ ist wiederum logisch wahr. Da c unabhängig von der Sprache durch das gleiche L charakterisiert ist, ist auch in dieser erweiterten Sprache c (Σ^{p^*}, T) mit dem gleichen Produkt von Brüchen identisch wie in jener zunächst betrachteten Teilsprache. Ist außerdem $p^* + p' = p$, so sind die Zustandsbeschreibungen dieser erweiterten Sprache (bis auf die Reihenfolge der Konjunktionsglieder) identisch mit denen der ursprünglichen Sprache. Damit ist auch dieser Folgesatz bewiesen.

Da in den Sätzen Ψ und Φ nur endlich viele Gegenstands- und Eigenschaftsausdrücke vorkommen (und nach Voraussetzung nur die junktorenlogischen Konstanten „nicht", „und", „oder", „wenn — dann" und „genau dann, wenn", nicht jedoch die Quantoren „jedes" und „es gibt" enthält), kann zur Berechnung von c (Ψ, Φ) bei einer adäquaten induktiven Methode c eine Teilsprache gewählt werden, die neben den junktorenlogischen Konstanten nur jene Ausdrücke enthält. Andererseits ist es oft unzweckmäßig, die Betrachtungen auf einen ganz bestimmten Bereich zu beschränken und auch nur endlich viele Eigenschaften für diesen Bereich in Erwägung zu ziehen. Da sich der c-Wert bei Erweiterungen des außerlogischen Vokabulars nicht ändert, darf dann unter Verwendung der unten charakterisierten Limeskonventionen angenommen werden, daß der Gegenstandsbereich und die Gesamtheit der Eigenschaften jeweils abzählbar unendlich sind (so daß also eine umkehrbar eindeutige Abbildung des Gegenstandsbereichs bzw. der Gesamtheit der Eigenschaften auf die Klasse der natürlichen Zahlen besteht). Es genügt dann also, *einen* der-

artigen Bereich und *eine* Gesamtheit von Eigenschaften anstelle einer Folge von endlichen Bereichen und einer Folge von endlichen Gesamtheiten von Eigenschaften den Überlegungen zugrundezulegen.
Gegeben sei eine Sprache mit abzählbar unendlich vielen Gegenstandsausdrücken; die Sprache ist daher, im Einklang mit der Voraussetzung aus dem 1. Abschnitt des III. Kapitels, daß eine umkehrbar eindeutige Abbildung der Gegenstandsausdrücke auf die Elemente von B besteht, über einem abzählbar unendlichen Gegenstandsbereich B interpretiert. Ein endlicher Abschnitt dieser Sprache für Gegenstandsausdrücke ist dann eine Teilsprache, die an Gegenstandsausdrücken alle Ausdrücke $\alpha_1, \ldots, \alpha_i$ für eine gewisse natürliche Zahl i, aber keinen Ausdruck a_j mit $j > i$, und ansonsten das gleiche Vokabular wie die vorgegebene Sprache enthält. Gegeben sei ferner eine Sprache mit abzählbar unendlich vielen voneinander logisch unabhängigen Grundprädikaten, die ja nach Voraussetzung Eigenschaftsausdrücke sind, kurz: mit abzählbar unendlich vielen einstelligen Grundprädikaten; die Gesamtheit der dadurch definierbaren V-Prädikate ist in diesem Fall gleichmächtig mit der Klasse der reellen Zahlen. Ein endlicher Abschnitt dieser Sprache für die Eigenschaftsausdrücke ist dann eine Teilsprache, die an einstelligen Grundprädikaten alle Ausdrücke $\Pi_1, \ldots, \Pi_i$ für eine gewisse natürliche Zahl i (und alle damit definierbaren V-Prädikate), aber kein einstelliges Grundprädikat Π_j mit $j > i$ (und kein dadurch definierbares V-Prädikat), ansonsten jedoch das gleiche Vokabular wie die vorgegebene Sprache enthält.
Es wird nun für Sprachen mit abzählbar unendlich vielen Gegenstandsausdrücken und/oder mit abzählbar unendlich vielen einstelligen Grundprädikaten gesagt, welche der Funktionen c, deren Definitionsbereich die Gesamtheit der geordneten Paare aus je einem Satz und einem erfüllbaren Satz und deren Wertbereich reelle Zahlen sind, adäquate induktive Methoden sind. Folgende Festsetzung liegt auf der Hand:

D 23 c ist eine adäquate induktive Methode (für eine Sprache mit unendlich vielen Gegenstandsausdrücken und/oder mit unendlich vielen einstelligen Beziehungsausdrücken) genau dann, wenn c für jeden endlichen Abschnitt dieser Sprache für Gegenstandsausdrücke und/oder für einstellige Grundprädikate

eine adäquate induktive Methode ist, dabei in der Normalform (vgl. Theorem 14 bzw. 15) durch die gleiche reelle Zahl L charakterisiert wird sowie die Limeskonvention für Gegenstandsausdrücke und/oder die für Eigenschaftsausdrücke erfüllt.

Diese beiden Limeskonventionen lassen sich, da man den Wert ∞ ja durch hinreichend große natürliche Zahlen beliebig approximieren kann, folgendermaßen charakterisieren:

D 24 c sei eine für alle endlichen Abschnitte der (abzählbar unendlich viele Gegenstandsausdrücke enthaltenden) Sprache eine adäquate induktive Methode und in der Normalform durch die reelle Zahl L determiniert.
Dann erfüllt c die Limeskonvention für Gegenstandsausdrücke genau dann, wenn für beliebige Sätze Ψ und Φ gilt: ${}_{\infty}c^{L}(\Psi, \Phi) = \lim_{i \to \infty} {}_{i}c^{L}(\Psi, \Phi)$, falls dieser Limes existiert; ansonsten hat ${}_{\infty}c^{L}(\Psi, \Phi)$ keinen Wert.
${}_{i}c^{L}$ ist dabei die durch L charakterisierte und auf den Abschnitt mit den Gegenstandsausdrücken $\alpha_1, \ldots, \alpha_i$ beschränkte Funktion c.

D 25 c sei eine für alle endlichen Abschnitte der (abzählbar unendlich viele einstellige Grundprädikate enthaltenden) Sprache eine adäquate induktive Methode und in der Normalform durch die reelle Zahl L determiniert.
Dann erfüllt c die Limeskonvention für Eigenschaftsausdrücke genau dann, wenn für beliebige Sätze Ψ und Φ gilt: ${}^{\infty}c^{L}(\Psi, \Phi) = \lim_{i \to \infty} {}^{i}c^{L}(\Psi, \Phi)$, falls dieser Limes existiert; ansonsten hat ${}^{\infty}c^{L}(\Psi, \Phi)$ keinen Wert.
${}^{i}c^{L}$ ist dabei die durch L charakterisierte und auf den Abschnitt mit den einstelligen Grundprädikaten $\Pi_1, \ldots, \Pi_i$ beschränkte Funktion c.

An sich hätte auch schon bisher statt „c" stets „${}^{p}_{n}c$" bzw. „${}^{p}_{n}c^{L}$" geschrieben werden müssen. Da jedoch in dieser Untersuchung bisher

aus allen Kontexten immer eindeutig hervorgegangen ist, wie groß die Gesamtheit der Gegenstandsausdrücke bzw. der einstelligen Grundprädikate ist, sind diese Indizes weggelassen worden. Dies geschieht auch im folgenden.
Eine Erweiterung dieses Ansatzes und insbesondere eine Charakterisierung des Wahrscheinlichkeitsbegriffs für Sprachen mit Relationsausdrücken wird im 1. Abschnitt des VI. Kapitels gegeben. Zunächst werden jedoch die bisher gewonnenen Ergebnisse vervollständigt und philosophisch ausgewertet.
$\Delta^m_{j_{k+1}}$ sei ein Satz, der an Gegenstandsausdrücken nur α_{k+1} enthält, also etwa die Aussage $\alpha_{k+1} \varepsilon \Pi^+$ mit einem Prädikat Π^+ der Weite m. Falls $m = 1$, so ist $\Delta^m_{j_{k+1}}$ ein V-Satz der beschriebenen Art. Aus den Theoremen 14 und 15 ergibt sich unmittelbar:

Korollar 7

c sei eine reguläre Bestätigungsmethode.
Dann ist c eine adäquate induktive Methode genau dann, wenn es eine positive reelle Zahl L gibt, so daß für alle natürlichen Zahlen k mit $0 \leqslant k < n$ und für alle natürlichen Zahlen j und m mit $0 \leqslant m \leqslant q$ (so daß k_j eine natürliche Zahl mit $0 \leqslant k_j \leqslant k$ ist und Φ_k sowie $\Delta^m_{j_{k+1}}$ Sätze der beschriebenen Art sind):

$$c\left(\Delta^m_{j_{k+1}}, \Phi_k\right) = \frac{k_j + \frac{m}{q} \cdot L}{k + L};$$

ist außerdem $0 < k$, so gilt für eine adäquate induktive Methode c:

$$c\left(\Delta^m_{j_{k+1}}, \Phi_k\right) = \frac{\frac{k_j}{k} \cdot k + \frac{m}{q} \cdot L}{k + L}.$$

Beweis

c sei eine reguläre Bestätigungsmethode.
(a) Die Implikation von rechts nach links ergibt sich unmittelbar aus Theorem 14 für $m = 1$.

(b) c sei eine adäquate induktive Methode. Es sei $0 \leqslant k < n$ und somit $0 \leqslant k_j \leqslant k$. $\Delta^m_{j_{k+1}}$ ist logisch äquivalent mit einer Adjunktion $\Psi_{j_1 k+1} \vee \ldots \vee \Psi_{j_m k+1}$ von V-Sätzen in α_{k+1}. Es ist ferner:

$k_j = \sum_{i=1}^{m} k_{j_i}$. Dann gilt:

$$
\begin{aligned}
c\,(\Delta^m_{j_{k+1}}, \Phi_k) &= c\,(\Psi_{j_1 k+1} \vee \ldots \vee \Psi_{j_m k+1}, \Phi_k) \\
&= c\,(\Psi_{j_1 k+1}, \Phi_k) + \ldots + c\,(\Psi_{j_m k+1}, \Phi_k) \\
&= \frac{\left(k_{j_1} + \frac{1}{q} \cdot L\right) + \ldots + \left(k_{j_m} + \frac{1}{q} \cdot L\right)}{k + L} \\
&= \frac{\sum_{i=1}^{m} k_{j_i} + m \cdot \frac{1}{q} \cdot L}{k + L} \\
&= \frac{k_j + \frac{m}{q} \cdot L}{k + L}
\end{aligned}
$$

Ist außerdem $0 < k$, so gilt:

$$
c\,(\Delta^m_{j_{k+1}}, \Phi_k) = \frac{\frac{k_j}{k} \cdot k + \frac{m}{q} \cdot L}{k + L}
$$

Durch die beiden folgenden Korollare wird gezeigt, daß (für gewisse Funktionen c) die Prämisse Φ_k des Wahrscheinlichkeitsarguments $c\,(\Delta^m_{j_{k+1}}, \Phi_k)$ durch schwächere Aussagen ersetzt werden kann, ohne daß dies den c-Wert berührt. Nach Folgesatz 9 ist von Φ_k nur jener Teil relevant, der besagt, welche Dinge die durch Π^+ ausgedrückte Eigenschaft haben, und nach Folgesatz 8 können individuelle Verteilungen durch statistische ersetzt werden (relevant ist für c also nur, *wieviele* Gegenstände die fragliche Eigenschaft haben, nicht jedoch, *welche* es sind).

Korollar 8

c erfülle die Bedingungen (I) und (II) (a). Φ sei eine individuelle Verteilung für Gegenstandsausdrücke $\beta_1, \ldots, \beta_k$, und Γ sei die entsprechende statistische Verteilung (d. h. Γ enthalte als erstes Adjunktionsglied den Satz Φ). In Ψ mögen keine der Gegenstandsausdrücke $\beta_1, \ldots, \beta_k$ vorkommen. Dann gilt: $c(\Psi, \Phi) = c(\Psi, \Gamma)$.

Beweis

Die Voraussetzungen seien gegeben.
Es sei Γ der Satz $\Phi_1 \vee \ldots \vee \Phi_j$, so daß Φ die Aussage Φ_1 ist. Φ' sei einer der Sätze $\Phi_2, \ldots, \Phi_j$, d. h. Φ' entstehe aus Φ durch Permutation der Gegenstandsausdrücke $\beta_1, \ldots, \beta_k$. Dann gilt wegen (II) (a): $c(\Phi, T) = c(\Phi', T)$, sowie: $c(\Psi \wedge \Phi, T) = c(\Psi \wedge \Phi', T)$. Es gilt folglich: $c(\Psi, \Phi_1) = c(\Psi, \Phi_2) = \ldots = c(\Psi, \Phi_j)$. Durch j—1-malige Anwendung von Theorem 5 (a) erhält man daraus:

$$\begin{aligned} c(\Psi, \Phi) &= c(\Psi, \Phi_1) \\ &= c(\Psi, \Phi_1 \vee \ldots \vee \Phi_j) \\ &= c(\Psi, \Gamma). \end{aligned}$$

Es sei $\Delta^m_{j_{k+1}}$ die Aussage $\alpha_{k+1} \varepsilon \Pi^+$ mit einem Prädikat, das die Weite m hat, und Φ^+_k sei ein Satz, der aus Φ_k auf folgende Weise entsteht: wenn eines seiner Konjunktionsglieder $\alpha_j \varepsilon \Pi^*_{i_j}$ den Satz $\alpha_j \varepsilon \Pi^+$ logisch impliziert, dann wird er durch $\alpha_j \varepsilon \Pi^+$ ersetzt; wenn er hingegen nicht $\alpha_j \varepsilon \Pi^*_{i_j}$ logisch impliziert (so daß er, da $\Pi^*_{i_j}$ ein V-Prädikat ist, $\neg(\alpha_j \varepsilon \Pi^*_{i_j})$ logisch impliziert), dann wird er durch $\neg(\alpha_j \varepsilon \Pi^+)$ ersetzt. Dann ergibt sich aus Folgesatz 7:

Korollar 9

c sei eine adäquate induktive Methode (und nach Theorem 14 bzw. 15 daher durch eine positive reelle Zahl L charakterisiert); k, m und k_j seien natürliche Zahlen mit $0 \leqslant k < n$, $0 \leqslant m \leqslant q$ und $0 \leqslant k_j \leqslant k$; $\Delta^m_{j_{k+1}}$, Φ_k und Φ^+_k seien Sätze der beschriebenen Art. Dann gilt: $c(\Delta^m_{j_{k+1}}, \Phi_k) = c(\Delta^m_{j_{k+1}}, \Phi^+_k)$.

Beweis

Die Voraussetzungen seien gegeben.
Φ^+_k ist logisch äquivalent mit der Adjunktion aller und nur der Sätze Φ'_k, die sich von Φ_k höchstens auf folgende Weise unterscheiden: Konjunktionsglieder $\alpha_j \in \Pi^*_{i_j}$ von Φ_k, die logisch $\alpha_j \in \Pi^+$ implizieren, können durch Sätze $\alpha_j \in \Pi^*_{i_j}$ ersetzt sein, die ebenfalls $\alpha_j \in \Pi^+$ logisch implizieren, und Konjunktionsglieder $\alpha_j \in \Pi^*_{i_j}$, die logisch $\neg (\alpha_j \in \Pi^+)$ implizieren, können durch Konjunktionsglieder $\alpha_j \in \Pi^+_{i_j}$ ersetzt sein, die ebenfalls $\neg (\alpha_j \in \Pi^+)$ logisch implizieren. Nach Korollar 7 hängt $c(\Delta^m_{j_{k+1}}, \Phi_k)$ bei festen k (und vorgegebenem L) nur von der Weite m von Π^+ und von der Anzahl k_j der k Gegenstände ab, die jene durch Π^+ ausgedrückte Eigenschaft haben. Folglich gilt: $c(\Delta^m_{j_{k+1}}, \Phi_k) = c(\Delta^m_{j_{k+1}}, \Phi'_k)$. Mit mehrmaliger Anwendung von Theorem 5 (a) erhält man daraus:
$c(\Delta^m_{j_{k+1}}, \Phi_k) = c(\Delta^m_{j_{k+1}}, \Phi^+_k)$.

Statt „c ist eine adäquate induktive Methode" wird im folgenden des öfteren gesagt: „c erfüllt die Bedingungen (I) bis (IV)", „c ist eine L-Funktion", „c ist im L-System enthalten" oder „c ist im Kontinuum enthalten". An einigen Stellen wird die Bedingung, daß c eine L-Funktion ist, auch durch „c^L" bezeichnet.

4. Erkenntnistheoretische Konsequenzen aus dieser Normalform

Über diese Darstellbarkeit des singulären Voraussageschlusses[13] lassen sich artige Betrachtungen anstellen, die ohne Zweifel erhebliche Folgen in Ansehung der logischen Form aller Wahrscheinlichkeitsschlüsse haben werden.

[13] Ein singulärer Voraussageschluß ist ein induktives Argument der in Theorem 14 und Korollar 7 benützten Form. Auf ihn können alle anderen induktiven Schlüsse zurückgeführt werden, da ja nach den Bemerkungen zu Beginn des vorigen Abschnittes die Aprioriwahrscheinlichkeit einer Zustandsbeschreibung das Produkt aus n derartigen singulären Voraussageschlüssen ist und nach D_2 und D_3 alle induktiven Argumente auf Aprioriwahrscheinlichkeiten von Zustandsbeschreibungen reduzierbar sind. Vgl. hierzu auch Carnap-Stegmüller [1], S. 81f., 226ff.

Der Zähler dieses Bruches ist demnach die gewogene Summe zweier Brüche. Der erste Bruch ist dabei die beobachtete relative Häufigkeit k_j/k, die mit dem empirischen Faktor k (der Anzahl der untersuchten Objekte) gewogen wird, und der zweite ist die apriorisch bestimmte Weite des Satzes $\Delta^m_{j_{k+1}}$ und des darin vorkommenden Prädikates Π^+, die mit dem logischen (besser: dem apriorischen) Faktor L gewogen wird. Gemittelt wird diese Summe durch die Summe aus dem empirischen und dem logischen Faktor. Zugrundegelegt wird dabei die Erhebung Φ_k, die die ersten k Individuen in allen Einzelheiten beschreibt, oder aber $\Phi^+{}_k$, die die ersten k Individuen nur im Hinblick auf die durch Π^+ ausgedrückte Eigenschaft beschreibt und also eine individuelle Verteilung für die Gegenstandsausdrücke $\alpha_1, \ldots, \alpha_k$ bezüglich Π^+ ist, oder schließlich die der individuellen Verteilung $\Phi^+{}_k$ entsprechende statistische Verteilung Γ_k bezüglich Π^+ (so daß $\Phi^+{}_k$ Adjunktionsglied von Γ_k ist).
Bei vorgegebenem Φ_k und $\Delta^m_{j_{k+1}}$ sind k und k_j fest vorgegeben. Der Wahrscheinlichkeitswert hängt dann lediglich von der Wahl von L ab. Ist $L \ll k$, so ist

$$\frac{\frac{k_j}{k}\cdot k + \frac{m}{q}\cdot L}{k+L} \approx \frac{\frac{k_j}{k}\cdot k}{k} = \frac{k_j}{k};$$

dann entspricht die Wahrscheinlichkeit annähernd der beobachteten relativen Häufigkeit. Ist hingegen $L \gg k$, so ist

$$\frac{\frac{k_j}{k}\cdot k + \frac{m}{q}\cdot L}{k+L} \approx \frac{\frac{m}{q}\cdot L}{L} = \frac{m}{q};$$

dann ist also die Wahrscheinlichkeit von $\Delta^m_{j_{k+1}}$ auf Grund des Datums Φ_k fast gleich der Aprioriwahrscheinlichkeit von $\Delta^m_{j_{k+1}}$; die Erfahrung wird dann also praktisch nicht berücksichtigt, sondern lediglich das Faktum, wie nahe $\Delta^m_{j_{k+1}}$ bei der Tautologie oder bei der Kontradiktion liegt, d. h. wo die damit charakterisierte Eigenschaft auf Grund eben dieser Charakterisierung zwischen den beiden Extremen der universellen und der leeren Eigenschaft liegt. Ist weder $L \ll k$ noch $L \gg k$, so fallen beide Faktoren bei der Ermittlung des Wahrscheinlichkeitswertes in Betracht.

L kann demnach als *Maß meiner Vorsicht gegenüber dem Beobachtungsdatum* betrachtet werden. Ist z. B. $L_1 < L_2$ und ist mein Wahrscheinlichkeitsbegriff durch L_1, der meines Freundes hingegen durch L_2 bestimmt, so kann ich früher den logischen Faktor vernachlässigen, als er es tun wird. Auf Grund dreier Versuche, die das gleiche Ergebnis haben, und keiner weiteren Vorkenntnis (z. B. bezüglich einer chemischen Substanz oder eines Würfels) werde *ich* eventuell zu der Überzeugung kommen, daß auch ein beliebiger vierter Versuch dieses Ergebnis haben wird, während *er* diesen Schluß für voreilig hält. Mit wachsendem k werden sich jedoch unsere c-Werte trotz verschiedenem L immer mehr annähern und können bei sehr großem k (relativ zu L_1 und L_2) als praktisch gleich angesehen werden. Je nachdem, wie der zugrundegelegte Gegenstandsbereich B beschaffen ist, kann sich nachträglich erweisen, daß meine Voreiligkeit oder seine Vorsicht unberechtigt gewesen sind. Da andererseits L, m und q konstant sind, während k mit zunehmenden Beobachtungen wächst, gleichen sich die beiden c-Werte trotz verschiedenem L immer mehr einander an. Ist $n \gg L_2$, so werden sie sich bei einem großen k nur noch um einen minimalen Betrag u voneinander unterscheiden, der in der praktischen Anwendung vernachlässigt werden kann.

L kann nicht nur als Maßstab meiner Vorsicht gegenüber dem Beobachtungsdatum, sondern auch als *Gradmesser meines Vertrauens in die a priori vorgenommene Bestimmung der relativen Weite* der Hypothese angesehen werden. Habe ich etwa einen Würfel auf seine physikalische Beschaffenheit hin untersucht und festgestellt, daß seine Flächen etwa gleich sind und sein Schwerpunkt etwa in der Mitte liegt, wobei die Abweichungen so minimal sind, daß ich sie vernachlässigen kann, und setze ich voraus, daß das Fallen des Würfels weder durch Magnetismus noch durch andere Störungen beeinflußt wird, da ich solche nicht beobachten kann, so handle ich vernünftig, wenn ich das L sehr groß wähle und mich damit durch eine zufällige Folge von Ergebnissen beim Würfeln nicht beeinflussen lasse. Habe ich etwa vor, nicht öfter als 10^6 Mal mit ihm zu würfeln, so kann sich die Wahl von $L = 10^{10}$ als zweckmäßig erweisen. Wenn bei den ersten 1000 Würfen dann vielleicht 970mal die Ziffer „6" oben liegt, so werde ich mich dadurch nicht irritieren lassen; ein solches Ergebnis ist zwar unter Verwendung dieser induktiven Methode a priori unwahr-

scheinlich, aber möglich, und ich werde sagen, daß es eben *zufällig* eingetreten ist. Dieses Beobachtungsdatum wird mich jedoch nicht dazu verleiten, der Hypothese, daß der nächste Wurf wiederum eine „6" ergibt, eine deutlich höhere Wahrscheinlichkeit zuzusprechen als der, daß eine „5" oben liegen wird.

Anders verhält es sich, wenn ich über den Würfel nichts weiß und daher zunächst einmal annehme, daß er korrekt und regelmäßig ist, wenn ich dieser Annahme jedoch nicht sehr traue, sondern vielmehr vermute, daß in einer (mir nicht bekannten) Seite eine Metallplatte eingebaut ist und in dem Wurfmechanismus sich ein Magnet befindet. Dann werde ich die Bewertung der relativen Weite der Prädikate nur vornehmen, um überhaupt anfangen zu können, und werde versuchen, aus der Erfahrung zu lernen, wie der Würfel beeinflußt wird und auf welcher Seite sich gegebenenfalls die Eisenplatte befindet. Das obige Datum wird dann für mich ein so starkes Argument sein, daß die relativen Weiten der Prädikate bei der Berechnung der Wahrscheinlichkeit fast keine Rolle mehr spielen und die Wahrscheinlichkeit praktisch identisch mit der beobachteten relativen Häufigkeit ist. Ich kann in einem solchen Fall $L = 2$ oder $L = 1$ wählen oder, wenn ich überhaupt kein Vertrauen in die relativen Weiten habe, vielleicht $L = {}^1/_{1000}$. Der Faktor L und die relativen Weiten können dann schon bei sehr kleinem k vernachlässigt werden, da sich der Wahrscheinlichkeitswert dann von der beobachteten relativen Häufigkeit praktisch nicht mehr unterscheidet. Wenn mein *Vertrauen in die Uniformität des Gegenstandsbereichs* berechtigt ist, so werde ich mit einem solchen Wahrscheinlichkeitsbegriff in der praktischen Anwendung (z. B. bei Wetten) wesentlich mehr Erfolg haben als mit der durch $L = 10^{10}$ charakterisierten induktiven Methode. Mit dieser allerdings wird der Erfolg größer sein, falls sich mein Vertrauen als trügerisch erweist.

Es besteht also ein Zusammenhang zwischen dem Uniformitätsgrad, den ein Gegenstandsbereich im Hinblick auf die durch die V-Prädikate bezeichneten Eigenschaften hat, und der optimalen (der fairen) induktiven Methode im Hinblick auf einen derartigen Uniformitätsgrad. Die Gesamtheit der Würfe, die mit einer regelmäßigen oder fast regelmäßigen Münze gemacht werden, bildet ohne Zweifel einen Gegenstandsbereich mit minimaler oder fast minimaler Uniformität.

Gewissen Gegenstandsbereichen der Physik wird hinsichtlich einiger Eigenschaften maximale Uniformität zugeschrieben (und wenn ein physikalisches Experiment dann doch einmal ein gegenteiliges Ergebnis hat, so vernachlässigt man es mit der Begründung, es sei durch unbekannte Störungen herbeigeführt worden). In anderen Fällen liegt der Uniformitätsgrad zwischen diesen Extremen; Lebewesen etwa haben hinsichtlich biologischer oder psychologischer Merkmale größere oder geringere Uniformität, ohne daß man deshalb vollständige Uniformität oder vollständige Disuniformität erwarten darf.
Je nachdem, welchen Uniformitätsgrad man kennt oder vermutet, wird man bei der Wahl einer Methode c^L einen großen oder geringen Wert L annehmen. Wenn ich etwa davon überzeugt bin, Würfe mit einer *idealen Münze* zu machen (die geometrisch ebenmäßig ist, deren Schwerpunkt mit dem geometrischen Mittelpunkt identisch ist usw.), so erscheint es als sinnvoll, die Wahrscheinlichkeit mit c^∞ zu messen, und umgekehrt ist offenbar die Verwendung von c^∞ auch *nur* dann vernünftig, wenn die Münze (und analog: der Würfel usw.) ideal ist; die Verwendung von c^∞ ist demnach äquivalent mit dem Postulat des durchgehenden Zufalls (bzw. im Einzelfall: der idealen Münze, des idealen Würfels usw.). Wenn ich hingegen davon ausgehen kann, daß ich mit gewissen physikalischen Dingen *störungsfrei* experimentieren kann und daß im Bereich dieser Dinge hinsichtlich der zu untersuchenden Eigenschaften vollständige Uniformität herrscht (daß Störungen zu anderen Ergebnissen führen, heißt ja nichts anderes, als daß unter den beschriebenen Umständen nicht alle Experimente das gleiche Ergebnis haben), dann ist es vernünftig, die Wahrscheinlichkeit mit c^0 zu messen und umgekehrt; die Verwendung von c^0 ist demnach äquivalent mit dem Postulat des störungsfreien Experiments. Wenn ich hingegen annehme, daß der Uniformitätsgrad weder minimal noch maximal ist, daß also weder alles Zufall ist, noch, daß der Gegenstandsbereich hinsichtlich der in Erwägung gezogenen Eigenschaften bereits vollständig determiniert ist und Störungen in Experimenten daher ausgeschlossen sind, so werde ich optimal mit einer induktiven Methode c^L mit $0 < L < \infty$ operieren; eine vorläufig noch offene Frage ist dabei, in welchem Verhältnis der Uniformitätsgrad w zum Parameter L steht.
Obwohl es nicht vernünftig erscheint, a priori die vollständige Uni-

formität eines Gegenstandsbereichs hinsichtlich der durch die V-Prädikate bezeichneten Eigenschaften oder aber seine vollständige Disuniformität anzunehmen, ist es doch nicht logisch ausgeschlossen, daß sein Uniformitätsgrad maximal bzw. minimal ist, so daß die Methoden c^0 bzw. c^∞ für ihn optimal sind.

Die durch $L = 0$ charakterisierte induktive Methode c^0 ist nicht im Kontinuum enthalten; sie kann jedoch durch beliebig kleine L beliebig angenähert werden. Es ist daher zweckmäßig, sie ebenfalls in Betracht zu ziehen. Unter Bezugnahme auf die L-Methoden kann sie wie folgt definiert werden:

$$c^0(\Delta^m_{j_{k+1}}, \Phi_k) = \lim_{L \to 0} c^L(\Delta^m_{j_{k+1}}, \Phi_k)$$

$$= \lim_{L \to 0} \frac{k_j + \frac{m}{q} \cdot L}{k + L}$$

Die Werte dieser Funktion ergeben sich dann auf einfache Weise:

$$c^0(\Delta^m_{j_{k+1}}, \Phi_k) = \begin{cases} k_j/k, \text{ falls } k > 0, \text{ und} \\ m/q, \text{ falls } k = 0. \end{cases}$$

Durch c^0 können also die Wahrscheinlichkeitswerte besonders einfach berechnet werden, weshalb diese Methode von Statistikern häufig benutzt wird. Dazu sind sie berechtigt, wenn entweder $k = 0$ oder $k \gg 0$ und die Erhebung Φ_k als im hohen Grad repräsentativ für die gesamte Population B angesehen werden kann, so daß Vorsicht nicht am Platze ist und ein L gewählt werden darf mit $k \gg L$. Für andere Werte k und für andere Erfahrungsdaten Φ_k interessieren sich Statistiker im allgemeinen nicht.

Für den Fall, daß $k \neq 0$ und daß nicht $k \gg 0$, führt c^0 jedoch zu inadäquaten Ergebnissen. Sind Φ_1 und $\Delta^m_{j_2}$ etwa die Sätze $\alpha_1 \varepsilon \Pi^*_1$ und $\alpha_2 \varepsilon \Pi^*_1$, so gilt: $c^0(\Delta^m_{j_2}, \Phi_1) = 1$. Nun folgt jedoch $\alpha_2 \varepsilon \Pi^*_1$ nicht logisch aus $\alpha_1 \varepsilon \Pi^*_1$; also ist die Bedingung (I) (6) verletzt, so daß c^0 wegen des Korollars 2 auch nicht streng kohärent ist. Wenn man die Frage, ob eine Wette akzeptabel ist, durch c^0 entscheidet, so kann man demnach dazu verleitet werden, ein Wettsystem zu akzeptieren, daß zwar einen Gewinn, nicht jedoch einen Verlust ausschließt. c^0 verstößt darüberhinaus auch gegen die Bedingung (IV) (b), wie bei der Analyse des Begriffs des Lernens aus der Erfahrung gezeigt

worden ist; sie gehört daher nicht zu den adäquaten induktiven Methoden. Trotzdem wird durch sie, wie oben gezeigt worden ist, in vielen Fällen eine gute und für praktische Zwecke genügende Approximation des tatsächlichen Wahrscheinlichkeitswertes gegeben, so daß sie sich in solchen Fällen als sehr brauchbar erweist.

c^0 ist eine merkwürdige Mischung aus Häufigkeitsdefinition und der in der klassischen Wahrscheinlichkeitslehre üblichen Gleichmöglichkeitsdefinition. Nach der klassischen Wahrscheinlichkeitslehre ist die Wahrscheinlichkeit der Quotient aus der Anzahl der dem Ereignis günstigen Fälle zur Gesamtheit aller gleichmög'ichen Fälle. In der hier verwendete Terminologie kann dies folgendermaßen interpretiert werden: die Wahrscheinlichkeit eines Satzes $\Delta^m_{j_{k+1}}$, der an Gegenstandsausdrücken nur α_{k+1} enthält, ist identisch mit dem Quotienten aus der Summe der Wahrscheinlichkeiten der Sätze $\Psi_{j_1 k+1}$, ..., $\Psi_{j_m k+1}$ zur Summe der Wahrscheinlichkeiten der Sätze $\Psi_{j_1 k+1}$, ..., $\Psi_{j_q k+1}$, wobei die $\Psi_{j_i k+1}$ an Gegenstandsausdrücken jeweils nur α_{k+1} enthalten; sind die Sätze $\Psi_{j_1 k+1}$, ..., $\Psi_{j_q k+1}$ gleichwahrscheinlich, so ist demnach die Wahrscheinlichkeit von $\Delta^m_{j_{k+1}}$ identisch mit der relativen Weite m/q dieses Satzes. Das Übergleiten von dieser Deutung des Begriffs der Wahrscheinlichkeit, die für die Aprioriwahrscheinlichkeit charakteristisch ist, zur Häufigkeitsinterpretation, geschieht relativ kontinuierlich durch die L-Funktionen und abrupt durch c^0. Da die Klassiker der Wahrscheinlichkeitsrechnung meist nur diejenigen Fälle im Auge hatten, bei denen entweder $k \ll L$ oder $L \ll k$, so daß der Fehler, der durch Vernachlässigung des empirischen oder des apriorischen Teils jener Normalform entsteht, minimal ist, war dieses Übergleiten im Gegensatz zur Auffassung Richard von Mises keine Inkonsequenz.[14] Woran man nur, und hier allerdings mit Recht, Kritik üben kann und muß, ist die naive und unreflektierte Anwendung des Begriffs „gleichmöglich" in jenen Theorien.

Entwickelt man, wie es Pascal getan hat, die Wahrscheinlichkeitsrechnung an Glücksspielen, sei es am Roulette, sei es am Münzwurf oder am Wurf des regelmäßigen Würfels, so wird man jede beliebige endliche Folge von beobachteten Ereignissen als zufällig ansehen. Die

---

[14] Vgl. Mises [2], S. 84.

Wahrscheinlichkeit ist dann der Quotient aus der Zahl m der günstigen und der Zahl q der möglichen Fälle und mit der die Bedingung (IV) (b) verletzenden Funktion c^∞ identisch:

$$c^\infty(\Delta^m_{j_{k+1}}, \Phi_k) = \lim_{L\to\infty} c^L(\Delta^m_{j_{k+1}}, \Phi_k) = \lim_{L\to\infty} \frac{k_j + \frac{m}{q}\cdot L}{k+L} = \frac{m}{q}.$$

Während der Benützer der Methode c^0 voraussetzt, daß die Welt völlig uniform ist, daß also in jedem Fall die tatsächliche relative Häufigkeit einer Eigenschaft in einem Gegenstandsbereich gleich der bis jetzt beobachteten relativen Häufigkeit ist, leugnet derjenige, der c^∞ verwendet, daß der in Frage kommende Bereich irgendeinen wenn auch noch so kleinen positiven Grad an Uniformität besitzt. Er ist davon überzeugt, daß die bisherigen Beobachtungen rein zufällige Ergebnisse sind und daß sich die relativen Häufigkeiten auf lange Sicht, wenn überhaupt, dann auf den Wert der Aprioriwahrscheinlichkeit einpendeln. Da k stets eine endliche Anzahl ist, sind die Werte k_j und k für die Berechnung des c^∞-Wertes irrelevant; dieser ist vielmehr durch die relative Weite m/q des Prädikates vollständig bestimmt. Trotz dieser Voraussetzung, die in den meisten, wenn nicht gar in allen Fällen unzutreffend sein dürfte, können die c^∞-Werte in vielen Fällen als gute Approximationen von c^L-Werten verwendet werden, z. B. beim Wurf eines Würfels, dessen Flächen und Material wir gut kennen. Wenn wir im Grad $10^6/(10^6 + 1)$ davon überzeugt sind, daß wir damit alle für den Wurf relevanten Merkmale zur Verfügung haben und diese durch die relativen Weiten der Prädikate ausgedrückt haben, so daß nach unserer Überzeugung alles übrige zufällig ist, dann werden wir $L = 10^6$ wählen; wenn wir weiterhin nicht mehr als 1000 Experimente mit diesem Würfel machen wollen, so wird die dadurch charakterisierte Funktion Werte liefern, die von den c^∞-Werten erst an Stellen hinter dem Komma abweichen, die wir sowieso vernachlässigen.

Neben der Tatsache, daß c^∞ bei Glücksspielen approximativ fair genannt werden darf, haben Ramsey, Wittgenstein und andere diese Funktion vor allem deshalb als Explikat für den Wahrscheinlichkeitsbegriff verwendet, weil ihre Definition besonders einfach ist und scheinbar nichts über die Natur voraussetzt: c^∞ ist nämlich dadurch

charakterisiert, daß durch sie alle Zustandsbeschreibungen (alle individuellen Verteilungen) die gleiche Anfangswahrscheinlichkeit erhalten.[15] Eine Funktion $c^\dagger$ mit diesen Eigenschaften kann folgendermaßen charakterisiert werden:

D 26 Die Anzahl der Gegenstandsausdrücke und der einstelligen Grundprädikate sei endlich.

(a) Ψ sei eine Zustandsbeschreibung und Φ sei logisch wahr; dann gilt: $c^\dagger(\Psi, \Phi) = 1/m$, wobei m die Anzahl der Zustandsbeschreibungen ist.

(b) Ψ sei ein Satz, aber keine Zustandsbeschreibung, und Φ sei logisch wahr; dann gilt: $c^\dagger(\Psi, \Phi) = \sum_{\Sigma \in Zb^\Psi} c^\dagger(\Sigma, \Phi)$.

(c) Ψ sei ein Satz und Φ sei erfüllbar, aber nicht logisch wahr; dann gilt:

$$c^\dagger(\Psi, \Phi) = \frac{c^\dagger(\Psi \wedge \Phi, T)}{c^\dagger(\Phi, T)}.$$

Durch diese Funktion $c^\dagger$ wird der Wahrscheinlichkeitswert durchgehend ohne Rücksichtnahme auf etwaige Erfahrungen bestimmt; denn es gilt:

Theorem 16

Die Anzahl der Gegenstandsausdrücke und der einstelligen Grundprädikate sei endlich. Dann ist $c^\dagger$ mit c^∞ identisch.

Beweis

Die Anzahl der Gegenstandsausdrücke und der einstelligen Grundprädikate sei endlich. Es genügt zu zeigen, daß $c^\dagger$ und c^∞ den Zustandsbeschreibungen gleiche Aprioriwerte zuordnen; nach D 2 und D 3 bzw. nach D 25 (b) und (c) sind dann auch die anderen Werte gleich.

[15] Vgl. das auf S. 148 gegebene Beispiel; vgl. ferner Carnap [11], S. 37ff., sowie Carnap-Stegmüller [1], S. 157f. und 220f.

Eine Zustandsbeschreibung Σ ist nun aber eine Konjunktion $\Psi_{i_1} \wedge \ldots \wedge \Psi_{i_n}$ der beschriebenen Art. Durch n-malige Anwendung der Bedingung (5) von D 4 erhält man:
$c^\infty (\Psi_{i_1} \wedge \ldots \wedge \Psi_{i_n}, T) = c^\infty (\Psi_{i_1}, T) \cdot \ldots \cdot c^\infty (\Psi_{i_n}, \Phi_{k-1})$.
Unter Berücksichtigung der Tatsache, daß
$c^\infty (\Psi_{i_1}, T) = \ldots = c^\infty (\Psi_{i_n}, \Phi_{k-1}) = 1/q$,
erhält man daraus: $c^\infty (\Sigma, T) = (1/q)^n$. Andererseits ordnet $c^\dagger$ den Zustandsbeschreibungen gleiche Aprioriwerte zu. Wie man sich leicht klarmachen kann, ist die Anzahl der Zustandsbeschreibungen $m = q^n$. Folglich gilt: $c^\dagger (\Sigma, T) = 1/q^n = (1/q)^n$, womit der Lehrsatz bewiesen ist.

Offenbar bleibt dieses Ergebnis auch erhalten, wenn man D 25 für Sprachen mit je abzählbar unendlich vielen Gegenstandsausdrücken und einstelligen Grundprädikaten erweitert. Die Funktion $c^\dagger$ macht also nur scheinbar keine Annahmen über die Uniformität der Welt; in Wirklichkeit setzt sie vollständige Disuniformität voraus.
Da diese auf der Hand liegende Definition des Wahrscheinlichkeitsbegriffs also zu Inadäquatheiten führt, liegt es nahe, sie folgendermaßen zu ändern: nicht die individuellen, sondern die statistischen Verteilungen erhalten gleiche Anfangswahrscheinlichkeiten. Eine statistische Verteilung Γ sagt ja lediglich aus, *wieviele* der Individuen die einzelnen Eigenschaften haben, nicht jedoch, *welche* Gegenstände unter die betreffenden Eigenschaften fallen; sie ist also darstellbar als die Standardadjunktion aller jener individuellen Verteilungen, die den einzelnen Prädikaten dieselben relativen Häufigkeiten zuschreiben wie Φ, aber noch zusätzlich ausdrücken, welche der Individuen die entsprechenden Eigenschaften besitzen. Da die einzelnen individuellen Verteilungen, deren Standardadjunktion eine statistische Verteilung ist, miteinander logisch unverträglich sind, ist es sinnvoll, weiterhin festzulegen, daß alle diese Zustandsbeschreibungen die gleichen Maße erhalten und ihre Summe identisch mit der Aprioriwahrscheinlichkeit der statistischen Verteilung ist. Diese Festlegung charakterisiert die Funktion c^*, die Carnap zunächst als Explikat für den Wahrscheinlichkeitsbegriff vorgeschlagen hat.[16] c^* nimmt die folgende Stelle im Kontinuum ein:

[16] Vgl. Carnap [9], S. 562ff., Carnap [11], S. 44ff., sowie Carnap-Stegmüller [1], S. 223ff.

$$c^*(\Delta^m_{j_{k+1}}, \Phi_k) = \frac{k_j + m}{k + q} = \frac{k_j + \frac{m}{q} \cdot q}{k + q}.$$

Obwohl bei dieser Definition also zunächst scheinbar ebenfalls keine Voraussetzungen über die Uniformität des Gegenstandsbereichs gemacht worden sind, erweist es sich, daß die relative Weite der Prädikate mit dem Gewicht $L = q$ belegt (und daher, wie aus den Ergebnissen des nächsten Abschnittes folgt, der Uniformitätsgrad $w = 2/(1 + q)$ vorausgesetzt wird). Für den Fall, daß $p = 1$, daß also nur ein Grundprädikat vorliegt, oder, was damit gleichbedeutend ist, daß $q = 2$, wobei die beiden Prädikate eine Familie bilden, gilt für Grundprädikate, deren relative Weite ja $1/2$ ist:

$$c^*(\Delta^1_{j_{k+1}}, \Phi_k) = \frac{k_j + 1}{k + 2}.$$

Dies ist jedoch gerade die Regel, die Laplace zur Berechnung von Wahrscheinlichkeiten angegeben hat. Man wird mit ihr (bzw. allgemeiner mit c^* für kleine p) wohl meist akzeptable Ergebnisse erhalten. Das gilt aber nicht für große p bzw. für große q. Denn angenommen, die Grundprädikate bilden eine Familie (eine Division), so daß auf Grund von Postulaten, die die Bedeutung der Grundprädikate teilweise beschreiben, jedes Ding eine und nur eine der Grundeigenschaften hat und damit alle bis auf p der q V-Prädikate leere Eigenschaften bezeichnen (dies wird im folgenden dadurch ausgedrückt, daß $p = q$ gesetzt wird, da die Grundprädikate einen Gegenstand dann bereits vollständig beschreiben und damit ein Übergang zu den V-Prädikaten überflüssig ist); angenommen ferner, es handelt sich dabei um Bezeichnungen für die Intervalle eines Metermaßes, das in Millimeter eingeteilt ist. Dann muß $q = 1000$ gesetzt werden, so daß z. B. die ersten 100 Messungen keinen nennenswerten Einfluß auf den c^*-Wert haben. Das gilt a fortiori für ein 100-m-Band mit mm-Einteilung bzw. für einen noch detaillierteren Maßstab. Solche Erweiterungen und gegebenenfalls Verfeinerungen des Meßverfahrens und die damit verbundene Aufteilung der Prädikate müssen jedoch vorgenommen werden, und man sollte mit Recht erwarten, daß damit keine Änderung der Schätzung der Homogenität verbunden ist.

Darüberhinaus verlangt der Naturwissenschaftler als Idealisierung, daß jene Verfeinerung beliebig fortgesetzt werden kann, so daß $L = q$ beliebig groß wird; als Grenzwert für $p \to \infty$ wird dann c^* mit c^∞ identisch.

Es setzt also jede Verwendung einer induktiven Methode aus dem Kontinuum eine Schätzung der Homogenität des den Untersuchungen zugrundegelegten Gegenstandsbereichs voraus. Diese kann, je nach den Vorkenntnissen, gut oder weniger gut vorgenommen werden. Es erhebt sich die Frage, wie man sie vornehmen soll, wenn über den Bereich (außer dessen Anzahl) überhaupt nichts bekannt ist. Da Aussagen $\alpha \varepsilon \Pi$ mit dem Grundprädikat Π durch alle Methoden c^L den Aprioriwert $1/2$ erhalten, wäre es sinnvoll und vernünftig, in diesem Fall ebenfalls das arithmetische Mittel aus maximaler und minimaler Uniformität zu nehmen, das sich, wie im nächsten Abschnitt gezeigt wird, durch: $w = (1 + 1/q)/2$ ergibt; mit Theorem 17 würde man dann daraus den Wert $L = 1$ erhalten. Dies ist jedoch kein Argument dafür, immer und in allen Fällen c^1 zu verwenden, da man häufig, wie etwa beim Münzwurf oder aber bei physikalischen Experimenten, größere Kenntnisse über den Gegenstandsbereich besitzt, so daß sich dann eine andere Wahl empfiehlt.

Mit seiner Explikation des logischen Wahrscheinlichkeitsbegriffs hat Carnap drei Ziele verfolgt:

(1) Eine Bestätigungsmethode c ist als eine logische Funktion einzuführen, deren Funktionswert bei einem vorgegebenen Paar von Argumenten ohne Zuhilfenahme von empirischem Wissen oder empirischen Hypothesen berechnet werden kann; diese Funktion soll dabei eine partielle logische Implikation kennzeichnen, d. h. sie soll als Grenzfall die logische Folgerung einerseits und die logische Unverträglichkeit andererseits enthalten. — Dieses Ziel hat Carnap ohne Zweifel erreicht. Charakterisiert man definitorisch irgendeine Funktion aus dem Kontinuum — sei es direkt, sei es über die Normalform — so ist für ein vorgegebenes Paar, bestehend aus einem Satz und einem erfüllbaren Satz, der Funktionswert logisch determiniert. Ein zusätzliches Wissen ist nicht erforderlich; dieses sollte vielmehr bereits in dem erfüllbaren Satz, der das Erfahrungsdatum beschreibt, enthalten sein.

(2) Mit den verschiedenen Adäquatheitsbedingungen hat Carnap zu-

nächst außerdem das Ziel verfolgt, die Klasse der in Frage kommenden Funktionen immer weiter einzuengen, so daß schließlich nur noch die Einerklasse einer Funktion übrigbleibt. Diese Funktion könnte dann *die* adäquate Explikation *des* Wahrscheinlichkeitsbegriffs genannt werden. — Dieses Ziel hat Carnap nicht erreicht, und es ist auch nicht erreichbar. Denn wenn die Fairneß von c^L davon abhängt, welchen Uniformitätsgrad der Gegenstandsbereich besitzt, dieser Uniformitätsgrad aber verschieden sein kann, dann muß es auch verschiedene faire Bestätigungsmethoden c^L geben, und dann ist es nicht möglich, weitere tatsächlich adäquate Adäquatheitsbedingungen anzugeben; denn solche würden ja Funktionen ausschließen, die man in einem Gegenstandsbereich, mit dem zu rechnen ist, als fair ansehen muß. Man muß sich daher mit dem Gedanken abfinden, daß es in der induktiven Logik, anders als in der klassischen deduktiven Logik, mehrere Systeme des Schließens gibt, die miteinander in einigen Fällen übereinstimmen, z. B. in den Grenzfällen, in denen logische Folgerungen oder logische Unverträglichkeiten vorliegen, sonst aber in größerem oder geringerem Umfang verschiedene Zahlenwerte liefern. *Die* adäquate induktive Methode gibt es nicht, sondern nur ein Kontinuum solcher adäquater Wahrscheinlichkeitsbegriffe.

(3) Schließlich hat Carnap zunächst mit seiner Wahrscheinlichkeitstheorie noch den Zweck verfolgt, einen Begriff der partiellen Implikation zu charakterisieren, der, wie der logische Folgerungsbegriff, keinerlei Voraussetzungen über die faktische Welt beinhaltet. — Dieses Ziel hat er nicht erreicht, und es ist ebenfalls nicht erreichbar; denn wenn eine Funktion c eine adäquate induktive Methode ist und also im Kontinuum liegt, dann kann sie durch eine positive reelle Zahl L eindeutig charakterisiert werden. Die durch die Gleichung: $w = (1 + L/q)/(1 + L)$ des Theorems 17 bestimmte Zahl w ist die Voraussetzung über den Uniformitätsgrad, die man zu machen hat, um die induktive Methode c^L als optimal und fair ansehen zu können. Diese Voraussetzung ist zwar eine *Aprioriannahme*, da sie der Charakterisierung und Anwendung dieser induktiven Methode vorangeht, sie ist jedoch ohne Zweifel eine *synthetische Hypothese* über die Welt und kann auch eine Fehleinschätzung sein.

Daß die beiden letzten Ziele nicht erreichbar sind, hat im übrigen

als erster Carnap selbst gesehen.[17] Daß man seine naheliegenden Vermutungen aber überhaupt klar und eindeutig hat widerlegen können, setzt einen präzisen und hinreichend entwickelten Begriffsapparat voraus, den in dieser Art ebenfalls erst er entwickelt hat.

Zu diskutieren bleibt noch die Frage, ob Kant mit seiner Behauptung, daß wir überhaupt erst auf Grund von synthetisch-apriorischen Voraussetzungen zu Wirklichkeitserkenntnissen gelangen, nicht wenigstens teilweise recht hat; denn die Aussage, daß ein vorgegebener Gegenstandsbereich hinsichtlich gewisser Eigenschaften einen Uniformitätsgrad w hat, ist ohne Zweifel kein analytisches, sondern ein synthetisches Urteil, dessen Wahrheit wir darüberhinaus, solange wir den Gegenstandsbereich (z. B. unsere faktische Welt) nicht vollständig auf die fraglichen Eigenschaften hin untersucht haben, nie sicher wissen, sondern bestenfalls hypothetisch annehmen können (induktiv erschließen können wir diese Behauptung deshalb nicht, weil wir sie ja zur Rechtfertigung eben dieser induktiven Schlüsse benötigen, wie schon Hume erkannt hat). Andererseits müssen wir eine derartige Aussage a priori als wahr annehmen, da unser Wissen sonst auf unsere Beobachtungen und deren deduktive Konsequenzen beschränkt bleibt. Offenbar setzen induktive Argumente also synthetisch-apriorische Urteile über die Welt voraus.

Wenn man, wie Kant, Mill, Laplace und zum Teil auch Hume, den Uniformitätsgrad als klassifikatorischen Begriff benützt und daher nur zwischen Uniformität und Disuniformität unterscheidet, hat man zwei Möglichkeiten: Entweder man nimmt die Disuniformität der Welt an und ist damit genötigt, auf jedes induktive Argumentieren, das die Erfahrung berücksichtigt, zu verzichten und sich damit auf die Erfahrung und deren deduktive Konsequenzen zu beschränken; oder man setzt die Uniformität voraus und kann dann induktive Methoden benützen. Jede Art von Induktion setzt demnach apodiktisch Uniformität voraus.

Diese Konklusion wird hier vermieden, da hier nicht nur maximale und minimale Uniformität, sondern auch alle dazwischenliegenden Grade berücksichtigt werden. Um eine adäquate induktive Methode

[17] Vgl. hierzu seine Bemerkungen in Carnap [11], S. 59—79, sowie seine Erwiderung auf die reichlich unklaren Ausführungen von Burks in Schilpp [1], S. 980—983.

anwenden zu können, ist niemand genötigt, die maximale Uniformität des Gegenstandsbereichs vorauszusetzen. Er braucht auch nicht *apodiktisch* irgend einen anderen Uniformitätsgrad w anzunehmen. Es genügt, wenn er ihn *hypothetisch* voraussetzt. Mit dieser Annahme ist dann nach Theorem 17 die Wahl einer Funktion c^L verbunden, aber diese Wahl ist dann genauso hypothetisch und genausowenig endgültig wie die des Uniformitätsgrades w. Wenn man aus irgendeinem (rationalen oder irrationalen) Grund vermutet, daß nicht w, sondern w′ der Uniformitätsgrad des Gegenstandsbereichs hinsichtlich seiner Eigenschaften ist, so wird man entsprechend statt c^L die Methode $c^{L'}$ benützen.

Es gibt also nicht *einen* Uniformitätsgrad für *alle* Gegenstandsbereiche, sondern *jeder* Gegenstandsbereich hat *seinen eigenen* Uniformitätsgrad. Darüberhinaus ist es, um induktiv schließen zu können, nicht notwendig, den faktischen Uniformitätsgrad zu kennen oder zufällig zu erraten. Selbstverständlich wird man schlechte Ergebnisse erhalten, wenn die Uniformität nahe beim Maximum liegt, wenn man jedoch annimmt, daß sie nicht weit vom Minimum entfernt ist (und umgekehrt); doch wird niemand daran gehindert, die Annahme zu verwerfen, wenn sie zu unplausiblen Resultaten führt, und sie durch eine andere zu ersetzen. Je näher die *geschätzte* Uniformität w′ bei der *tatsächlichen* liegt, um so bessere Ergebnisse wird man mit der Funktion $c^{L'}$, die man auf Grund der geschätzten Uniformität wählt, erhalten. Wenn die Schätzung w′ nahe beim tatsächlichen Wert w liegt, so wird man mit $c^{L'}$ *praktisch* die gleichen Werte erhalten wie mit c^L, so daß man $c^{L'}$ als *praktisch optimal und fair* ansehen kann, obwohl die Annahme, daß w′ die Uniformität ist, *an sich falsch* ist.

Ohne gewisse Annahmen über die Welt (z. B. über den Uniformitätsgrad des Bereichs der Dinge oder eines Teilbereichs davon[18]) können wir also nicht aus der Erfahrung lernen und insbesondere nicht von Bekanntem auf Unbekanntes induktiv schließen. Jede Annahme über den Uniformitätsgrad des Gegenstandsbereichs legt eine bestimmte induktive Methode als fair fest, die (abgesehen von c^∞) im

18 Im letzten Kapitel werden noch andere Annahmen aufgezeigt, die wir machen müssen, wenn wir Kenntnisse von den Dingen der Welt erlangen wollen.

Sinne Reichenbachs *selbstkorrigierend*[19] ist, d. h. bei der die durch die Prämissen beschriebenen Erfahrungen jene Annahme über den Uniformitätsgrad teilweise wieder korrigieren, indem sie den c-Wert in Richtung auf die beobachteten relativen Häufigkeiten hin abändern, und die jene Annahme um so besser korrigieren, je umfangreicher die Erfahrungsdaten sind. Solche Annahmen müssen zwar *a priori gemacht* werden, sie brauchen jedoch nicht *a priori begründet* zu werden; sie *können* auch nicht a priori begründet werden, und werden von uns manchmal durch andere ersetzt, wenn sie im Lichte neuerer Erfahrungen unplausibel erscheinen.

Zur Beantwortung der Frage, ob es sich bei ihnen um synthetisch-apriorische Urteile handelt, ist es nützlich, sich zu vergegenwärtigen, was mit den Begriffen „analytisch", „synthetisch", „a priori" und „a posteriori" (oder „empirisch") überhaupt intendiert ist. Der Einfachheit halber wird nur der Fall betrachtet, daß das vorliegende Urteil, dem eine dieser Eigenschaften zukommen soll, wahr ist.

Daß eine wahre Aussage *analytisch wahr*[20] ist, kann heißen, (α) daß sie logisch wahr ist[21], (β) daß sie aus Definitionen logisch folgt (= daß sie logisch wahr ist oder aus Definitionen logisch folgt), oder (γ) daß sie aus Definitionen und *Analytizitäts*postulaten logisch folgt. Die Bestimmung (α) ist offenbar zu eng, da der Begriff „analytisch wahr" dann mit „logisch wahr" zusammenfällt und uninteressant wird. Die Bestimmung (γ) ist umgekehrt sicherlich zu weit, da Postulate, die keine Definitionen sind, Aussagen über die Welt und nicht nur über begriffliche Zusammenhänge machen; sie ist ferner zumindest teilweise zirkulär, da man wissen muß, wodurch sich Analytizitätspostulate von anderen Aussagen unterscheiden, um sagen zu können, ob ein vorgelegter wahrer Satz analytisch wahr ist oder nicht, und sie ist schließlich für bestimmte erkenntnistheoretische Problemstellungen

---

19 Vgl. Reichenbach [3], S. 446.

20 Statt „analytisch wahr" wird oft einfach „analytisch" und statt „analytisch falsch" dann „kontradiktorisch" gesagt.

21 Auf die subtilen Untersuchungen von Jaakko Hintikka zu diesem Problem auf S. 193—253 des Buches „Deskription, Analytizität und Existenz" (hrsg. von Paul Weingartner, Salzburg—München 1966), in denen er zu zeigen versucht, daß nicht einmal alle logisch wahren Sätze im Sinne Kants analytisch wahr sind, möchte ich hier nicht eingehen; vgl. meine Besprechung dieses Buches in Synthese 18 (1968), S. 109—117, insbesondere S. 111—114.

ungeeignet, da sie das Problem der synthetisch-apriorischen Urteile sozusagen *wegdefiniert* (durch (γ) werden Urteile dieser Art ja automatisch zu *analytischen* Sätzen). Ich möchte daher in diesem Zusammenhang an der Bestimmung (β) festhalten. Eine wahre Aussage ist dann *synthetisch wahr* genau dann, wenn sie nicht (im Sinne von (β)) analytisch wahr ist.

Daß eine wahre Aussage *a priori wahr* ist, kann heißen, (a) daß sie logisch wahr ist, (b) daß sie als wahr erweisbar ist, ohne daß hierzu Prämissen benötigt werden, die Erfahrungen beschreiben, (c) daß sie als wahr angenommen wird, ohne daß Erfahrungsaussagen logisch notwendige Bedingungen für sie sind, (d) daß ihre Wahrheit eine notwendige Voraussetzung dafür ist, daß überhaupt Erfahrungen gemacht werden können (daß sie Erfahrungen überhaupt erst ermöglicht), oder (e) daß sie den üblichen Sprachgebrauch und den Gebrauch der in ihr vorkommenden Begriffe korrekt wiedergibt.

Die Bestimmung (a) ist offenbar zu eng und, wie auch (α), daher ungeeignet. Unbrauchbar ist auch (b), solange nicht gesagt wird, was mit „erweisbar" gemeint ist; wenn dieser Begriff mit „beweisbar" im Sinne von „logisch wahr" zusammenfällt, ist (a) mit (b) identisch, und wenn nicht, so muß, bevor die Methode des *Erweisens* verwendet wird, erst geprüft werden, ob sie logisch korrekt ist bzw. ob ihre Korrektheit die Gültigkeit bestimmter Annahmen über die Welt voraussetzt. Ferner muß gelten, daß ein Satz auch dann noch a priori wahr ist, wenn man ihn formalisiert, d. h. wenn man von seiner Interpretation abstrahiert;[22] dann ist jedoch kein synthetisch wahres Urteil a priori wahr, denn ein uninterpretierter nicht logisch wahrer Satz kann verschieden interpretiert werden, und unter anderem auch so, daß er hierbei zu einer falschen Aussage wird. Unbrauchbar ist auch die Festlegung (d); denn man kann im Alltag wie in den Wissenschaften auch die Intensionen der grundlegendsten Begriffe ändern, indem man die Regeln für ihren Gebrauch modifiziert und damit andere Urteile als a priori wahr annimmt. Gelegentlich wird daher argumentiert, daß die Aprioriguültigkeit einer derartigen Aussage aus den

[22] Die Interpretation ordnet ja den in dem Satz vorkommenden außerlogischen Ausdrücken Extensionen wie Dinge, Klassen von Dingen und Klassen von geordneten r-Tupeln von Dingen zu, die für die Frage seiner Apriorität ohne Bedeutung sind.

Regeln der Verwendung der in ihr vorkommenden Begriffe ersehen werden könne, was auf Bestimmung (e) hinausläuft; auf diese Festlegung möchte ich jedoch nicht eingehen, da durch sie die augenblickliche Art der Verwendung unserer Begriffe fixiert und als für alle Zeiten a priori gültig deklariert wird, wohingegen der wissenschaftliche Fortschritt oft darin bestanden hat, daß die Intensionen gewisser grundlegender Begriffe verändert worden sind.
Kant hat den Begriff „a priori wahr" üblicherweise im Sinne von (b), gelegentlich aber auch in Übereinstimmung mit (d) benützt. Bei diesen Festsetzungen (wie immer man sie präzisieren mag) sind die Annahmen über den Uniformitätsgrad des Gegenstandsbereichs zwar synthetische, aber zweifellos nicht apriorische Behauptungen; jede von ihnen kann stets verworfen und durch eine andere ersetzt werden. Als a priori wahre Urteile können sie gelten, wenn man diesen Begriff im Sinne von (c) benützt, was ich als zweckmäßig empfinde.[23] Hier wird „a priori" in einem unschuldigen pragmatischen Sinn verstanden: Es wird nicht vorausgesetzt, daß man einen *a priori wahren* Satz *a priori begründen* kann; vielmehr wird er in diesem *Zusammenhang und in dieser Situation a priori akzeptiert*, ohne daß man dafür logische oder empirische Gründe angibt (normalerweise wird man jedoch Zweckmäßigkeitsgründe aufführen können). Dies darf nicht so verstanden werden, daß Erfahrungen (z. B. Erfahrungen, die man mit einer solchen Annahme bei ihrer Anwendung auf die Wirklichkeit gemacht hat), es nicht *nahelegen* können, diese Apriorivoraussetzungen durch andere zu ersetzen; wesentlich ist dabei lediglich, daß Erfahrungen uns zu diesem Schritt niemals *zwingen* können, d. h. daß Erfahrungsaussagen niemals logisch notwendige Bedingungen für jene Voraussetzungen sind.

---

[23] Ein wahres Urteil ist dann *a posteriori wahr* (oder *empirisch wahr*) genau dann, wenn es nicht (im Sinne von (c)) a priori wahr ist.

V. Das System der Adäquatheitsbedingungen

1. Widerspruchsfreiheit und Vollständigkeit der Adäquatheitsbedingungen

Die im letzten Kapitel formulierten Adäquatheitsbedingungen für den logischen Wahrscheinlichkeitsbegriff können als ein System von Postulaten aufgefaßt werden. Es stellen sich uns dann die Fragen, ob dieses System vollständig ist, ob es syntaktisch und semantisch widerspruchsfrei ist und ob die einzelnen Postulate voneinander logisch unabhängig sind. Zur Beantwortung der Vollständigkeitsfrage sei nochmals und in einer weniger intuitiven Form das Verhältnis der adäquaten induktiven Methoden und des Uniformitätsgrades des Gegenstandsbereichs zueinander untersucht.

Das Problem, wie man die Uniformität des Gegenstandsbereichs am besten mißt, wird an einer Folge von Beispielen erörtert. Dabei ist zu berücksichtigen, daß es durchaus möglich ist, verschiedene Funktionen zu bestimmen, die die Uniformitätsgrade verschieden festlegen. Man kann ja auch etwa die Relevanz von zusätzlichen Erfahrungsdaten sowohl durch die Differenz zweier Wahrscheinlichkeiten messen, wie im 3. Abschnitt des II. Kapitels, als auch (und mit gleichem Recht) durch den Quotienten dieser beiden Wahrscheinlichkeiten, wobei man andere Werte erhält als im ersten Fall. Man wird jedoch, wenn verschiedene Methoden zur Verfügung stehen, die einfachste wählen, und diese ist hier ohne Zweifel die Differenzbildung. Ähnliche Erwägungen werden auch im folgenden angestellt.

Gegeben seien drei Maschinen, von denen die erste rote Kugeln und die zweite nichtrote Kugeln herstellt, während die dritte in jede der ihr zur Verfügung stehenden Urnen Mischungen von 100 roten und nichtroten Kugeln füllt. Eine dieser Urnen voll Kugeln sei für das Folgende der Bereich, über dem wir eine Sprache mit den Gegenstandsausdrücken $\alpha_1, \ldots, \alpha_{100}$ und dem Eigenschaftsausdruck Π_1 interpretieren. Den Gegenstandsausdrücken ordnen wir dabei umkehrbar eindeutig die Kugeln aus der Urne zu (z. B. in der Reihenfolge, in der wir die Kugeln bei der Überprüfung ihrer Farbe der Urne entneh-

men), und Π_1 bezeichne bei dieser Interpretation die Eigenschaft *rot* (bzw. die Klasse der roten Dinge). Eine individuelle Verteilung Φ beschreibt dann, welche Kugeln aus der Urne rot sind und welche nicht, und eine statistische Verteilung Γ sagt aus, wieviele Kugeln dieser Urne rot sind und wieviele nicht.

Durch den Uniformitätsgrad soll nun angegeben werden, wie häufig wir im Durchschnitt eine durch ein V-Prädikat bezeichnete Eigenschaft (kurz: eine V-Eigenschaft), in diesem Fall *rot* oder *nichtrot*, wiederfinden werden, wenn wir sie schon einmal angetroffen haben. Wenn alle Kugeln der Urne rot sind oder wenn alle diese Kugeln nichtrot sind, dann ist der Uniformitätsgrad maximal; denn wenn wir ein Ding mit einer der beiden Eigenschaften beobachten, so wird auch der nächste Gegenstand diese Eigenschaft haben, und dann werden wir überhaupt kein Ding antreffen, das die andere Eigenschaft besitzt. Es ist naheliegend, als maximalen Uniformitätsgrad 1 zu wählen. Zur Frage, welche Zahl den minimalen Uniformitätsgrad charakterisieren soll, sind etwas detailliertere Überlegungen notwendig.

Zunächst scheint es auf der Hand zu liegen, daß nur 0 als minimaler Grad infrage kommt. Dies würde aber bedeuten, daß wir im Durchschnitt bei keiner der beiden Eigenschaften (in absehbarer Zeit oder überhaupt) damit rechnen können, noch auf ein Ding zu stoßen, das die Eigenschaft des gerade beobachteten Gegenstandes hat. Das ist jedoch nicht der Fall; wenn unter den 100 Kugeln 50 rote und 50 nichtrote sind, so werden wir im Durchschnitt bei jedem zweiten Zug eine Kugel mit der gleichen Eigenschaft erhalten. Bei einem Grundprädikat und zwei V-Prädikaten ist demnach die Zahl $1/2$ als untere Grenze der Uniformitätsgrade anzusehen.

Gegeben seien fünf Maschinen, von denen die beiden ersten einfarbige Kugeln, nämlich Kugeln mit den Farben *rot* bzw. *orange* herstellen, die dritte zweifarbige Kugeln mit den Farben *rot und orange* produziert und die vierte Kugeln herstellt, die *weder rot noch orange* sind. Die fünfte Maschine füllt Urnen mit Mischungen aus je 100 Kugeln. Wenn der Gegenstandsbereich (eine bestimmte Urne mit Kugeln) bezüglich der vier Eigenschaften *rot und orange*, *rot und nicht orange*, *orange und nicht rot* und *weder rot noch orange* völlig disuniform ist, so bedeutet dies, daß je 25 Kugeln eine dieser vier Eigenschaften besitzen. Hier charakterisiert die Zahl $1/4$ den minimalen Uniformitätsgrad.

Eine Sprache, die diesen Gegenstandsbereich adäquat beschreibt, besitzt neben den Gegenstandsausdrücken $\alpha_1, \ldots, \alpha_{100}$ die einstelligen Grundprädikate Π_1 und Π_2 und somit die vier V-Prädikate $\Pi^*_1, \ldots, \Pi^*_4$. Π_1 bezeichnet bei der vorgegebenen Interpretation die Eigenschaft *rot* und Π_2 *orange*, während die V-Prädikate $\Pi^*_1, \ldots, \Pi^*_4$ die oben erwähnten V-Eigenschaften *rot und orange*, ..., *weder rot noch orange* bezeichnen.

Nimmt man noch die Farbe *gelb* hinzu, so erhält man $2^3 = 8$ V-Eigenschaften; um dann begründen zu können, daß hier die Zahl $1/8$ der kleinste Uniformitätsgrad ist, benötigt man eine Anzahl von Kugeln, die nicht kleiner als 8 und durch 8 teilbar ist. Mit wachsendem p übersteigt die Anzahl q jede endliche Grenze, weshalb der Bruch $1/q$ gegen 0 konvergiert. Werden bei einem endlichen Gegenstandsbereich unendlich viele Grundeigenschaften berücksichtigt (so daß eine Sprache mit unendlich vielen Gegenstandsausdrücken und unendlich vielen Grundprädikaten benützt wird), dann erhält man durch Anwendung der beiden Limeskonventionen das Ergebnis, daß 0 die untere Grenze aller Uniformitätsgrade ist.

Es bleibt noch das Problem, eine geeignete Funktion zu finden, die als Gradmesser der Uniformität des Gegenstandsbereichs hinsichtlich seiner V-Eigenschaften angesehen werden kann und die als kleinsten bzw. größten Wert die Zahlen $1/q$ bzw. 1 hat, wobei $q = 2^p$ und p die Anzahl der Grundprädikate (Grundeigenschaften) ist. Zwei naheliegende Verfahren erweisen sich als inadäquat. Das eine besteht darin, als Uniformitätsgrad die Summe der relativen Häufigkeiten der V-Eigenschaften im Gegenstandsbereich zu wählen; diese Summe ist jedoch stets mit 1 identisch, so daß sie für diese Zwecke nicht infrage kommt. Die andere besteht darin, als Uniformitätsgrad das arithmetische Mittel aus den einzelnen relativen Häufigkeiten zu nehmen; doch dieses ist wiederum stets mit $1/q$ identisch und muß deshalb ebenfalls außer Betracht bleiben.

Offensichtlich sind lineare Funktionen der relativen Häufigkeiten nicht zur Messung der Uniformitätsgrade geeignet. Es wird deshalb die nächsteinfachere Funktion genommen und die Uniformität eines endlichen Gegenstandsbereichs bezüglich endlich vieler V-Eigenschaften durch die Summe der *Quadrate* jener relativen Häufigkeiten gemessen; diese Funktion nimmt tatsächlich Werte von $1/q$ bis 1 an.

Je mehr Grundbegriffe die Sprache enthält, desto mehr kann die untere Grenze 1/q der 0 angenähert werden (genaugenommen geht man mit der schrittweisen Vermehrung der Grundbegriffe zu immer ausdrucksreicheren Sprachen über). Aus zwei Gründen empfiehlt es sich, hier Sprachen mit unendlich vielen Grundprädikaten und somit auch mit unendlich vielen V-Prädikaten in Erwägung zu ziehen: 1. Man ist dann nicht genötigt, von Fall zu Fall die vorgegebene Sprache durch eine ausdrucksreichere zu ersetzen, und 2. es ist dann auch die 0 ein möglicher Uniformitätsgrad.

Wenn die Anzahl der Grundprädikate endlich ist und man dieser Sprache schrittweise neue Gegenstandsausdrücke hinzufügt, so erhält man eine Folge von Sprachen, durch die es möglich wird, alle und nur die rationalen Zahlen zwischen 1/q und 1 als mögliche Uniformitätsgrade anzusehen (besitzt die Sprache unendlich viele einstellige Grundprädikate, so gilt gleiches für die rationalen Zahlen zwischen 0 und 1). Nun können irrationale Zahlen als Folgen von rationalen Zahlen aufgefaßt und dargestellt werden. Deshalb erweist es sich als zweckmäßig, auch eine abzählbar unendliche Gesamtheit von Gegenstandsausdrücken (und entsprechend von Elementen des Gegenstandsbereichs) vorauszusetzen.

Der Uniformitätsgrad eines Gegenstandsbereichs hängt davon ab, welche V-Eigenschaften seine Elemente besitzen (wie die Welt beschaffen ist). Nun charakterisieren Zustandsbeschreibungen die Gegenstände in dieser Hinsicht vollständig; daher ist der Begriff der Uniformität in Abhängigkeit von einer vorgegebenen Zustandsbeschreibung einzuführen.

Zu beantworten bleibt noch die Frage, was unter einer Zustandsbeschreibung für eine Sprache mit unendlich vielen Gegenstandsausdrücken und/oder unendlich vielen einstelligen Grundprädikaten zu verstehen ist. Man kann darunter eine Folge von Zustandsbeschreibungen verstehen, die entsteht, wenn man eine Folge von Sprachen betrachtet, die aus einer vorgegebenen durch Hinzufügung von immer neuen Gegenstandsausdrücken und/oder neuen einstelligen Grundprädikaten entsteht. Es kann darunter auch ein Satz von unendlicher Länge verstanden werden; wenn man den Begriff des Aktualunendlichen für problematisch hält, soll damit nichts anderes als jene Folge von Zustandsbeschreibungen gemeint sein, und im anderen

Fall das Ergebnis dieser unendlich vielen schrittweisen Hinzufügungen bzw. Änderungen.[1] Der Begriff der Uniformität eines Gegenstandsbereichs auf Grund einer Zustandsbeschreibung bezüglich der V-Eigenschaften kann dann folgendermaßen definiert werden:

D 27 (a) Die Anzahl n der Gegenstandsausdrücke und die Anzahl p der einstelligen Grundprädikate der Sprachen seien endlich, und n_i sei die Anzahl der Gegenstände, die bei einer vorgegebenen Zustandsbeschreibung Σ (sowie bei der entsprechenden statistischen Verteilung) die durch das V-Prädikat Π^*_i bezeichnete Eigenschaft besitzen.
Dann ist w die Uniformität, die der Gegenstandsbereich auf Grund von Σ bezüglich aller 2^p V-Eigenschaften hat, genau dann, wenn

$$w = \sum_{i=1}^{2^p} \left(\frac{n_i}{n}\right)^2 .$$

(b) Die Anzahl n der Gegenstandsausdrücke sei endlich und die Anzahl der einstelligen Grundprädikate abzählbar unendlich. Die Zustandsbeschreibung Σ sei ein Satz von unendlicher Länge, der die Konjunktion von n unendlich langen Aussagen ist; jede von ihnen habe die Form $(\neg)\,(\alpha_i \varepsilon \Pi_1) \wedge (\neg)\,(\alpha_i \varepsilon \Pi_2) \wedge \ldots \wedge (\neg)\,(\alpha_i \varepsilon \Pi_j) \wedge \ldots$, d. h. jede von ihnen sei eine unendliche Konjunktion, die für jede natürliche Zahl $j \neq 0$ entweder den Satz $\alpha_i \varepsilon \Pi_j$ oder $\neg\,(\alpha_i \varepsilon \Pi_j)$ enthält. Ein endlicher Teil Σ^p von Σ sei eine Konjunktion von n Sätzen der Form $(\neg)\,(\alpha_i \varepsilon \Pi_1) \wedge (\neg)\,(\alpha_i \varepsilon \Pi_2) \wedge \ldots \wedge (\neg)\,(\alpha_i \varepsilon \Pi_p)$ für jedes i mit $1 \leqslant i \leqslant n$, wobei die Teilsätze von Σ^p genau dann negiert sind, wenn die entsprechenden Teilsätze von Σ es sind; die Uniformität des Gegenstandsbereichs auf Grund eines derartigen endlichen Teils Σ^p von Σ bezüglich der 2^p V-Eigenschaften ergibt sich in diesem Fall nach (a). Dann ist w die Uniformität, die der Gegenstandsbereich auf Grund von Σ bezüglich aller V-Eigenschaften hat, genau dann, wenn

[1] Zur Behandlung von Sätzen mit unendlicher Länge vgl. z. B. Carol R. Karp, Languages with Expressions of Infinite Length, Amsterdam 1964.

$$w = \lim_{p \to \infty} \sum_{i=1}^{2^p} \left(\frac{n_i}{n}\right)^2 .$$

(c) Die Anzahl der Gegenstandsausdrücke sei abzählbar unendlich, und die Anzahl p der einstelligen Grundprädikate sei endlich. Die Zustandsbeschreibung Σ sei eine unendliche Konjunktion von Sätzen, die für jede natürliche Zahl $i \neq 0$ die Form $(\neg)\,(\alpha_i \varepsilon \Pi_1) \wedge (\neg)\,(\alpha_i \varepsilon \Pi_2) \wedge \ldots \wedge (\neg)\,(\alpha_i \varepsilon \Pi_p)$ haben. Ein endlicher Abschnitt Σ_n von Σ sei die Konjunktion $(\neg)\,(\alpha_1 \varepsilon \Pi_1) \wedge \ldots \wedge (\neg)\,(\alpha_n \varepsilon \Pi_p)$, wobei die Teilsätze von Σ_n genau dann negiert sind, wenn es die entsprechenden Teilsätze von Σ sind. Die Uniformität des Gegenstandsbereichs auf Grund eines derartigen endlichen Abschnittes Σ_n von Σ bezüglich der V-Eigenschaften ergibt sich dann wiederum nach (a).
Dann ist w die Uniformität, die der Gegenstandsbereich auf Grund von Σ bezüglich aller 2^p V-Eigenschaften hat, genau dann, wenn

$$w = \lim_{n \to \infty} \sum_{i=1}^{2^p} \left(\frac{n_i}{n}\right)^2 .$$

(d) Die Anzahl der Gegenstandsausdrücke und die Anzahl der einstelligen Grundprädikate seien beide abzählbar unendlich. Die Zustandsbeschreibung Σ sei eine unendliche Konjunktion von unendlichen Sätzen der Form $(\neg)\,(\alpha_i \varepsilon \Pi_1) \wedge (\neg)\,(\alpha_i \varepsilon \Pi_2) \wedge \ldots \wedge (\neg)\,(\alpha_i \varepsilon \Pi_j) \wedge \ldots$ für alle natürlichen Zahlen i und j mit $i \neq 0$ und $j \neq 0$, Σ_n sei ein endlicher Abschnitt von Σ (enthält also nur die ersten n unendlichen Konjunktionen) und $\Sigma^p{}_n$ sei ein endlicher Teil von Σ_n; die Uniformität des auf die ersten n Glieder beschränkten Gegenstandsbereichs auf Grund von $\Sigma^p{}_n$ bezüglich der 2^p V-Eigenschaften ergibt sich auch hier nach (a).
Dann ist w die Uniformität, die der Gegenstandsbereich auf Grund von Σ bezüglich aller V-Eigenschaften hat, genau dann, wenn

$$w = \lim_{n \to \infty} \lim_{p \to \infty} \sum_{i=1}^{2^p} \left(\frac{n_i}{n}\right)^2 .$$

Der hier definierte Begriff der Uniformität unterscheidet sich von den Vorstellungen des Philodemus und Mills, die auch Carnap explizieren wollte,[2] nicht nur dadurch, daß diese lediglich einen klassifikatorischen Begriff der Uniformität im Auge hatten, der nur zwischen Uniformität und Disuniformität zu unterscheiden gestattet, während hier ein metrischer Begriff definiert wird, mit dessen Hilfe Grade an Uniformität angegeben werden können, sondern auch und vor allem in folgender Hinsicht: Nach den Vorstellungen jener Autoren ist in der Welt um so mehr Uniformität vorhanden, je mehr Naturgesetze gelten, d. h. je mehr Allbehauptungen von gewisser Form wahr sind. Nun hat Carnap gezeigt, daß um so mehr Naturgesetze gelten, je mehr V-Eigenschaften leer sind.[3] Der in D 27 definierte Begriff der Uniformität gibt dieses Verhältnis nicht wieder. Man kann sich leicht an Beispielen klarmachen, daß der Uniformitätsgrad nach dieser Definition größer ist, wenn zwar keine V-Eigenschaft leer ist, aber nur entweder wenige Beispiele oder wenige Gegenbeispiele zu den einzelnen V-Eigenschaften existieren, als wenn zwar eine oder einige wenige V-Eigenschaften leer sind, die Gegenstände sich jedoch auf die übrigen V-Eigenschaften etwa in gleichen Anzahlen verteilen.

So interessant eine Explikation des traditionellen Uniformitätsbegriffs sein mag, so wertlos will sie mir in diesem Zusammenhang erscheinen. Man kann adäquate induktive Methoden anwenden und mit ihnen vernüftige Ergebnisse erhalten, auch wenn in unserer Welt kein einziges Naturgesetz gilt; zur Begründung des induktiven Schließens braucht also keine Uniformität im traditionellen Sinn vorausgesetzt zu werden. Wohl aber muß eine andere Uniformität dazu angenommen werden, die durch Theorem 17 in eine Beziehung zu fairen induktiven Methoden gesetzt wird und die ich durch D 27 sowohl adäquat als auch einfach expliziert zu haben glaube.[4]

Es bereitet nun keine Schwierigkeiten, den Begriff der *fairen* induktiven Methode zu definieren. Dazu ist zu berücksichtigen, daß man den

---

[2] Vgl. Carnap [11], S. 66.

[3] Vgl. Carnap [9], § 37.

[4] Allerdings könnte Theorem 17 auch bei gewissen Modifikationen von D 27 gelten. Eine zusätzliche Stütze erhält meine obige Vermutung durch die Tatsache, daß sie auch durch die Ergebnisse Carnaps in [11], S. 65—79, nahegelegt werden; vgl. auch S. 353—368 dieses Buches.

singulären Voraussageschluß ja auch als Schätzung der relativen Häufigkeit der in der Hypothese erwähnten Eigenschaft im zugrundegelegten Gegenstandsbereich im Hinblick auf das Erfahrungsdatum ansehen kann.[5]

Gegeben sei die L-Methode c; der Gegenstandsbereich (und damit die Gesamtheit der Gegenstandsausdrücke) sei hinreichend groß, so daß es sinnvoll ist, die willkürlich vorgegebene reelle Zahl w mit $1/q \leqslant w \leqslant 1$ wenigstens approximativ als Schätzung der relativen Häufigkeit jener Eigenschaft im Gegenstandsbereich anzusehen (oder als Schätzung des Grenzwertes der relativen Häufigkeiten, falls dieser Bereich unendlich ist). Wenn noch kein Erfahrungsdatum zur Verfügung steht, muß diese Schätzung im Hinblick auf die Tautologie vorgenommen werden und führt bei allen adäquaten induktiven Methoden dann zum gleichen Ergebnis; dieser Fall kann also außer Betracht gelassen werden. Wenn die Schätzung bezüglich eines Erfahrungsdatums vorgenommen worden ist, das mehr als einen Gegenstand erwähnt, so ist sie (bei endlich vielen Grundprädikaten) nach den Theoremen 14 und 15 eine Funktion von L und somit von l sowie von c ($\alpha_2 \varepsilon \Pi^*_i$, $\alpha_1 \varepsilon \Pi^*_i$). Ob c fair ist, hängt also davon ab, ob c zu fairen Schätzungen führt, wenn man ein einziges Ding auf seine Eigenschaften hin überprüft hat.

Zu einer fairen Schätzung führt c in diesem Fall offenbar dann, aber auch nur dann, wenn der c-Wert mit der Uniformität des Gegenstandsbereichs (so wie er von der vorgegebenen und in diesem Zusammenhang als wahr angenommenen Zustandsbeschreibung Σ beschrieben wird) identisch ist. Denn durch den Uniformitätsgrad soll ja charakterisiert werden, mit welcher durchschnittlichen Häufigkeit man rechnen kann, wiederum ein Ding mit der vorgegebenen Eigenschaft zu beobachten, wenn der gegenwärtig untersuchte Gegenstand diese Eigenschaft besitzt.[6] Es kann daher definiert werden:

D 28 c sei eine adäquate induktive Methode und im Sinne der Theoreme 14 und 15 durch die positive reelle Zahl L charakteri-

---

[5] Vgl. Theorem 23 auf S. 241 f. dieser Arbeit.

[6] Eine weniger intuitive Begründung für diese Wahl ergibt sich aus der Argumentation Carnaps in [11], S. 69 ff., insbesondere aus der Gleichung (22-5); vgl. auch Gleichung (40) auf S. 361 dieses Buches.

siert; die Gesamtheit der Gegenstandsausdrücke sei endlich oder unendlich, desgleichen die Gesamtheit der einstelligen Grundprädikate; Σ sei eine Zustandsbeschreibung.
Dann ist c fair relativ zum Uniformitätsgrad w, den der Gegenstandsbereich auf Grund der Zustandsbeschreibung Σ bezüglich aller V-Eigenschaften hat, genau dann, wenn $c(\alpha_2 \varepsilon \Pi^*_i, \alpha_1 \varepsilon \Pi^*_i) = w$ für jedes beliebige V-Prädikat Π^*_i, wobei w die Uniformität ist, die der Gegenstandsbereich auf Grund von Σ bezüglich aller V-Eigenschaften hat.

Mit dieser Definition kann folgender Lehrsatz bewiesen werden:

Theorem 17

c sei eine adäquate induktive Methode und im Sinne der Theoreme 14 und 15 durch die positive reelle Zahl L charakterisiert; die Gesamtheit der Gegenstandsausdrücke sei endlich oder abzählbar unendlich.
(a) Die Gesamtheit der einstelligen Grundprädikate sei endlich.
Dann ist die Bedingung, daß c fair relativ zum Uniformitätsgrad w ist, den der Gegenstandsbereich auf Grund der Zustandsbeschreibung Σ bezüglich aller V-Eigenschaften hat, logisch gleichwertig mit jeder der drei folgenden Gleichungen:

$$(\alpha) \quad w = \frac{1 + L/q}{1 + L}$$

$$(\beta) \quad L = \frac{1 - w}{w - 1/q}$$

$$(\gamma) \quad l = \frac{1 - w}{q - 1}$$

(b) Die Gesamtheit der einstelligen Grundprädikate sei abzählbar unendlich.
Dann ist die Bedingung, daß c fair relativ zum Uniformitätsgrad w ist, den der Gegenstandsbereich auf Grund der Zustandsbeschreibung Σ bezüglich aller V-Eigenschaften hat, logisch gleichwertig mit jeder der zwei folgenden Gleichungen:

$$(\alpha)\quad w = \frac{1}{1+L}$$

$$(\beta)\quad L = \frac{1-w}{w}$$

Die Beweise für die Teilbehauptungen (a) erhält man durch elementare algebraische Umformungen unter Zugrundelegung der Festsetzungen, daß $c(\alpha_2 \varepsilon \Pi^*_i, \alpha_1 \varepsilon \Pi^*_1) = f(1,1) = 1 - (q-1) \cdot f(1,0) = 1 - (q-1) \cdot l$ und daß $L = ql/(1-ql)$. Bei den Teilbehauptungen (b) ist außerdem noch zu berücksichtigen, daß bei festem L und beliebig großem p die Werte $1/q$ und L/q beliebig klein werden; daher gilt hier: $l = 0$, woraus nun aber kein Schluß auf die anderen Gleichungen gezogen werden kann.

Zu jedem Uniformitätsgrad w mit $1/q \leqslant w \leqslant 1$ bzw. $0 \leqslant w \leqslant 1$ gibt es also genau eine Funktion c, nämlich c^L mit $L = (1-w)/(w-1/q)$, die fair ist und die in der Anwendung (z. B. bei Wetten) optimal ist, d. h. den größten Gewinn verspricht, und umgekehrt. Zwischen Uniformitätsgraden und Zustandsbeschreibungen besteht allerdings keine Eineindeutigkeit; vielmehr können Gegenstandsbereiche auf Grund von *verschiedenen* Zustandsbeschreibungen hinsichtlich gewisser V-Eigenschaften den *gleichen* Uniformitätsgrad w haben.

Die Funktionen c^0 und c^∞ sind zwar selbst keine adäquaten induktiven Methoden (im definierten Sinn), sie können aber durch adäquate induktive Methoden beliebig approximiert werden. Wie Theorem 17 zeigt, ist c^0 (im intuitiven Sinn oder, bei entsprechender Erweiterung von D 27, auch im formalen Sinn) fair relativ zum Uniformitätsgrad w (den der Gegenstandsbereich auf Grund der Zustandsbeschreibung Σ bezüglich aller V-Eigenschaften hat) genau dann, wenn $w = 1$, und ist c^∞ fair relativ zu Uniformitätsgrad w genau dann, wenn $w = 1/q$ bzw., bei abzählbar unendlich vielen einstelligen Grundprädikaten, wenn $w = 0$.

Auf die Frage, ob die Adäquatheitsbedingungen vollständig sind oder durch zusätzliche Forderungen, die den logischen Wahrscheinlichkeitsbegriff betreffen, ergänzt werden müssen, gibt Theorem 17 eine klare Antwort: Da zusätzliche Forderungen, die nicht aus den bisherigen Bedingungen logisch folgen, Funktionen c^L ausschließen,

denen ein möglicher Uniformitätsgrad w entspricht, können sie keine Forderungen sein, die den Wahrscheinlichkeitsbegriff betreffen, sondern sind Behauptungen über die Welt bzw. über den Gegenstandsbereich, über den die Sprache interpretiert ist, und über dessen Uniformität. *Die Gesamtheit der Adäquatheitsbedingungen ist also vollständig.*

Sie ist auch syntaktisch widerspruchsfrei, d. h. aus ihr kann nicht logisch eine Kontradiktion deduziert werden. Denn nach den Theoremen 14 und 15 erfüllen so viele reguläre Bestätigungsmethoden die vier Adäquatheitsbedingungen, wie es positive reelle Zahlen gibt (daran ändern auch die beiden Limes-Konventionen nichts). Daß der Widerspruchsfreiheitsbeweis hier nicht mit finitären Mitteln erbracht worden ist, die Hilbert zunächst als die einzig zulässigen proklamiert hat, ist dabei ohne Belang; denn einerseits kann nicht mit finitären Mitteln widerlegbar sein, was mit klassischen beweisbar ist, und andererseits hat man heute seit den Ergebnissen Gödels keine Bedenken mehr, auch klassische Mittel zu verwenden (man denke etwa an Takahashis Widerspruchsfreiheitsbeweis für die einfache Typentheorie).

Eine andere Frage ist, ob die Adäquatheitsbedingungen auch tatsächlich *nur* Forderungen an den Wahrscheinlichkeitsbegriff sind oder ob sie nicht teilweise synthetische Aussagen über den Gegenstandsbereich und dessen Uniformitätsgrad hinsichtlich seiner V-Eigenschaften darstellen, *ob sie also auch korrekt oder semantisch widerspruchsfrei sind.* Auf den ersten Blick will es scheinen, daß sie nicht analytisch sein können, da ja den Uniformitätsgraden 1 und 1/q (bzw. 0 bei abzählbar unendlich vielen Grundprädikaten) nach Theorem 17 keine adäquaten induktiven Methoden entsprechen. Dieser Eindruck wird noch durch die ganz andersartigen Überlegungen verstärkt, die im nächsten Abschnitt geschildert werden.

2. Der Gehalt der Adäquatheitsbedingungen

Gegen einzelne der Adäquatheitsbedingungen sind Einwände denkbar, die darauf hinauslaufen, daß diese Bedingungen nur in *einigen* faktischen Situationen (nur bei einigen Interpretationen der Sprache)

erfüllt sind, daß es jedoch Situationen (Interpretationen der Sprache) gibt, bei denen sie offensichtlich falsch sind. Wenn man diesen Einwänden recht gibt, so leugnet man damit, daß diese Bedingungen tat ächlich Adäquatheitsbedingungen für den Begriff der logischen Wahrscheinlichkeit sind; man behauptet dann, daß man sie als synthetische Aussagen über die Welt aufzufassen hat.

Gegen die Adäquatheit der Bedingung (II) (b) kann man folgendermaßen argumentieren: Der Gegenstandsbereich, über dem die Sprache interpretiert ist, sei die Gesamtheit der Kugeln einer Urne; die Eigenschaftsausdrücke $\Pi_1, \Pi_2, \ldots, \Pi_6$ bezeichnen dabei die Eigenschaften *rot*, *orange*, ..., *violett*. Es ist zwar denkbar, daß die Kugeln verschiedenfarbig sind, im vorliegenden Fall mögen sie jedoch genau eine dieser Spektralfarben besitzen. Dann sind alle bis auf 6 der V-Eigenschaften leer, und die Zustandsbeschreibungen mit den entsprechenden V-Prädikaten sind bei einer solchen Interpretation von vornherein falsch, was jedoch durch die Bedingung (II) (b) unberücksichtigt bleibt. Wenn ich etwa eine Funktion c^L aus dem Kontinuum mit $L = 10^{10}$ wähle und die Urne 100 Kugeln enthält, so werden die Funktionswerte praktisch mit denen von c^∞ zusammenfallen. Dann bemißt sich die Wahrscheinlichkeit nach der relativen Weite der V-Prädikate, die hier $1/2^6 = 1/64$ ist, während *in Wirklichkeit* nur 6 Möglichkeiten bestehen (da die Kugeln ja einfarbig sind) und daher alle bis auf 6 V-Prädikate die relative Weite 0 haben und diesen 6 daher die relative Weite $1/6$ zukommt.

Weiter könnte gegen die Verwendung von (II) (b) als Adäquatheitsbedingung für den logischen Wahrscheinlichkeitsbegriff geltend gemacht werden, daß zwar in *einigen*, aber durchaus nicht in *allen* Fällen die relative Weite der Spektralfarben durch $1/6$ angegeben werden sollte. In manchen Fällen, etwa bei gewissen Experimenten der Optik, könnte es sich als zweckmäßig erweisen, ihnen als relative Weiten das Verhältnis der Länge, das jede von ihnen im Farbspektrum einnimmt, zur Gesamtlänge des Farbspektrums zu geben; da die Farben im Spektrum nicht die gleiche Länge haben, werden damit die relativen Weiten im allgemeinen vom Wert $1/6$ abweichen. Ähnliches gilt z. B. auch, wenn die Sprache über der Gesamtheit der mit einem bestimmten Würfel gemachten Würfe interpretiert ist und die Eigenschaftsausdrücke $\Pi_1, \ldots, \Pi_6$ die Eigenschaften „1" *oben liegend*, ..., „6" *oben*

liegend bezeichnen, wobei die Kanten dieses Würfels nicht gleich lang und die Seitenflächen nicht gleich groß sind.

Mit den folgenden Überlegungen kann man Argumente dagegen vorbringen, daß (II) (a) in allen Situationen adäquat ist. Der Gegenstandsbereich sei die Gesamtheit jener Menschen, denen eine gewisse Rechenaufgabe, etwa die Lösung eines bestimmten Integrals, vorgelegt worden ist; das einzige Grundprädikat Π_1 bezeichnet die Eigenschaft, die Aufgabe gelöst zu haben. Der Gegenstandsbereich umfasse die Personen Theophil Müller, Dankwart Meier, Norbert Wiener und John von Neumann. Der Person x, die nach einer (im intuitiven Sinn) adäquaten induktiven Methode für diesen Gegenstandsbereich sucht, sei unbekannt, welche und wieviele Rechenaufgaben diese Personen im Laufe ihres Lebens gelöst haben (kaum ein Mensch wird dieses tatsächlich wissen); x kann sich bei der Angabe des Wahrscheinlichkeitsgrades also weder auf eine individuelle noch auf eine statistische Verteilung (in gewissen Gegenstandsausdrücken) stützen, sie hat, mit anderen Worten, überhaupt keine brauchbare Prämisse zur Verfügung. Trotzdem wird sie, was vernünftig erscheint, die Aprioriwahrscheinlichkeit, daß Norbert Wiener jene Aufgabe gelöst hat, als verschieden ansehen von der Aprioriwahrscheinlichkeit, daß Dankwart Meier eine Lösung gefunden hat, und gleiches gilt für Theophil Müller und John von Neumann. Wenn Δ der Satz „Dankwart Meier hat jene Aufgabe gelöst" und Ψ die Aussage „Norbert Wiener hat jene Aufgabe gelöst" ist, so wird x also nach einer Wahrscheinlichkeitsfunktion c suchen, für die sich: $c(\Psi, T) \neq c(\Delta, T)$ ergibt, und analog für die Gegenstandsausdrücke „Theophil Müller" und „John von Neumann"; diese Funktion c wird also die Bedingung (II) (a) verletzen.

Alle Würfel und alle Münzen werden beim Werfen deformiert. Die Deformation wird normalerweise (wenn der Würfel oder die Münze nicht aus Kaugummi besteht) so gering sein, daß sie mit bloßem Auge nicht wahrgenommen werden kann, doch wird man sicherlich bei 1000000 Würfen schon Spuren der Abnützung erkennen können. Solange der Würfel bzw. die Münze bei einer Anzahl k von Würfen noch eine genügend große Ähnlichkeit mit seinem bzw. ihrem Ausgangszustand hat, wird man von einem *praktischen* Standpunkt aus mit Recht sagen können, daß es sich bei den einzelnen Würfen um

Experimente der gleichen Art handelt. Da vom *theoretischen* Standpunkt aus allerdings nach jedem Wurf ein anderer Würfel bzw. eine andere Münze mit einer anderen Gestalt zur Verfügung steht, liegen eben *nicht Experimente der gleichen Art* vor; konsequenterweise müßte eine Person x, die die Sprache über dem Bereich dieser Würfe interpretiert hat, dann eine Funktion c benützen, die ebenfalls die Bedingung (II) (a) verletzt (denn Erfahrungsdaten über die Art und Richtung der Deformation stehen ihr, wie man mit Recht voraussetzen darf, nicht zur Verfügung).

Das Beispiel des deformierten Würfels legt einen Einwand gegen die Bedingung (III) nahe. Bei einem normalen Würfel ist für die Beurteilung der Wahrscheinlichkeit, daß beim nächsten Wurf die Ziffer „6" obenliegt, sicherlich nur das Verhältnis der Anzahlen der bisherigen Würfe, die die „6" ergeben haben, zu denen, bei denen nicht die „6" oben lag, relevant. Gegeben sei jedoch ein Würfel, von dem wir lediglich wissen, daß sein Schwerpunkt nahe bei dem Mittelpunkt einer seiner Seiten liegt, ohne daß uns bekannt ist, um welche es sich handelt. Das aus 50 Würfen bestehende Erfahrungsdatum besage, daß dreizehnmal die „6", achtmal die „5", siebenmal die „4", dreizehnmal die „3", achtmal die „2" und einmal die „1" oben gelegen hat. Dann wird man intuitiv die Wahrscheinlichkeit, daß beim nächsten Wurf die „6" oben liegt, hinsichtlich dieses Datums höher bewerten als die Wahrscheinlichkeit, daß die „3" oben liegt, da die „1" diametral entgegengesetzt zur „6" liegt und der Schwerpunkt offensichtlich nahe der Seite liegt, auf der die Ziffer „1" steht.

Ein ähnliches Beispiel kann man sich auch bei verschiedenfarbigen Kugeln einer Urne ausdenken. An Grundeigenschaften seien die Farben *rot* und *grün* gegeben, so daß *rot und grün*, *rot und nichtgrün*, *grün und nichtrot* und *weder rot noch grün* die V-Eigenschaften sind. Die Person x habe die nicht ungerechtfertigte Vermutung, daß die Kugeln meistens entweder beide Farben oder eine der beiden Farben oder keine von beiden Farben haben. Von 20 beobachteten Kugeln seien nun zwei *rot und grün*, vier *weder rot noch grün*, vier *rot und nichtgrün* und zehn *grün und nichtrot*. Dann wird x vermutlich eine Funktion c als Wahrscheinlichkeitsfunktion wählen, für die gilt: c („die nächste Kugel ist rot und nichtgrün", T) > c („die nächste Kugel ist weder rot noch grün", T), die also die Bedingung (III) verletzt.

Während gegen die Bedingung (IV) (a) nur Glücksspieler Einwände erheben dürften, die etwa nach einer Serie von 50 Würfen mit einem regelmäßigen Würfel, in der keinmal die „6" oben gelegen hat, es als fast notwendig ansehen, daß beim nächsten Wurf die „6" oben liegt, kann gegen (IV) (b) immerhin der ernsthafte Einwand erhoben werden, daß diese Bedingung offensichtlich die Möglichkeit ausschließt, daß in der Welt alles Zufall ist (daß der Würfel vollkommen regelmäßig ist) und daß die Welt deshalb vollständig disuniform ist; (IV) (b) ist dann scheinbar nicht eine Adäquatheitsbedingung für den Begriff der logischen Wahrscheinlichkeit, sondern eine synthetische Voraussetzung über die Welt, die vielleicht in unserer Welt meistens oder immer erfüllt ist, die aber trotzdem etwas *über die Welt* und nicht *über den Wahrscheinlichkeitsbegriff* aussagt.
Es gibt also scheinbar reguläre Bestätigungsmethoden c, die keine (im formalen Sinn) adäquaten induktiven Methoden sind, die jedoch für gewisse Situationen (im intuitiven Sinn) adäquater sind als alle Funktionen des L-Systems. Scheinbar engen die Bedingungen (I) bis (IV) also die Gesamtheit der in Frage kommenden induktiven Methoden zu sehr ein.

3. Die Rechtfertigung der Adäquatheitsbedingungen

Gegeben seien als Gegenstandsbereiche zwei Urnen mit einer bestimmten Anzahl n von Kugeln. Die Grundprädikate $\Pi_1, \Pi_2, \ldots, \Pi_6$ der Sprache mögen die Farben *rot, orange, ..., violett* bezeichnen. Der Uniformitätsgrad der Gegenstände hinsichtlich der V-Eigenschaften sei in beiden Fällen w. Uns sei das folgende Zusatzwissen gegeben: Die Kugeln der ersten Urne sind alle einfarbig, die der zweiten Urne hingegen alle zweifarbig. In der Objektsprache seien Γ' und Γ'' die Sätze, die diese Sachverhalte beschreiben.
Dann gibt es nach den Überlegungen des vorigen Abschnittes zwei Funktionen c' und c'', die offensichtlich voneinander verschieden sind, so daß c' bezüglich der über dem ersten Gegenstandsbereich und c'' bezüglich der über dem zweiten Gegenstandsbereich interpretierten Sprache jeweils geeignetere induktive Methoden sind als c^L mit

$L = (1 - w)/(w - 1/64)$. Die Funktion c' enthält das Wissen Γ' und c'' das Datum Γ'', denn c' gibt allen Zustandsbeschreibungen, die einem Gegenstand nicht genau eine Farbe zuschreiben, den Apriori-wert 0, und c'' tut Gleiches für Zustandsbeschreibungen, die einem Gegenstand nicht genau zwei Farben zuschreiben.

Das erste Problem, das sich uns hier stellt, ist die Frage, ob diese Zusatzdaten als analytische oder als synthetische Urteile anzusehen sind. Carnap hat Sätze dieser Art „Bedeutungspostulate" bzw. „Analytizitätspostulate" genannt[7] und damit indirekt behauptet, es wären Aussagen ohne synthetischen Gehalt, die nichts über die Welt aussagen, sondern lediglich die Bedeutung der in ihnen vorkommenden Begriffe näher charakterisieren. Dies ist jedoch im vorliegenden Fall ohne Zweifel falsch; die Tatsache, daß die Kugeln einfarbig bzw. zweifarbig sind, hat mit der Bedeutung der Begriffe $\Pi_1, \ldots, \Pi_6$ (unter der vorgegebenen Interpretation) nichts zu tun. Ebenso sehe ich nicht ein, warum es nicht logisch möglich (wenn auch vielleicht faktisch ausgeschlossen) sein soll, daß ein Würfel auf der Kante stehenbleibt und also keine Seite oben liegt oder daß er sich bei einem Wurf so deformiert, daß mehrere Seiten oben liegen. Es ist also äußerst unzweckmäßig und irreführend, in Zusammenhängen dieser Art die Begriffe „Analytizitätspostulate" und „Bedeutungspostulate" zu verwenden. Eine gewisse Ähnlichkeit der Bedeutungspostulate Carnaps mit den synthetisch-apriorischen Urteilen Kants ist im übrigen nicht zu übersehen. Doch darf „a priori", wie schon erwähnt, nicht im apodiktischen Sinn wie bei Kant, sondern nur im hypothetischen Sinn verwendet werden, da sich diese Voraussetzungen als falsch erweisen können.

Wenn zugestanden wird, daß Sätze wie Γ' und Γ'' synthetisch sind, so erwartet uns ein zweites Problem: Ist es dann notwendig, die Funktion c' bzw. c'' als synthetischen Begriff einzuführen, der das (in diesem Zusammenhang a priori angenommene) Wissen Γ' bzw. Γ'' bereits implizit enthält, oder ist es möglich, die *beiden* Funktionen c' und c'' in *eine* analytische Methode c^L für ein gewisses L und in die *beiden* synthetischen Aussagen Γ' und Γ'' aufzuspalten? Wenn eine Person x auf Grund eines ihr bewußten Zusatzwissens, das in der Objektsprache durch den Satz Γ beschrieben wird, eine nicht im Kontinuum enthaltene induktive Methode c einer jeden Methode c^L

[7] Vgl. z. B. Carnap-Stegmüller [1], S. 245.

vorzieht, so ist es nicht schwer, c in einen analytischen und einen synthetischen Teil zu trennen: Man erforscht, welcher Uniformitätsgrad w von x vorausgesetzt wird und bestimmt nach Theorem 17 die Zahl L. Wenn dann Γ die gesamten Gründe[8] dafür sind, daß x keine Funktion aus dem Kontinuum benützt, dann ergibt es sich, daß für beliebige Sätze Ψ und Φ: $c(\Psi, \Phi) = c^L(\Psi, \Phi \wedge \Gamma)$, falls Φ und Γ simultan erfüllbar sind. Denn wenn Γ und w (und damit L) fest vorgegeben sind, dann kann c auf diese Weise definiert werden und führt in diesem Fall zu besseren Ergebnissen als alle anderen, im Kontinuum enthaltenen Methoden.

Es kann nun der Fall eintreten, daß das Wissen in einer Sprache mit endlich vielen Gegenstandsausdrücken und einstelligen Grundprädikaten nicht formulierbar ist, da z. B. vorausgesetzt werden muß, daß die relative Weite einer V-Eigenschaft durch eine irrationale Zahl zu charakterisieren ist, wie etwa im Fall der Spektralfarben. Auf zwei Arten kann der Ausdrucksreichtum der Sprache so erweitert werden, daß dieses Wissen dann durch einen ihrer Sätze formulierbar ist: Irrationale Zahlen können, wie schon angedeutet worden ist, durch Sätze von unendlicher Länge charakterisiert werden; sie können aber auch in der bekannten Weise in einer Sprache der einfachen Typenlogik dargestellt werden (im 4. Abschnitt des folgenden Kapitels wird der Weg angedeutet, auf dem man den vorliegenden Ansatz auch auf solche ausdrucksreichere Sprachen erweitern kann).

Sind w und Γ vorgegeben, so kann eine nicht im Kontinuum enthaltene Funktion c in der oben geschilderten Weise mittels einer Methode c^L definiert werden, wobei c nicht explizit auf das Datum Γ Bezug nimmt, es aber trotzdem verwertet. c ist also in diesem Fall eindeutig in eine analytische Methode c^L und in ein synthetisches Wissen Γ zerlegbar. Es stellt sich uns die Frage, ob man ganz allgemein, ausgehend von c das Datum Γ und den Uniformitätsgrad w finden kann.

Wenn c und zusätzlich noch w vorgegeben sind, so bereitet es bei Sprachen mit endlich vielen Gegenstandsausdrücken und einstelligen

[8] Bernoulli hat nach Philodemus als erster darauf hingewiesen, daß zur Berechnung des Wahrscheinlichkeitswertes das gesamte Wissen zu verwenden ist; vgl. Bernoulli [1], S. 215: „Non sufficit expendere unum alterumve argumentum, sed conquirenda sunt omnia, quae in cognitionem nostram venire possunt, atque ullo modo ad probationem rei facere videntur."

Grundprädikaten keine prinzipiellen Schwierigkeiten, das Wissen Γ ausfindig zu machen (wenn es auch vielleicht wegen der erheblichen praktischen Schwierigkeiten zweckmäßiger ist, diese Arbeit einem Automaten zu überlassen). Man braucht lediglich die Sätze der Sprache in eine lineare Anordnung zu bringen und kann dann für jede Zustandsbeschreibung Σ entscheiden, ob $c(\Sigma, T) = c^L(\Sigma, \Gamma)$ oder nicht.

Wenn lediglich c vorgegeben ist, bestimmt man zunächst die endlich vielen Uniformitätsgrade, deren der Gegenstandsbereich hinsichtlich seiner V-Eigenschaften bei den einzelnen (endlich vielen) Zustandsbeschreibungen fähig ist, und wendet dann für jeden Grad w dieses Verfahren an; dieses ist dann allerdings nicht mehr notwendigerweise eindeutig, d. h. es ist nicht ausgeschlossen, daß es wenigstens zwei Grade w' und w'' sowie zwei Daten Γ' und Γ'' gibt, so daß sowohl $c(\Sigma, T) = c^{L'}(\Sigma, \Gamma')$ als auch $c(\Sigma, T) = c^{L''}(\Sigma, \Gamma'')$ für jede Zustandsbeschreibung Σ.

Bei Sprachen mit unendlichem Vokabular und mit Sätzen von unendlicher Länge sind derartige Verfahren offensichtlich nicht mehr mechanisch anwendbar. Hier kann man sich folgendermaßen zu helfen versuchen: Wenn eine Person x eine solche Funktion c an Stelle einer L-Methode benützt, so wird sie gebeten, ihre Gründe für eine derartige Wahl vorzulegen. Wenn sie ihre Gründe korrekt und vollständig darlegt, so werden diese sowohl ihre Voraussetzungen über die Uniformität des Bereichs als auch ihr Zusatzwissen Γ enthalten; damit ist die Aufgabe, w und Γ zu finden, in trivialer Weise gelöst.

In manchen Fällen wird x aber den von ihr vorausgesetzten Grad w und ihr Zusatzwissen Γ gar nicht kennen: Sie wendet c instinktiv an, ohne sich große Gedanken darüber zu machen, ob dieser Funktion bei einer Zerlegung eine L-Methode entspricht oder nicht; sie ist der Ansicht, daß sie mit der Wahl von c in dieser Situation sozusagen *genau richtig liegt*. Wenn dies der Fall ist, bleibt einem nichts anderes übrig als entweder x zu bitten, sich doch einmal ihre Gründe zu überlegen, da diese doch leichter zu finden seien als die genauen Werte $c(\Sigma, T)$ für die einzelnen Zustandsbeschreibungen Σ, oder, wenn dies nicht weiterführt, sich selbst hinzusetzen und mit etwas Glück und Geschick nach den Voraussetzungen der Person x zu suchen.

Gegen diesen pragmatischen Lösungsvorschlag könnte man einwen-

den, daß doch üblicherweise den meisten Menschen nicht genau bewußt sein wird, welches Zusatzwissen Γ sie bei Wahrscheinlichkeitsargumenten benützen, und daß eine Analyse des Begriffs der logischen Wahrscheinlichkeit dies eigentlich zu berücksichtigen hätte. Dem möchte ich folgendes erwidern: Auch beim deduktiven Argumentieren sind den wenigsten Menschen sämtliche Prämissen bekannt, die logisch zu ihren jeweiligen Konklusionen führen. Trotzdem hält man es für sinnvoll, einen Begriff der logischen Folgerung zu entwickeln, der nicht bereits implizit auf synthetische Aussagen als verborgene Prämissen Bezug nimmt, selbst wenn man zugibt, daß er dann eigentlich nur auf Rechenautomaten zutrifft. Analoges gilt für die induktiven Argumente: So wie man die deduktive Logik einerseits als eine reine Theorie entwickeln und sie andererseits in eine angewandte überführen kann, indem man synthetische Aussagen zu ihren Grundsätzen hinzunimmt, so kann man auch die induktive Logik zunächst als ein analytisches System entwickeln und sie dann in eine angewandte Disziplin überführen, indem mittels eines Zusatzwissens (einer synthetischen Theorie) und einer L-Methode eine Funktion c definiert wird, die nicht im Kontinuum liegt. Daß der Übergang von einer reinen zu einer angewandten induktiven Logik in dieser klaren Form nur von Rechenautomaten vollzogen wird, ist dabei kein theoretischer Einwand, sondern kann vielmehr als Aufforderung aufgefaßt werden, das im faktischen Leben geübte Schließen immer mehr den Normen der deduktiven wie der induktiven Logik anzugleichen.

Während man die Bedingungen (II) und (III) also durch Hinweis auf die Forderung des Gesamtdatums (der vollständigen Erwähnung aller relevanten synthetischen Aussagen in der Prämisse des Wahrscheinlichkeitsschlusses) rechtfertigen kann, zeigt man die Adäquatheit der Linearitätsbedingung (der in D 20 formulierten Bedingung (III')) am einfachsten durch Reduktion auf das folgende Prinzip: Die Funktion c muß so eingeführt sein, daß der Funktionswert c (Ψ, Φ) nur von den Sätzen Ψ und Φ abhängt und unabhängig davon ist, was die Sprache sonst noch an Aussagen enthält.

Es ist schwer, direkte Gründe für die Adäquatheit der Linearitätsbedingung zu liefern. Zwar sprechen Einfachheitsüberlegungen für sie, doch könnte man sich in diesem Fall auch Funktionen vorstellen, die

zwar nicht linear von k_i abhängen, die aber trotzdem hinreichend einfach sind, die etwa vom Quadrat von k_i abhängen. Diese würden allerdings gegen jenes Prinzip verstoßen; denn wenn zwei Sätze Ψ und Φ an einstelligen Grundprädikaten nur Π_1 enthalten, so hängt der Funktionswert c (Ψ, Φ) dann davon ab, ob die Sprache nur Π_1 oder zusätzlich noch Π_2 enthält: Wenn sie mehr als ein Grundprädikat (und also Sätze mit diesen zusätzlichen Prädikaten) enthält, so ist c ja bereits wegen (III) eine lineare Funktion von k_i; nimmt man diese zusätzlichen Ausdrücke (Sätze), die mit Ψ und Φ garnichts zu tun haben, aus der Sprache heraus, so muß sich der Wert c (Ψ, Φ) ändern, wenn c gegen (III′) verstößt.

Die Adäquatheit der beiden Limeskonventionen ergibt sich unmittelbar, wenn man bedenkt, daß Sätze von endlicher Länge auf Grund jenes Prinzips in Sprachen mit unendlich vielen Gegenstandsausdrücken und/oder unendlich vielen einstelligen Grundprädikaten den gleichen c-Wert haben müssen wie in Sprachen, die nur das in diesen Sätzen vorkommende Vokabular enthalten, und daß Sätze von unendlicher Länge als Folgen von Sätzen mit endlicher Länge aufgefaßt werden können und die c-Werte für unendliche Sätze daher durch die entsprechenden Folgen von c-Werten für endliche Sätze eindeutig bestimmt sind.

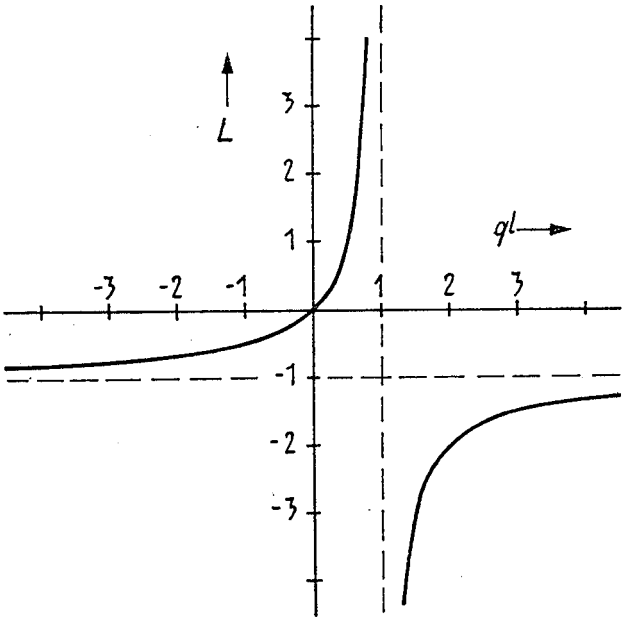
Zur Bedingung (IV) (a) ist bisher im Grunde nur polemisch bemerkt worden, daß ihre Verletzung das Verhalten von gewissen Glücksspielern kennzeichnet, die annehmen, daß etwa beim Würfelwurf die Hypothese, daß beim nächsten Wurf die „6" oben liegt, um so wahrscheinlicher wird, je länger es schon zurückliegt, daß eine „6" beobachtet worden ist. So richtig diese Beobachtung ist, so wenig trägt sie dazu bei, (IV) (a) als für den Wahrscheinlichkeitsbegriff adäquat zu erweisen. Einen Nachweis liefert jedoch der folgende Lehrsatz, der die Adäquatheit von (IV) (a) auf die anderer Bedingungen zurückführt:

Theorem 18

B sei abzählbar unendlich, und c erfülle die Bedingungen (I), (II) und (III) bzw. (III′) sowie die Limeskonvention für Gegenstandsausdrücke; dann erfüllt c auch (IV) (a).

Beweis

Die Voraussetzungen seien gegeben.
Auf Grund einer früheren Festsetzung gilt: $L = ql/(1 - ql)$. Die Abhängigkeit von L von ql wird dann durch die folgende Figur veranschaulicht:



L in Abhängigkeit von ql

Fig. 4

Dann ergibt sich für L:
(1) $L \neq 0$: Wegen (6).
(2) $L \neq -1$. Denn angenommen, $L = -1$, dann gilt: $ql = ql - 1$. Für endliches ql ist dies ein Widerspruch; folglich gilt: $ql = \infty$. Da q nach Voraussetzung endlich ist, folgt daraus, daß $l = \infty$, im Widerspruch dazu, daß $l < 1$ für eine reguläre Bestätigungsmethode c.
(3) Es ist $L = -\infty$ genau dann, wenn die Bedingung (IV) (b) verletzt ist; denn es gilt:

$$\lim_{L\to+\infty} c^L(\Psi_{i_{k+1}}, \Phi_k) = \lim_{L\to+\infty} \frac{k_i + L/q}{k+L}$$
$$= \frac{1}{q}$$
$$= \lim_{L\to-\infty} \frac{k_i + L/q}{k+L}$$
$$= \lim_{L\to-\infty} c^L(\Psi_{i_{k+1}}, \Phi_k).$$

(4) Nicht: $-1 < L < 0$. Denn angenommen, $-1 < L < 0$, dann gilt: $ql = L(1 - ql) = L/(1 + L) < 0$; da q endlich und positiv ist, folgt daraus: $l < 0$, was der Voraussetzung widerspricht, daß c eine reguläre Bestätigungsmethode ist und also stets $l > 0$ gilt.
(5) Nicht: $-\infty < L < -1$. Denn angenommen, L ist eine feste Zahl im offenen Intervall $(-\infty, -1)$; dann liegt L zwischen zwei benachbarten ganzen Zahlen K und $K+1$, d. h. es ist $K \leqslant L < K+1$. Es ist also auf alle Fälle $K < -1$. Nun ist $1/q \leqslant 1/2$; es wird daher angenommen, daß $m/q = 1/2$. Es sei ferner $k_j = k = -(K+1)$; dann gilt für den Nenner des Bruchs von Korollar 7:
$k + L = -(K+1) + L < -(K+1) + (K+1) = 0$, und damit $k + L < 0$. Für den Zähler gilt darüber hinaus:
$k_j + L/2 = -(K+1) + L/2$.
Fall (a): Es sei $-2 < L < -1$, so daß $L/2 > -1$ sowie $K = -2$. Dann gilt: $-(K+1) + L/2 = 1 + L/2 > 0$, und folglich:

$\frac{k_j + (m/q)\cdot L}{k+L} < 0$, was den Voraussetzungen widerspricht.

Fall (b): Es sei $L = -2$, so daß also $K = -2$ und damit $k = 1$; dann gilt: $-(K+1) + L/2 = 1 - 2/2 = 0$. Daraus folgt:

$\frac{k_j + (m/q)\cdot L}{k} = 0$, was der Regularität widerspricht.

[Für $k + 1 = 2$ erhält dieser Bruch darüber hinaus den Wert $(2-1)/(2-2) = \infty$, was ebenfalls den Voraussetzungen widerspricht.]
Fall (c): Es sei $-\infty < L < -2$, so daß $K < -2$. Auf Grund der Limeskonvention für Gegenstandsausdrücke darf angenommen wer-

den, daß der Gegenstandsbereich unendlich ist; k kann dann beliebig groß gewählt werden, so daß die Voraussetzung, daß $k_j = k = -(K+1)$ in jedem Fall erfüllt bleibt. Es gilt: $K/2 \leqslant L/2$, und damit: $-(K+1) + L/2 \geqslant -(K+1) + K/2 > 0$ [wegen $K < -2$, so daß $-(K+1) > K/2$]. Also ist:

$$\frac{k_j + (m/q) \cdot L}{k + L} < 0,$$ was den Voraussetzungen widerspricht.

Damit ist das Theorem bewiesen.

Reichenbach hat versucht, die Wahrscheinlichkeit eines Ereignisses als den Grenzwert einer unendlichen Folge von relativen Häufigkeiten zu explizieren. Er ist sich dabei allerdings im klaren darüber gewesen, daß uns stets nur endliche Beobachtungen zur Verfügung stehen; deshalb hat er den Wahrscheinlichkeitswert mittels c^0 berechnet (ihn also mit der beobachteten relativen Häufigkeit identifiziert) und dies folgendermaßen begründet: Ohne Zweifel wird der c^0-Wert vom *objektiven Wahrscheinlichkeitswert* (vom Grenzwert der relativen Häufigkeiten) verschieden sein; aber c^0 ist *selbstkorrigierend*, d. h. mit wachsendem k gleicht sich der c^0-Wert k_j/k dem Grenzwert der relativen Häufigkeiten (falls dieser existiert) immer mehr an.[9]
In einer etwas schwächeren Formulierung, die *nicht* die Existenz eines Grenzwertes der relativen Häufigkeiten voraussetzt (so daß diese relativen Häufigkeiten auch divergieren können), besagt diese Forderung des Selbstkorrigierens der Methode c: Mit wachsendem k wird die Differenz zwischen dem Wahrscheinlichkeitswert $c\,(\Delta^m_{j_{k+1}}, \Phi_k)$ und der beobachteten relativen Häufigkeit k_j/k immer kleiner (vorausgesetzt wird dabei, daß die Anzahl n der Gegenstände größer ist als die einzelnen Anzahlen k der jeweiligen Beobachtungen).
Diese Forderung, die in der Literatur meist „Reichenbachprinzip" genannt wird, ist zweifellos eine notwendige Bedingung dafür, daß eine induktive Methode c adäquat ist; denn wenn man auch aus den im I. Kapitel dargelegten Erwägungen heraus Hemmungen haben wird, die Wahrscheinlichkeit mit einem Grenzwert von relativen Häufigkeiten zu identifizieren und wenn man auch bei einem beschei-

[9] Vgl. Reichenbach [10], §§ 85—89.

denen Beobachtungsdatum mit Recht der beobachteten relativen Häufigkeit kein zu großes Gewicht beimißt, so wird man doch von einem adäquaten Wahrscheinlichkeitsbegriff verlangen, daß sich seine Werte mit wachsendem k immer mehr der beobachteten relativen Häufigkeit angleichen und so die a priori vorgenommene Bestimmung der relativen Weite des Prädikates a posteriori korrigiert wird. Unter Verwendung der in Korollar 7 beschriebenen Sätze $\Delta^{m}_{j_{k+1}}$ und Φ_k kann das Reichenbachprinzip wie folgt definiert werden:

D 29 B sei ein abzählbar unendlicher Bereich, und c sei eine reguläre Bestätigungsmethode, die die Bedingungen (II) und (III) bzw. (III') sowie die Limeskonvention für Gegenstandsausdrücke erfüllt.
Dann erfüllt c das Reichenbachprinzip genau dann, wenn für alle Sätze $\Delta^{m}_{j_{k+1}}$ und Φ_k der beschriebenen Art gilt:
$\lim_{k \to \infty} [c(\Delta^{m}_{j_{k+1}}, \Phi_k) - k_j/k] = 0.$

Offensichtlich ist nicht nur c^0, sondern jede L-Methode (und darüber hinaus auch manche andere Funktion, die nicht im Kontinuum liegt) im Sinne Reichenbachs selbstkorrigierend; das Reichenbachprinzip ist also zwar eine *notwendige* Bedingung für den Wahrscheinlichkeitsbegriff (für den Satz „c ist eine adäquate Wahrscheinlichkeitsfunktion"), stellt jedoch keine *hinreichende* Charakterisierung desselben dar. Es kann nun gezeigt werden, daß die Methode c^∞ nicht selbstkorrigierend ist und daß (IV) (b) auf der Grundlage der anderen Bedingungen mit dem Reichenbachprinzip äquivalent ist:

Theorem 19

B sei abzählbar unendlich, und c erfülle die Bedingungen (I), (II) und (III) bzw. (III') sowie die Limeskonvention für Gegenstandsausdrücke. Dann genügt c der Bedingung (IV) (b) genau dann, wenn c das Reichenbachprinzip erfüllt.

Beweis

c erfülle die Limeskonvention für Gegenstandsausdrücke sowie (I) bis (III).

(a) c genüge der Forderung (IV) (b). Mit Korollar 7 gilt dann:

$$\lim_{k \to \infty} \left[\frac{k_j + (m/q) \cdot L}{k + L} - \frac{k_j}{k} \right] = 0 .$$

(b) Angenommen, c verletzt die Bedingung (IV) (b), so daß nach Theorem 14 also $L = \infty$; dann ist $c\,(\Delta^m_{j_{k+1}}, \Phi_k) = m/q$. Es sei $m = 1$ und $\Delta^m_{j_{k+1}}$ also einer der Sätze $\Psi_{i_{k+1}}$. Es seien Φ_k und $\Psi_{i_{k+1}}$ Konjunktionsglieder einer unendlichen Standardkonjunktion von Sätzen der Form $\alpha_1 \in \Pi^*_i \wedge \alpha_2 \in \Pi^*_i \wedge \ldots$ (oder aber Konjunktionsglieder eines Satzes aus einer Folge von immer größeren endlichen Konjunktionen dieser Art). Dann ist die relative Häufigkeit in jedem Fall 1, und es gilt:

$\lim_{k \to \infty} \left[\frac{1}{q} - 1 \right] \neq 0$, womit das Theorem bewiesen ist.

Die Bedingungen (II), (III) und (IV) sind also adäquate Forderungen an den logischen Wahrscheinlichkeitsbegriff und brauchen nicht als empirische Voraussetzungen gedeutet zu werden. Gleiches gilt von der Bedingung (I), die ja nach Korollar 3 mit der strengen Kohärenz von c äquivalent ist: Eine Bestätigungsmethode, die (I) verletzt, verleitet uns dazu, Wetten abzuschließen, die zwar sicherlich mit keinem Gewinn, möglicherweise jedoch mit einem Verlust enden. Eine solche Methode wird man jedoch nicht als adäquat ansehen können.

Die Bedingung (I), bzw. genauer: die Bedingung (6) von D 4, und die Forderung (IV) (b) geben uns die folgende Paradoxie zu lösen auf: Einerseits sind sie als *analytische* Folgerungen aus der Behauptung „c ist eine adäquate induktive Methode" anzusehen, aber andererseits scheinen sie synthetische Aussagen über den Gegenstandsbereich zu sein, da sie nach Theorem 17 besagen, daß der Gegenstandsbereich nicht völlig disuniform und auch nicht vollständig uniform ist, da sie also vollständigen Zufall und störungsfreies Experimentieren ausschließen. Auch wenn beides in unserer *faktischen* Welt vielleicht nicht anzutreffen ist und die Bedingungen daher vielleicht gelten, ist doch eine

Welt logisch *möglich*, in der es durchgehenden Zufall oder aber störungsfreies Experimentieren gibt, so daß jene Bedingungen demnach als *synthetische* Aussagen über die Welt anzusehen sind.
Von den unendlich vielen Funktionen c, die (I) (b) verletzen, ist jedoch nur eine interessant, nämlich c^0, und desgleichen von den unendlich vielen, die dem Reichenbachprinzip nicht genügen, nur c^∞. Statt auf diese Bedingungen zu verzichten, ist es deshalb zweckmäßiger, die Methoden c^0 und c^∞ den L-Methoden hinzuzufügen (sie stellen ja sozusagen die beiden Endpunkte des Kontinuums dar) und etwa zu definieren, daß eine Funktion c im weiteren Sinn adäquat ist, wenn sie im früher definierten Sinn adäquat ist oder wenn sie mit c^0 oder mit c^∞ identisch ist.
Ich glaube jedoch, daß man auf eine solche Definition verzichten kann. Denn diese beiden extremen Methoden c^0 und c^∞ können ja durch L-Methoden beliebig approximiert werden; in vielen Fällen etwa wird man $c^{1/10^{10}}$ als guten Ersatz für c^0 und $c^{10^{10}}$ als guten Ersatz für c^∞ ansehen können. Ob man dann den Wahrscheinlichkeitswert tatsächlich durch $c^{1/10^{10}}$ oder durch c^0 bzw. durch $c^{10^{10}}$ oder durch c^∞ berechnet, ist dabei praktisch ohne Belang, da sich die Werte erst an Stellen hinter dem Komma voneinander unterscheiden, die in der Anwendung ohnehin nicht berücksichtigt werden. Da man die beiden Endpunkte des Kontinuums auf diese Weise beliebig approximieren kann, hat man sie bereits im Griff; ob man sie dann auch noch in ein erweitertes Kontinuum aufnehmen soll, halte ich für zweitrangig.
Zur Frage der Unabhängigkeit der einzelnen Adäquatheitsbedingungen voneinander sei noch bemerkt, daß für Sprachen mit endlich vielen Gegenstandsausdrücken (IV) (a) teilweise von den übrigen Bedingungen abhängt, wie Theorem 18 zeigt, und daß sie bei einem abzählbar unendlichen Gegenstandsbereich aus jenen und der Limeskonvention für Gegenstandsausdrücke logisch folgt. Bedingung (IV) (b) folgt teilweise aus (I), kann jedoch dahingehend abgeschwächt werden, daß sie von ihr unabhängig ist.[10] Im übrigen sind die einzelnen Bedingungen voneinander logisch unabhängig, wie man leicht dadurch zeigen kann, daß man für jede einzelne von ihnen Funktionen c angibt, die zwar alle anderen, nicht jedoch sie erfüllen.

---

10 Vgl. die Bemerkungen im Anschluß an D 19 auf S. 147 dieser Arbeit.

4. *Die Wahrscheinlichkeit genereller Hypothesen*

In der traditionellen wie auch teilweise in der modernen Philosophie wird die Angabe des Bestätigungsgrades genereller Hypothesen auf Grund von endlichen Erfahrungsdaten als die Hauptaufgabe, wenn nicht gar als das einzige Betätigungsgebiet der induktiven Logik angesehen. Aristoteles hat ja *ἐπαγωγή* (inductio) als *Herbeischaffung* der allgemeinen Prinzipien (von einem endlichen Erfahrungsdatum aus) aufgefaßt. Es sei abschließend noch untersucht, ob die L-Methoden zu diesem Problem etwas beitragen und wie es in dieser Hinsicht um ihre Adäquatheit (und damit die der Adäquatheitsbedingungen) bestellt ist.

Die Bedeutung eines Satzes $\Delta^{m}_{j_{k+1}}$ ändert sich nicht, wenn dem Bereich B Gegenstände und der Sprache damit Gegenstandsausdrücke hinzugefügt werden. Wohl aber ändern sich bei solchen Erweiterungen die Bedeutungen genereller Aussagen, also der All- und Existenzbehauptungen: In der über B_n (dem aus n Gegenständen bestehenden Bereich B) interpretierten Sprache kann $\Lambda\, \xi\, \Psi\, (\xi)$ durch $\Psi\, (\alpha_1) \wedge \ldots \wedge \Psi\, (\alpha_n)$ sowie $V\, \xi\, \Psi\, (\xi)$ durch $\Psi\, (\alpha_1) \vee \ldots \vee \Psi\, (\alpha_n)$ ausgedrückt werden, in der über B_{n+1} interpretierten hingegen durch $\Psi\, (\alpha_1) \wedge \ldots \wedge \Psi\, (\alpha_n) \wedge \Psi\, (\alpha_{n+1})$ sowie durch $\Psi\, (\alpha_1) \vee \ldots \vee \Psi\, (\alpha_n) \vee \Psi\, (\alpha_{n+1})$. In der über B_∞ interpretierten Sprache müssen sie dann sinngemäß durch unendliche Konjunktionen $\Psi\, (\alpha_1) \wedge \Psi\, (\alpha_2) \wedge \ldots$ sowie durch unendliche Adjunktionen $\Psi\, (\alpha_1) \vee \Psi\, (\alpha_2) \vee \ldots$ wiedergegeben werden. Angenommen, k ist eine feste Anzahl von Beobachtungen und n ist endlich; ist n dabei nur um weniges größer als k, dann sagt $\Psi\, (\alpha_1) \wedge \ldots \wedge \Psi\, (\alpha_k) \wedge \ldots \wedge \Psi\, (\alpha_n)$ im intuitiven Sinn nicht wesentlich mehr aus als $\Psi\, (\alpha_1) \wedge \ldots \wedge \Psi\, (\alpha_k)$, so daß zu erwarten ist, daß die Wahrscheinlichkeit dieser Allaussage relativ zu dem vorgegebenen Datum sehr groß ist. Läßt man n anwachsen, wobei k konstant bleibt, so wächst der Gehalt der Allaussage $\Psi\, (\alpha_1) \wedge \ldots \wedge \Psi\, (\alpha_k) \wedge \ldots \wedge \Psi\, (\alpha_n)$ relativ zum Datum $\Psi\, (\alpha_1) \wedge \ldots \wedge \Psi\, (\alpha_k)$, so daß der Wahrscheinlichkeitswert dann immer kleiner wird. Ist der Bereich B unendlich, so behauptet der Allsatz unvergleichbar mehr als die endliche Beobachtung; es ist daher zu erwarten, daß die Allbehauptung relativ zur endlichen Beobachtung den Wahrscheinlichkeitswert 0 erhält. Dies kommt durch den folgenden Lehrsatz zum Ausdruck:

Theorem 20

B sei unendlich, gegeben sei ein Allsatz $\Lambda\,\xi\,\Psi(\xi)$, ausgedrückt in der Sprache durch eine unendliche Konjunktion $\Psi(\alpha_1) \wedge \Psi(\alpha_2) \wedge \ldots \wedge \Psi(\alpha_k) \wedge \ldots$, deren Konjunktionsglieder $\Psi(\alpha_i)$ von endlicher Länge sind und keine Quantoren enthalten, die relative Weite von $\Psi(\alpha)$ sei größer als 0 und kleiner als 1, und c sei eine adäquate induktive Methode, die die Limeskonvention für Gegenstandsausdrücke erfüllt. Dann gilt für jede beliebige natürliche Zahl k:
$c(\Lambda\,\xi\,\Psi(\xi), \Psi(\alpha_1) \wedge \ldots \wedge \Psi(\alpha_k)) = 0.$

Beweis

Die Voraussetzungen seien gegeben. Dann gilt:

$$c(\Lambda\,\xi\,\Psi(\xi), \Psi(\alpha_1) \wedge \ldots \wedge \Psi(\alpha_k)) = \frac{c(\Lambda\,\xi\,\Psi(\xi) \wedge \Psi(\alpha_1) \wedge \ldots \wedge \Psi(\alpha_k), T)}{c(\Psi(\alpha_1) \wedge \ldots \wedge \Psi(\alpha_k), T)} = \frac{r}{s}.$$

Wegen der Regularität von c ist auf alle Fälle $s \neq 0$. Zu zeigen ist, daß $r = 0$. Nun ist:

$$\begin{aligned} r &= c(\Lambda\,\xi\,\Psi(\xi) \wedge \Psi(\alpha_1) \wedge \ldots \wedge \Psi(\alpha_k), T) \\ &= c(\Lambda\,\xi\,\Psi(\xi), T) \\ &= c(\Psi(\alpha_1) \wedge \Psi(\alpha_2) \wedge \ldots \wedge \Psi(\alpha_k) \wedge \ldots, T) \\ &= c(\Psi(\alpha_1), T) \cdot c(\Psi(\alpha_2), \Psi(\alpha_1)) \cdot c(\Psi(\alpha_3), \Psi(\alpha_1) \wedge \Psi(\alpha_2)) \cdot \ldots \\ &= \prod_{k=0}^{\infty} \frac{k + (m/q) \cdot L}{k + L} \\ &= \prod_{k=0}^{\infty} \left[1 - \frac{(1 - m/q) \cdot L}{k + L}\right]. \end{aligned}$$

[Der Übergang zur Normalform ist dabei durch Korollar 9 gerechtfertigt.]

Nun ist $\frac{(1 - m/q) \cdot L}{k + L} > 0$ für alle $k \geqslant 0$. Dann gilt:[11]

[11] Vgl. Konrad Knopp, Theorie und Anwendung der unendlichen Reihen, Berlin und Heidelberg [4]1947, S. 227, Satz 4. (Diesen Hinweis verdanke ich Fräulein Dipl.-Math. Veronika Reiß.)

$0 < \prod_{k=0}^{\infty} \left[1 - \frac{(1 - m/q) \cdot L}{k + L}\right] < \infty$ genau dann, wenn

$\sum_{k=0}^{\infty} \frac{(1 - m/q) \cdot L}{k + L} < \infty$.

Nun gilt jedoch: $\sum_{k=0}^{\infty} \frac{(1 - m/q) \cdot L}{k + L} = (1 - m/q) \cdot L \cdot \sum_{k=0}^{\infty} \frac{1}{k + L}$.

L sei eine natürliche Zahl; dann ist $\sum_{k=0}^{\infty} \frac{1}{k + L}$ die um ein endliches Anfangsstück verminderte harmonische Reihe und also unendlich. Ist L keine natürliche Zahl und ist L^* die auf L unmittelbar folgende natürliche Zahl, so ist jeder der einzelnen Brüche $\frac{1}{k + L^*} < \frac{1}{k + L}$ und damit $\infty = \sum_{k=0}^{\infty} \frac{1}{k + L^*} \leqslant \sum_{k=0}^{\infty} \frac{1}{k + L}$. Also ist auch $\sum_{k=0}^{\infty} \frac{(1 - m/q) \cdot L}{k + L} = \infty$.

Nun gilt: $\prod_{k=0}^{\infty} \frac{k + (m/q) \cdot L}{k + L} \leqslant \frac{m}{q} < 1 < \infty$, und damit:

$\prod_{k=0}^{\infty} \frac{k + (m/q) \cdot L}{k + L} = r = 0$. Damit ist das Theorem bewiesen.

Korollar 10

Die Voraussetzungen des Theorems 20 seien gegeben.
Dann gilt für jede beliebige Zahl k:
$c(V\,\xi\,\Psi(\xi), \neg\,\Psi(\alpha_1) \wedge \ldots \wedge \neg\,\Psi(\alpha_k)) = 1$.

Beweis

Dieser Folgesatz ergibt sich unmittelbar aus den Theoremen 20 und 1 (d) unter Berücksichtigung der Tatsache, daß $V\,\xi\,\Psi(\xi)$ logisch äquivalent mit $\neg\,\Lambda\,\xi\,\neg\,\Psi(\xi)$ ist.

Es gibt also nicht nur kein Verfahren, das mechanisch von Beobachtungen zu Allhypothesen führt, sondern es können derartige generelle Hypothesen (bei einem unendlichen Gegenstandsbereich) auf Grund

einer endlichen Beobachtung darüber hinaus durch eine adäquate induktive Methode c nicht als mehr oder weniger gut bestätigt erkannt werden: Sie erhalten in solchen Bereich unterschiedslos den Wahrscheinlichkeitswert 0, gleichgültig, wie groß Beobachtung und relative Weite der Prädikate sind.[12] Dies scheint allen Prinzipien der Induktion zu widersprechen, und es sind daher auch gelegentlich Versuche unternommen worden, durch Modifikationen zu scheinbar adäquateren Bestätigungsmethoden zu gelangen.[13]

Carnap selbst hat vorgeschlagen, die Bestätigung eines Gesetzes (einer Allaussage) an der Bestätigung ihrer Einzelfälle zu messen.[14] Daß der Bestätigungsgrad eines Allsatzes auf Grund eines endlichen Erfahrungsdatums mit der Wahrscheinlichkeit eines beliebigen Einzelfalls dieser Allbehauptung hinsichtlich des gleichen Datums identisch ist, kann man schon aus den Ausführungen des Aristoteles und des Philodemus herauslesen und ist klar und bewußt von Mill formuliert worden.[15] Es erweist sich (im Einklang mit Korollar 7), daß der Einzelfall eine hohe Bestätigung erhalten kann, auch wenn der Wahrscheinlichkeitswert der generellen Hypothese 0 ist.

Dennoch liefert dieser Ausweg nach meinem Dafürhalten keine Lösung des Problems. Wenn z. B. das Erfahrungsdatum aus 100000 Beobachtungen besteht, die zwei negative Instanzen enthalten, so wird sich der Wahrscheinlichkeitswert für den Einzelfall (vorausgesetzt, daß $L \ll 100000$) nur unmerklich von dem Wahrscheinlichkeitswert unterscheiden, den der gleiche Einzelfall bezüglich einer Beobachtung von 100000 ausschließlich positiven Instanzen besitzt. Im einen Fall ist die Allhypothese widerlegt, im anderen ist sie es aber (noch) nicht.

Die Inadäquatheit dieses Lösungsvorschlags kann z. B. dadurch beseitigt werden, daß man ihn folgendermaßen modifiziert: Eine Allbehauptung wird durch ein Erfahrungsdatum im Grad 0 bestätigt, wenn

---

12 In endlichen Bereichen können sich freilich Unterschiede im Wahrscheinlichkeitswert ergeben, doch werden diese mit wachsendem n immer kleiner.

13 Vgl. z. B. Hintikka [1], [2], [3] und [4] (vgl. S. 342—350 dieses Buches).

14 Vgl. Carnap [9], S. 517ff.

15 Vgl. das auf den Seiten 17, 19 und 30f. dieser Arbeit Gesagte. Auch Popper ist vermutlich hierher zu rechnen, da sein Bewährungsgrad einer Theorie ja das Vertrauen ausdrücken soll, mit dem wir damit rechnen, daß sie auch weiteren Falsifikationsversuchen standhält (gleichgültig, wie groß dieses Vertrauen ist).

sie mit ihm logisch unverträglich ist, und sie erhält ansonsten den Bestätigungsgrad ihrer Einzelfälle. Doch auch hiergegen sind ernsthafte Einwände möglich. Die Wahrscheinlichkeit einer Allaussage läßt sich sicherlich nicht durch die ihrer Einzelfälle messen, wie die folgende Überlegung zeigt.

L kann jede positive reelle Zahl sein, je nachdem, wie man die Uniformität des Gegenstandsbereichs einschätzt. $\Psi(\alpha)$ besage bei J, daß a die Eigenschaft F^+ habe. Dann ist, wie im nächsten Abschnitt gezeigt wird, $c(\Psi(\alpha_{100001}, \Psi(\alpha_1) \wedge \ldots \wedge \Psi(\alpha_{100000}))$ mit der Schätzung der (tatsächlichen) relativen Häufigkeit von F^+ im noch nicht beobachteten Teil von B identisch. Diese Schätzung ergibt den Wert m/q der relativen Weite des Satzes $\Psi(\alpha)$ bei $L = \infty$, den Wert 1 bei $L = 0$ und bei positivem und endlichem L den dazwischenliegenden Wert $[10^5 + (m/q) \cdot L]/[10^5 + L]$, der also jedenfalls kleiner als 1 ist. Wir rechnen demnach bei Benützung einer adäquaten induktiven Methode c^L durchaus damit, daß einmal ein Gegenbeispiel beobachtet wird, wenn der Gegenstandsbereich nur groß genug ist (nach Voraussetzung ist er sogar unendlich).

Wenn r/s der von 1 verschiedene c^L-Wert ist (nach Korollar 7 kommen als Werte bei endlichem B und rationalem L nur rationale Zahlen in Frage), so schätzen wir, daß sich unter den nächsten s Gegenständen, die wir beobachten werden, s — r Dinge befinden werden, die *nicht* die Eigenschaft F^+ besitzen, unter den nächsten 2 s Gegenständen demnach 2 (s — r) widersprechende Instanzen usw., und unter den unendlich vielen nicht beobachteten Gegenständen daher sogar unendlich viele Gegenbeispiele. Je kleiner L ist, um so näher liegt der Wert r/s bei 1, um so später erwarten wir also das Auftreten einer Gegeninstanz und um so mehr vertrauen wir der Uniformität des Gegenstandsbereichs (der Natur) hinsichtlich der in Erwägung gezogenen V-Eigenschaften, und umgekehrt. Welches statistische Gesetz durch jenes Beobachtungsdatum geschätzt wird, hängt also davon ab, wie man die Uniformität der Natur einschätzt. Nur wer ein blindes Vertrauen in sie hat, wird $L = 0$ setzen und damit auch in einem unendlichen Bereich nicht mehr mit einem Gegenbeispiel rechnen.

Doch dieser naheliegende Ausweg, an Stelle von Funktionen aus dem Kontinuum die durch $L = 0$ charakterisierte Proportionalregel zu wählen, führt ebenfalls zu keiner Lösung des Problems. Solange das

Erfahrungsdatum nämlich keine negative Instanz und wenigstens eine positive enthält, ordnet c^0 dem Allsatz auf Grund dieses Datums den Wert 1 zu, gleichgültig, um welche Hypothese es sich handelt, wie umfangreich die bestätigenden Instanzen sind und welche relative Weite das Prädikat der Hypothese besitzt. Bei wenigstens einem Gegenbeispiel hingegen ist der c^0-Wert wiederum 0, da sich Hypothese und Erfahrungsdatum in diesem Fall widersprechen. Es ist also nicht möglich, mittels c^0 etwa die Wahrscheinlichkeiten dreier Hypothesen auf Grund eines vorgegebenen Erfahrungsdatums zu vergleichen und zu dem Ergebnis zu gelangen, daß die eine besser bestätigt sei als die zweite und diese wiederum besser als die dritte.

Damit soll nicht gesagt werden, es habe keinen Sinn, davon zu reden, daß der Grad an Gewißheit der einen Hypothese größer sei als der einer anderen; infrage gestellt ist damit nur die Ansicht, daß sich dieser Grad an Gewißheit durch eine Funktion aus dem Kontinuum und, wenn die obigen Überlegungen in bezug auf die Adäquatheit zutreffen, mit dem Wahrscheinlichkeitsbegriff messen läßt. Wir ziehen die eine Theorie der anderen nicht deshalb vor, weil jene vielleicht einige positive Instanzen mehr aufweist als diese, sondern wir tun dies aus Gründen, die auf Beobachtungen oft nur indirekt Bezug nehmen und bei denen die Frage der Einfachheit eine große Rolle spielt. Doch darauf komme ich an späterer Stelle noch zurück.

Theorem 20 stellt also nicht das Ende aller Induktion dar, wie es zunächst erscheinen mag. Allerdings verschärft es Humes Ergebnis zum Problem, ob irgendwelche Argumente von der Erfahrung zu den allgemeinen Gesetzen führen. Wie Hume gezeigt hat, können Allsätze und damit Naturgesetze nicht deduktiv aus einer endlichen Gesamtheit von Erfahrungen abgeleitet werden, da sie mit diesen Erfahrungen nicht logisch äquivalent sind, und sie können ohne Zusatzannahmen über die Uniformität der Natur auch nicht induktiv aus ihnen erschlossen werden.

Setzt man nun einen Uniformitätsgrad voraus, der kleiner als 1 ist (offensichtlich gilt Theorem 20 auch für die Methode c^∞, die bei völliger Disuniformität fair ist), so ist der Bestätigungsgrad der Allbehauptung im unendlichen Bereich 0; in allen interessanten Fällen können wir den Gegenstandsbereich jedoch als praktisch unendlich ansehen. Nimmt man 1 als Uniformitätsgrad und benützt daher c^0,

so erhält man damit das paradoxe Ergebnis, daß ein und dasselbe Erfahrungsdatum zwei sich teilweise widersprechende generelle Hypothesen jeweils im maximalen Grad 1 bestätigen können; so ist etwa sowohl

$$c^0 (\Lambda \xi [\Phi (\xi) \rightarrow \Psi (\xi)], \neg \Phi (\alpha_1) \wedge (\neg) \Psi (\alpha_1) \wedge \ldots \wedge \neg \Phi (\alpha_k) \wedge (\neg) \Psi (\alpha_k)) = 1$$

als auch

$$c^0 (\Lambda \xi [\Phi (\xi) \rightarrow \neg \Psi (\xi)], \neg \Phi (\alpha_1) \wedge (\neg) \Psi (\alpha_1) \wedge \ldots \wedge \neg \Phi (\alpha_k) \wedge (\neg) \Psi (\alpha_k)) = 1,$$

und es ist sowohl

$$c^0 (\Lambda \xi [\Phi (\xi) \vee \Psi (\xi)], \Phi (\alpha_1) \wedge (\neg) \Psi (\alpha_1) \wedge \ldots \wedge \Phi (\alpha_k) \wedge (\neg) \Psi (\alpha_k)) = 1$$

als auch

$$c^0 (\Lambda \xi [\Phi (\xi) \vee \neg \Psi (\xi)], \Phi (\alpha_1) \wedge (\neg) \Psi (\alpha_1) \wedge \ldots \wedge \Phi (\alpha_k) \wedge (\neg) \Psi (\alpha_k)) = 1.$$

Die Ptolemäischen und die Keplerschen Gesetze hatten etwa zur Zeit der Formulierung der letzteren gleich viele bestätigende und widerlegende Instanzen, trotzdem einigte sich die Fachwelt relativ rasch darauf, Keplers Theorie als die bessere anzusehen. Durch c^0 wären also beide Theorien zur Zeit der Formulierung der Keplerschen Gesetze entweder im Grad 1 oder im Grad 0 bestätigt.

Bei der Beurteilung der Frage, welche naturwissenschaftliche Theorie als die plausibelste anzusehen ist, spielt ohne Zweifel die Frage eine Rolle, in welchem Grad ihre Einzelfälle durch die Erfahrung bestätigt werden; dies ist aber sicherlich nicht das einzige und auch nicht das entscheidende Kriterium für die Annahme oder Verwerfung einer solchen allgemeinen Hypothese. Die Frage, aus welchen Gründen wir allgemeine Theorien akzeptieren oder nicht akzeptieren, wird am Ende des VII. Kapitels nochmals aufgeworfen.

Theorem 20 zeigt im übrigen, daß die Bedingung (6) von D 4 nur für endliche Bereiche gefordert werden darf; hält man an der Limeskonvention fest (und nichts scheint gerechtfertigter als dies), so bekommt man bei unendlichen Bereichen den Wahrscheinlichkeitswert 0, wenn diese für alle endlichen Teilbereiche zwar wegen (6) größer als 0 sind, sich der 0 mit wachsender Anzahl der Gegenstände jedoch immer mehr annähern.

Auch unter dem Gesichtspunkt der Anwendung des Wahrscheinlich-

keitsbegriffs, etwa bei Wetten, erscheint es sinnvoll, daß bei unendlichen Bereichen Allbehauptungen hinsichtlich endlicher Erfahrungsdaten den Wahrscheinlichkeitswert 0 erhalten. Denn im anderen Fall müßte ein rational Wettender genötigt sein, mit einem von 0 verschiedenen Einsatz auf eine derartige unendliche Konjunktion zu wetten. Dann können folgende Fälle eintreten: Der Allsatz erweist sich als falsch, da ein Gegenbeispiel beobachtet wird; er hat dann die Wette verloren und muß bezahlen. Oder das generelle Urteil erweist sich bei allen bisherigen Experimenten nicht als falsch; dann hat er die Wette noch nicht gewonnen, denn es kann sich ja bei künftigen Untersuchungen als irrig erweisen. Eine Wahrscheinlichkeitsfunktion, die hier nicht den Wert 0 liefert, verleitet ihn also dazu, ein Wettsystem einzugehen, das ihm zwar einen Verlust, nicht jedoch einen Gewinn ermöglicht; sie ist im Sinne der Definition nicht streng kohärent.

Schließlich sei noch bemerkt, daß man die Behauptung, der Gegenstandsbereich sei (abzählbar) unendlich, in zweierlei Weise verstehen kann: Man kann annehmen, daß er aktual unendlich ist, also als fertige Gesamtheit vorliegt, und man kann voraussetzen, daß dem nicht so ist, daß er vielmehr eine Folge von immer umfangreicheren endlichen Klassen darstellt.[16]

Vom Standpunkt der induktiven Logik aus ist dazu folgendes zu bemerken: Entweder die eine Auffassung erlaubt mehr Definitionen und Lehrsätze als die andere; dann besteht zwischen beiden ein erkennbarer Unterschied, und man hat nach Gründen zu suchen, um sich für die eine oder die andere zu entscheiden. Oder beide Auffassungen führen zu den gleichen Ergebnissen, wobei lediglich die eine mit komplizierteren Formulierungen verbunden ist als die andere; dann liegt, zumindest vom Standpunkt der induktiven Logik aus, kein echtes Problem, sondern nur ein Wortstreit oder ein Scheinproblem vor.

---

[16] Nach Hao Wangs Kriterium ist der eine Standpunkt für den Platonismus, der andere für den Nominalismus charakteristisch, vgl. Hao Wang, What is an Individual?, in: The Philosophical Review 62 (1953) S. 413—420.

VI. Induktive Logik und Wahrscheinlichkeitslehre

1. Logische und objektive Wahrscheinlichkeit

Der Schluß von den relativen Häufigkeiten eines Gegenstandsbereichs auf die einer (durch Zufall ausgewählten) Stichprobe dieser Gesamtheit — von Carnap „direkter Schluß" genannt — nimmt in den Lehrbüchern der Wahrscheinlichkeitslehre eine zentrale Stellung ein. Er wird im folgenden für induktive Methoden entwickelt.[1] Dabei beschreibt die Prämisse die relativen Häufigkeiten von elementfremden Eigenschaften (Klassen) im Gegenstandsbereich und die Hypothese deren relative Häufigkeiten in der Stichprobe.

Normalerweise wird bei den Lehrsätzen der Gegenstandsbereich B endlich sein. Für abzählbar unendliche Bereiche ergeben sich die Wahrscheinlichkeitswerte dann nach einer noch zu formulierenden Limeskonvention für Gegenstandsausdrücke. Für endliche Bereiche braucht c nur die Bedingungen (I) und (II) (a) zu erfüllen; für abzählbar unendliche Bereiche muß c noch dieser Limeskonvention genügen, die nicht die Gültigkeit von (II) (b) und (III) voraussetzt. Jene Lehrsätze, die nur für unendliche Gegenstandsbereiche formuliert werden, gelten approximativ auch für endliche Bereiche, und zwar ist die Approximation um so besser, je größer die Anzahl n der Gegenstände im Vergleich zu den anderen Anzahlen ist.

Da in diesem Abschnitt von c nur vorausgesetzt wird, daß es den Bedingungen (I) und (II) (a) genügt, kann die Limeskonvention für Gegenstandsausdrücke, so wie sie in D 24 eingeführt worden ist, nicht übernommen werden; reguläre Bestätigungsmethoden, die hinsichtlich der Gegenstandsausdrücke symmetrisch sind, können ja auch außerhalb des Kontinuums liegen. Da im folgenden jedoch nur diese beiden Adäquatheitsbedingungen verwendet und gelegentlich auch abzählbar unendliche Bereiche in Betracht gezogen werden, ist

---

[1] Ich übernehme in diesem Abschnitt die für die vorliegende Untersuchung wichtigsten Ergebnisse aus Carnap [9], S. 483—503.

es notwendig, solche Funktionen c auch für unendliche Sprachen zu definieren.

Nach Korollar 5 ist die Gesamtheit der Funktionen $_1c^1$, $_2c^1$, $_3c^1$, ... für die immer umfangreicheren endlichen Teilbereiche B_1, B_2, B_3, ... von B in dem Sinn eine passende Folge von c-Funktionen, daß $_{n+1}c^1$ den Sätzen über B_n den gleichen Wert zuordnet wie $_nc^1$. Wenn nicht alle Adäquatheitsbedingungen benützt werden, kann c jedoch nicht durch einen Parameter L charakterisiert werden. Deshalb muß die Bedingung, daß eine passende Folge von c-Funktionen existiert, eigens in die erweiterte Limeskonvention aufgenommen werden. Es wird also gefordert, daß Sätze Ψ und Φ, die bezüglich jener Sprache, die an Gegenstandsausdrücken nur die darin vorkommenden n Ausdrücke enthält, den c-Wert $_nc$ (Ψ, Φ) erhalten, auch bezüglich aller Sprachen mit m zusätzlichen Gegenstandsausdrücken den gleichen Wert $_{n+m}c$ (Ψ, Φ) erhalten, daß also Gegenstandsausdrücke, die nicht in Ψ oder Φ vorkommen, auf den c-Wert keinen Einfluß haben.

Wegen D 3 und D 2 kann diese Forderung auf die Aprioriwerte von c hinsichtlich jeder Zustandsbeschreibung $_n\Sigma$ jener Sprache beschränkt werden, die an Gegenstandsausdrücken nur α_1, ..., α_n enthält; in der Sprache, die zusätzlich noch den Gegenstandsausdruck α_{n+1} enthält, ist $_n\Sigma$ natürlich keine Zustandsbeschreibung mehr, da sie nicht alle Gegenstände des neuen Bereichs vollständig charakterisiert, aber sie ist identisch mit einer Adjunktion von Zustandsbeschreibungen dieser neuen Sprache. Es wird daher definiert:

D 30 Der Gegenstandsbereich B und die Gesamtheit der Gegenstandsausdrücke der Sprache seien beide abzählbar unendlich; $_nc$ sei eine reguläre Bestätigungsmethode für jene endliche Teilsprache, die nur die ersten n Gegenstandsausdrücke enthält (und zwar gelte das für alle natürlichen Zahlen n, die größer als 0 sind).

Dann erfüllt $_\infty c$ die Limeskonvention für Gegenstandsausdrücke hinsichtlich einer passenden Folge von c-Funktionen genau dann, wenn

(1) für jede endliche Anzahl n von Gegenstandsausdrücken (die alle und nur die Ausdrücke α_1, ..., α_n enthält) gilt, daß

nach Hinzufügung einer beliebigen endlichen Anzahl m der nächsten Gegenstandsausdrücke sich der c-Wert für Zustandsbeschreibungen $_n\Sigma$ aus der Sprache mit n Ausdrücken nicht ändert, daß also $_nc\,(_n\Sigma, T) = {}_{n+m}c\,(_n\Sigma, T)$, und wenn

(2) $_\infty c\,(\Psi, \Phi) = \lim_{n\to\infty} {}_nc\,(\Psi, \Phi)$ für jeden beliebigen Satz Ψ und jeden beliebigen erfüllbaren Satz Φ.

Die Hypothese Ψ der folgenden induktiven Argumente nimmt auf eine Stichprobe Bezug. In den Lehrsätzen wird über die Art, wie man zu diesen gelangt, nichts ausgesagt. Da ein induktives Argument *bei seiner Anwendung* der Forderung des Gesamtdatums zu genügen hat und in den Prämissen nicht formuliert wird, welche Teilklasse der Gesamtheit zu wählen ist, ist die Auswahl so vorzunehmen, daß man über sie tatsächlich nichts Näheres weiß, daß sie also regellos (etwa durch Losentscheid) zustande gekommen ist.

Daß eine Stichprobe eine Zufallsauswahl ist, besagt dasselbe, wie daß bei unendlich oftmaliger Wiederholung jenes Auswahlverfahrens alle Gegenstände mit dem gleichen Grenzwert der relativen Häufigkeit ausgewählt werden. Daraus folgt, daß wir niemals *wissen* können, ob eine Stichprobe eine Zufallsauswahl ist, da uns Grenzwerte der relativen Häufigkeiten nicht gegeben sind; wir können diese nur induktiv *schätzen* und können olglich nur sagen, daß eine gewisse *logische Wahrscheinlichkeit* dafür besteht, daß eine bestimmte Auswahl tatsächlich durch Zufall zustandegekommen ist.[2]

Den folgenden Überlegungen wird stets eine Gesamtheit $\Pi'_1, \ldots, \Pi'_d$ von Prädikaten zugrundegelegt, die eine Division bilden, die also Eigenschaften (Klassen) $F'_1, \ldots, F'_d$ bezeichnen, die wechselseitig elementfremd sind und zusammen den Bereich B ausmachen. Die q V-Prädikate $\Pi^*_1, \ldots, \Pi^*_q$ bilden etwa eine derartige Familie. Dann gilt der folgende Hilfssatz:

---

[2] Vgl. Keynes [1], S. 290ff., und Carnap [9], S. 493ff., sowie Carnap-Stegmüller [1], S. 185f.

Lemma 6

Der Gegenstandsbereich B sei endlich oder abzählbar unendlich. Gegeben sei die Division $\Pi'_1, \ldots, \Pi'_d$. Ψ sei eine individuelle Verteilung für endlich viele Gegenstandsausdrücke $\beta_1, \ldots, \beta_m$ bezüglich dieser Division $\Pi'_1, \ldots, \Pi'_d$, und Δ sei die entsprechende statistische Verteilung (die Standardadjunktion aus Ψ und jenen individuellen Verteilungen, die mit Ψ isomorph sind). Ψ wie Δ besagen also für jedes i mit $1 \leqslant i \leqslant d$, daß m_i der m Gegenstände der Stichprobe die Eigenschaft F'_i haben. Φ sei irgendein Satz, der keinen der Gegenstandsausdrücke $\beta_1, \ldots, \beta_m$ enthält. Die Funktion c erfülle die Bedingungen (I) und (II) (a). Dann gilt:

$$c(\Phi \wedge \Delta, T) = \frac{m!}{m_1! \cdot m_2! \cdot \ldots \cdot m_d!} \cdot c(\Phi \wedge \Psi, T).$$

Beweis

Die Voraussetzungen seien gegeben.

Es gibt m! Möglichkeiten, um m Gegenstandsausdrücke in eine lineare Anordnung zu bringen; denn man hat zuerst m Möglichkeiten zur Auswahl des ersten Gegenstandsausdrucks, sodann $m - 1$ Möglichkeiten zur Auswahl des zweiten aus den restlichen $m - 1$ Gegenstandsausdrücken usw., also insgesamt $m \cdot (m - 1) \cdot \ldots \cdot 1$, was per definitionem mit m! identisch ist.

Die Anzahl j jener individuellen Verteilungen, die die Adjunktionsglieder von Δ sind und die F'_1 m_1 Dinge, F'_2 m_2 Dinge, ..., und F'_d m_d Dinge zuschreiben, ist dann $m!/(m_1! \cdot m_2! \cdot \ldots \cdot m_d!)$. Denn wenn man die m_1 Gegenstandsausdrücke, die die Dinge aus F'_1 bezeichnen, alle als verschieden betrachtet, also annimmt, daß die durch sie bezeichneten Dinge neue Eigenschaften $H^1_1, \ldots, H^1_{m_1}$ haben, dann gibt es $m_1!$ Anordnungen dieser Ausdrücke, und man erhält dann statt der Anzahl j von individuellen Verteilungen die Anzahl $j \cdot m_1!$. Wenn man zusätzlich die m_2 Gegenstandsausdrücke, die die Dinge aus F'_2 bezeichnen, als verschieden betrachtet, also annimmt, daß die durch sie bezeichneten m_2 Dinge verschiedene Eigenschaften $H^2_1, \ldots, H^2_{m_2}$ haben, erhält man den neuen Wert $j \cdot m_1! \cdot m_2!$, usw. Betrachtet man also alle Gegenstandsausdrücke als verschieden (so

daß die durch sie bezeichneten Dinge alle verschiedene Eigenschaften haben), so erhält man den Wert $j \cdot m_1! \cdot m_2! \cdot \ldots \cdot m_d!$. Andererseits muß dies aber die Anzahl aller möglichen Anordnungen der m Gegenstandsausdrücke sein (es ist ja $m = m_1 + m_2 + \ldots + m_d$), so daß $m! = j \cdot m_1! \cdot m_2! \cdot \ldots \cdot m_d!$ und damit $j = m!/(m_1! \cdot m_2! \cdot \ldots \cdot m_d!)$. Entsteht Ψ' aus Ψ durch Vertauschung der Gegenstandsausdrücke und Neuordnung der Konjunktionsglieder in lexikographischer Anordnung, so ist nach (3) und (II) (a) $c(\Psi', T) = c(\Psi, T)$, und damit nach Bedingung (4) von D 4, da Ψ und Ψ' nicht simultan erfüllbar sind, $c(\Psi \vee \Psi', T) = c(\Psi, T) + c(\Psi', T) = 2 \cdot c(\Psi, T)$. Wiederholt man dies für die übrigen $j - 2$ individuellen Verteilungen, die aus Ψ durch die angegebene Art hervorgehen, so erhält man $c(\Delta, T) = c(\Psi_1 \vee \ldots \vee \Psi_j, T) = j \cdot c(\Psi, T)$. Da nun $\Phi \wedge \Delta$ logisch äquivalent mit $(\Phi \wedge \Psi_1) \vee \ldots \vee (\Phi \wedge \Psi_j)$ ist und diese Adjunktionsglieder ebenfalls miteinander logisch unverträglich sind und da wegen (II) (a) gilt: $c(\Phi \wedge \Psi_1, T) = \ldots = c(\Phi \wedge \Psi_j, T)$, erhält man schließlich:

$$c(\Phi \wedge \Delta, T) = j \cdot c(\Phi \wedge \Psi, T) = c(\Phi \wedge \Psi, T) \cdot m!/(m_1! \cdot m_2! \cdot \ldots \cdot m_d!).$$

Im nächsten Hilfssatz werden die beiden folgenden mathematischen Definitionen benützt:

$$\binom{r}{s} = \frac{r!}{s! \cdot (r-s)!} \quad \text{für natürliche Zahlen r und s mit } s \leqslant r;$$

$$\begin{bmatrix} r \\ s \end{bmatrix} = \frac{r!}{(r-s)!} \quad \text{für natürliche Zahlen r und s mit } s \leqslant r.$$

Lemma 7

Der Gegenstandsbereich B sei endlich oder abzählbar unendlich. Gegeben sei die Division $\Pi'_1, \ldots, \Pi'_d$. Ψ sei eine individuelle Verteilung für die endlich vielen Gegenstandsausdrücke $\gamma_1, \ldots, \gamma_k$ bezüglich dieser Division und besage für jede natürliche Zahl i mit $1 \leqslant i \leqslant d$, daß k_i der k Gegenstände die Eigenschaft F'_i haben, und Ψ' sei eine individuelle Verteilung für endlich viele andere Gegenstandsausdrücke $\beta_1, \ldots, \beta_m$ bezüglich dieser Division und besage ana-

log, daß m_i der m Dinge die Eigenschaft F'_i haben. Δ sei die Ψ entsprechende statistische Verteilung (und enthalte daher Ψ als Adjunktionsglied), und Δ' sei die Ψ' entsprechende statistische Verteilung. Γ sei jene statistische Verteilung, die als Adjunktionsglied einen bis auf die Anordnung der Konjunktionsglieder mit $\Psi \wedge \Psi'$ identischen Satz enthält. c erfülle die Bedingungen (I) und (II) (a). Dann gilt:

$$\text{(a)} \quad c(\Psi \wedge \Delta', T) = \frac{m!}{m_1! \cdot \ldots \cdot m_d!} \cdot c(\Psi \wedge \Psi', T)$$

$$\text{(b)} \quad c(\Gamma, T) = \frac{(k+m)!}{(k_1+m_1)! \cdot \ldots \cdot (k_d+m_d)!} \cdot c(\Psi \wedge \Psi', T)$$

$$\text{(c)} \quad c(\Delta \wedge \Delta', T) = \frac{k!}{k_1! \cdot \ldots \cdot k_d!} \cdot c(\Psi \wedge \Delta', T)$$

$$= \frac{k!}{k_1! \cdot \ldots \cdot k_d!} \cdot \frac{m!}{m_1! \cdot \ldots \cdot m_d!} \cdot c(\Psi \wedge \Psi', T)$$

$$\text{(d)} \quad c(\Psi, \Gamma) = \frac{m!}{m_1! \cdot \ldots \cdot m_d!} \cdot \frac{(k_1+m_1)! \cdot \ldots \cdot (k_d+m_d)!}{(k+m)!}$$

$$= \frac{\begin{bmatrix} k_1+m_1 \\ k_1 \end{bmatrix} \cdot \ldots \cdot \begin{bmatrix} k_d+m_d \\ k_d \end{bmatrix}}{\begin{bmatrix} k+m \\ k \end{bmatrix}}$$

$$\text{(e)} \quad c(\Delta, \Gamma) = \frac{k!}{k_1! \cdot \ldots \cdot k_d!} \cdot \frac{m!}{m_1! \cdot \ldots \cdot m_d!} \cdot \frac{(k_1+m_1)! \cdot \ldots \cdot (k_d+m_d)!}{(k+m)!}$$

$$= \frac{\binom{k_1+m_1}{k_1} \cdot \ldots \cdot \binom{k_d+m_d}{k_d}}{\binom{k+m}{k}}$$

$$= \frac{k!}{k_1! \cdot \ldots \cdot k_d!} \cdot c(\Psi, \Gamma) .$$

Beweis

Die Voraussetzungen seien gegeben.
(a) Δ' hat $m!/(m_1! \cdot \ldots \cdot m_d!)$ Adjunktionsglieder. Ferner folgt aus Δ und Δ' logisch Γ, da $\Delta \wedge \Delta'$ für jedes i mit $1 \leqslant i \leqslant d$ aussagt, daß $k_i + m_i$ der $k + m$ Gegenstände die Eigenschaft F'_i haben, was genau der Gehalt von Γ ist. Ähnlich kann auf Grund der Aussagen Γ und Δ der Gehalt von Δ' berechnet werden, so daß aus Γ und Δ logisch Δ' folgt. Also ist $\Gamma \wedge \Delta$ logisch äquivalent mit $\Delta \wedge \Delta'$. Daraus ergibt sich, daß auch $\Gamma \wedge \Psi$ logisch äquivalent ist mit $\Psi \wedge \Delta'$ (denn Ψ impliziert logisch Δ, dessen Adjunktionsglied es ist, also impliziert $\Psi \wedge \Delta'$ logisch $\Delta \wedge \Delta' \wedge \Psi$ und damit auch $\Gamma \wedge \Psi$; umgekehrt folgt aus $\Gamma \wedge \Psi$ logisch $\Gamma \wedge \Delta$ und daraus Δ', weshalb man schließlich $\Psi \wedge \Delta'$ erhält). Mit Bedingung (3) von D 4 erhält man daher: $c(\Gamma \wedge \Psi, T) = c(\Psi \wedge \Delta', T)$. Da Ψ und Δ' keinen gemeinsamen Gegenstandsausdruck haben, folgt mit Lemma 6:

$$c(\Psi \wedge \Delta', T) = \frac{m!}{m_1! \cdot \ldots \cdot m_d!} \cdot c(\Psi \wedge \Psi', T).$$

(b) Da Γ $(k + m)!/[(k_1 + m_1)! \cdot \ldots \cdot (k_d + m_d)!]$ Adjunktionsglieder hat, ergibt sich diese Teilbehauptung unmittelbar aus Lemma 6.
(c) Da Δ $k!/(k_1! \ldots k_d!)$ Adjunktionsglieder hat und kein Gegenstandsausdruck sowohl in Ψ als auch in Δ' vorkommt, ergibt sich die erste Gleichung dieser Teilbehauptung aus Lemma 6; die zweite folgt daraus mit (a).
(d) Diese Teilbehauptung folgt unmittelbar aus (a) und (b).
(e) Da $\Gamma \wedge \Delta$ logisch äquivalent mit $\Delta \wedge \Delta'$ ist, gilt mit Bedingung (3) von D 4: $c(\Gamma \wedge \Delta, T) = c(\Delta \wedge \Delta', T)$. Individuelle und damit auch statistische Verteilungen sind erfüllbar, so daß mit Theorem 2 (b): $c(\Gamma, T) > 0$. Mit D 3 ist daher $c(\Delta, \Gamma) = c(\Delta \wedge \Delta', T)/c(\Gamma, T)$. Daraus ergeben sich die beiden ersten Gleichungen mit (c) und (b) und hieraus wiederum die dritte mit (d).

Die nächsten Lehrsätze zeigen Beispiele für direkte Schlüsse. Die Grundgesamtheit, auf die dabei Bezug genommen wird, ist ein echter oder unechter Teil des Gegenstandsbereichs B. Die Stichprobe wird (in der Anwendung) durch ein Zufallsverfahren aus der Grund-

gesamtheit ausgewählt, da wegen der Forderung des Gesamtdatums vorausgesetzt werden muß, daß die Aussage, ein Gegenstand der Stichprobe sei durch ein solches Verfahren ausgewählt, für die Hypothese irrelevant ist. Wenn der Gegenstandsbereich etwa die Klasse der Menschen und die Grundgesamtheit die Klasse der wahlberechtigten Bürger der Bundesrepublik Deutschland ist, wobei die Prädikate die Eigenschaften bezeichnen, bei der letzten Bundestagswahl die eine oder die andere der bestehenden Parteien oder keine von ihnen gewählt zu haben, so ist es nicht ratsam, als Stichprobe die Gesamtheit jener Westdeutschen zu nehmen, die am Wahlsonntag in der Kirche waren, da es möglicherweise Erfahrungen über das Wahlverhalten dieser Kirchgänger gibt, die den Wahrscheinlichkeitswert beeinflussen.
Die folgenden Lehrsätze zeigen, wie man die Werte $c(\Psi', \Gamma)$ und $c(\Delta, \Gamma)$ berechnen kann, wenn Γ eine statistische Verteilung über der Grundgesamtheit, Δ eine statistische Verteilung über der Stichprobe und Ψ' eine individuelle Verteilung über der Stichprobe ist.

Theorem 21

Der Gegenstandsbereich B sei endlich oder unendlich. Die endlich vielen Ausdrücke $\beta_1, \ldots, \beta_m$ bezeichnen Elemente aus B. Gegeben sei die Division $\Pi'_1, \ldots, \Pi'_d$. Γ sei eine statistische Verteilung für die $\beta_1, \ldots, \beta_m$ bezüglich dieser Division und besage für jedes i mit $1 \leqslant i \leqslant d$, daß m_i der m Gegenstände die durch Π'_i bezeichnete Eigenschaft F'_i haben. Es sei Ψ' eine individuelle Verteilung für k der m betrachteten Gegenstände (so daß $k \leqslant m$ und damit auch $k_i \leqslant m_i$ für jedes i mit $1 \leqslant i \leqslant d$). Δ sei die statistische Verteilung, die Ψ' als Adjunktionsglied enthält. c erfülle die Bedingungen (I) und (II) (a). Dann gilt:

$$(a)\quad c(\Psi', \Gamma) = \frac{(m-k)!}{(m_1-k_1)! \cdot \ldots \cdot (m_d-k_d)!} \cdot \frac{m_1! \cdot \ldots \cdot m_d!}{m!}$$

$$= \frac{\begin{bmatrix} m_1 \\ k_1 \end{bmatrix} \cdot \ldots \cdot \begin{bmatrix} m_d \\ k_d \end{bmatrix}}{\begin{bmatrix} m \\ k \end{bmatrix}}$$

(b) $$c(\Delta, \Gamma) = \frac{k!}{k_1! \cdot \ldots \cdot k_d!} \cdot c(\Psi, \Gamma)$$
$$= \frac{\binom{m_1}{k_1} \cdot \ldots \cdot \binom{m_d}{k_d}}{\binom{m}{k}}.$$

Beweis

Die Voraussetzungen seien gegeben.
(a) Man erhält diese Teilbehauptung aus Lemma 7 (d), wenn man den dortigen Satz Ψ auf die Stichprobe und Ψ' auf den restlichen Teil der Grundgesamtheit anwendet und also die Ziffer „m" durch „$m - k$" und „$m + k$" durch „m" ersetzt (und analog für die Ziffern „m_i" und „$m_i + k_i$").
(b) Durch die gleichen Ersetzungen erhält man diese Teilbehauptung aus Lemma 7 (e).

Wenn man berücksichtigt, daß stets $\begin{bmatrix} m \\ 0 \end{bmatrix} = 1$ und $\begin{bmatrix} m \\ 1 \end{bmatrix} = m$, ergibt sich aus Theorem 21 (a) unmittelbar der wichtige Folgesatz:

Korollar 11

Gegeben seien die Voraussetzungen des Theorems 21. Zusätzlich sei $d = 2$, so daß Π'_1 und Π'_2 eine Eigenschaft (Klasse) und deren Komplement bezeichnen. Es sei $k = 1$, d. h. die Stichprobe enthalte nur ein Element, so daß Ψ sowohl als individuelle als auch als statistische Verteilung aufgefaßt werden kann.
(a) Ist $k_1 = 1$ und $k_2 = 0$, so gilt: $c(\Psi, \Gamma) = m_1/m$.
(b) Ist $k_1 = 0$ und $k_2 = 1$, so gilt: $c(\Psi, \Gamma) = m_2/m$.

Dieses Korollar stellt einen wichtigen Zusammenhang zwischen logischer und objektiver Wahrscheinlichkeit her. Da die Grundgesamtheit endlich ist, ist die objektive Wahrscheinlichkeit der Eigenschaft F'_1 in dieser Grundgesamtheit identisch mit der relativen Häufigkeit

von F'_1 in ihr. Bezüglich dieses Datums ist dann die *logische Wahrscheinlichkeit* dafür, daß ein in der Grundgesamtheit enthaltener Gegenstand die Eigenschaft F'_1 hat, identisch mit der beobachteten *objektiven Wahrscheinlichkeit*. Wenn also 0,39 die objektive Wahrscheinlichkeit der SPD-Wähler (bei der letzten Bundestag wahl) in der Grundgesamtheit der wahlberechtigten Bundesbürger ist, so ist die Wahrscheinlichkeit, daß ein beliebig vorgegebener Bundesbürger (bei der letzten Bundestagswahl) SPD gewählt hat, hinsichtlich dieses Datums 0,39.

Wählt man die Gesamtheit der Gegenstände $b_1, b_2, \ldots$ und damit die Gesamtheit der Gegenstandsausdrücke $\beta_1, \beta_2, \ldots$, auf die die statistische Verteilung Γ Bezug nimmt, als unendliche Klasse, so daß Γ also ein Satz von unendlicher Länge ist, dann kann das Korollar 11 für reguläre Bestätigungsmethoden, die die Bedingung (II) (a) und die die Limeskonvention für Gegenstandsausdrücke hinsichtlich einer passenden Folge von c-Funktionen erfüllen, bewiesen werden, vorausgesetzt, daß die relativen Häufigkeiten, die Γ mit wachsendem m erwähnt, gegen einen Limes u_i konvergieren. Es gilt also auch dann, wenn die objektive Wahrscheinlichkeit mit dem Limes von relativen Häufigkeiten identisch ist, da die *logische Wahrscheinlichkeit* dafür, daß ein in der (unendlichen) Grundgesamtheit enthaltener Gegenstand die Eigenschaft F'_i hat, bezüglich dieser *objektiven Wahrscheinlichkeit* den gleichen Wert hat wie dieser selbst.

Allerdings bleibt die Schwierigkeit nach wie vor bestehen, ob man berechtigt ist, jemals eine solche Prämisse Γ zu verwenden, die über unendlich viele Einzelerfahrungen berichtet. Da uns nur endliche Erfahrungen gegeben sind, können wir eine derartige Prämisse höchstens mittels eines logischen Wahrscheinlichkeitsbegriffs schätzen. Der Schätzwert der objektiven Wahrscheinlichkeit wird jedoch von dieser selbst (falls sie überhaupt existiert), im allgemeinen zumindest geringfügig abweichen, also normalerweise von dieser verschieden sein. Der Übergang vom Schätzwert zum tatsächlichen Wert (die Identifizierung beider) ist demnach ein Argument, das weder durch eine deduktive noch durch eine induktive Logik gerechtfertigt werden kann (wiewohl es sich in sehr vielen praktischen Fällen als unproblematisch erweist, da der Fehler, den man dabei macht, so klein ist, daß er vernachlässigt werden kann).

Bevor der Begriff der Schätzung der relativen Häufigkeit eingeführt wird, seien noch zwei weitere Verfahren zur Berechnung des Wahrscheinlichkeitswertes für den direkten Schluß genannt. Sie gelten approximativ, wenn nur endlich viele Gegenstandsausdrücke $\beta_1, \ldots, \beta_m$ betrachtet werden, wobei die Approximation mit wachsendem m immer größer wird (m muß groß im Hinblick auf k^2 sein), und sie gelten ohne Einschränkung, wenn abzählbar unendlich viele Gegenstandsausdrücke $\beta_1, \beta_2, \ldots$ in Erwägung gezogen werden (hier muß noch zusätzlich vorausgesetzt werden, daß die relativen Häufigkeiten gegen einen Limes konvergieren).

Theorem 22

Der Gegenstandsbereich B sei abzählbar unendlich. Die unendlich vielen Ausdrücke $\beta_1, \beta_2, \ldots$ bezeichnen Elemente aus B. Gegeben sei die Division $\Pi'_1, \ldots, \Pi'_d$. Γ sei eine statistische Verteilung für die $\beta_1, \beta_2, \ldots$ bezüglich dieser Division und besage, daß für jedes i mit $1 \leqslant i \leqslant d$ mit wachsendem m die relative Häufigkeit m_i/m der Gegenstände, die die durch Π'_i bezeichnete Eigenschaft F'_i haben, gegen den Wert u_i konvergiert. Es sei Ψ eine individuelle Verteilung für k Gegenstände aus der Grundgesamtheit, wobei k (und somit auch jedes der k_i) endlich ist. Δ sei die statistische Verteilung, die Ψ als Adjunktionsglied enthält. c erfülle die Bedingungen (I), (II) (a) und die Limeskonvention für Gegenstandsausdrücke hinsichtlich einer passenden Folge von c-Funktionen. Dann gilt:

(a) $c(\Psi, \Gamma) = u_1^{k_1} \cdot \ldots \cdot u_d^{k_d}$;

(b) $c(\Delta, \Gamma) = \dfrac{k!}{k_1! \cdot \ldots \cdot k_d!} \cdot u_1^{k_1} \cdot \ldots \cdot u_d^{k_d}$.

Beweis

Die Voraussetzungen seien erfüllt.

(a) Diese Teilbehauptung ergibt sich aus Theorem 21 (a) unter Berücksichtigung der approximativen Gleichung $\begin{bmatrix} r \\ s \end{bmatrix} \cong r^s$ (falls $r \gg s^2$), die man mit Hilfe der Stirlingschen Näherungsformel für s! beweisen kann.

(b) Dieser Teilsatz folgt aus (a) und Theorem 21 (b).

Korollar 12

Gegeben seien die Voraussetzungen von Theorem 22. Es sei $d = 2$. Dann gilt:

(a) $c(\Psi, \Gamma) = u_1^{k_1} \cdot (1 - u_1)^{k-k_1}$

(b) $c(\Delta, \Gamma) = \binom{k}{k_1} \cdot u_1^{k_1} \cdot (1 - u_1)^{k-k_1}$

Korollar 12 (b) wird üblicherweise wegen der Analogie zum binomischen Lehrsatz der Algebra „binomisches Gesetz" genannt. Da Δ $\binom{k}{k_1}$ Adjunktionsglieder hat, worunter auch Ψ ist, wird der Wert $u_1^{k_1} \cdot (1 - u_1)^{k-k_1}$ mit diesem Faktor multipliziert.

Vergleicht man die Folgesätze 12 (a) und 11, so stellt man folgendes fest: Wenn $\Psi_1, \Psi_2, \ldots, \Psi_k$ die k Konjunktionsglieder sind, aus denen Ψ besteht, so ist $c(\Psi, \Gamma) = c(\Psi_1, \Gamma) \cdot c(\Psi_2, \Gamma) \cdot \ldots \cdot c(\Psi_k, \Gamma)$. Andererseits gilt mit k—1-maliger Anwendung von Bedingung (5) von D 4:

$$c(\Psi, \Gamma) = c(\Psi_1 \wedge \ldots \wedge \Psi_k, \Gamma)$$
$$= c(\Psi_1, \Gamma) \cdot c(\Psi_2, \Gamma \wedge \Psi_1) \cdot \ldots \cdot c(\Psi_k, \Gamma \wedge \Psi_1 \wedge \ldots \wedge \Psi_{k-1}).$$

Also gilt für jedes i mit $1 < i \leqslant k$: $c(\Psi_i, \Gamma \wedge \Psi_1 \wedge \ldots \wedge \Psi_{i-1}) = c(\Psi, \Gamma)$, d. h. $\Psi_1 \wedge \ldots \wedge \Psi_{i-1}$ ist für Ψ_i bezüglich Γ *irrelevant*.

Dies ist jedoch nur auf den ersten Blick überraschend. Denn c erfüllt ja die Bedingung (II) (a) und ist deshalb, wie schon durch Korollar 8 gezeigt worden ist, nur von den beobachteten relativen Häufigkeiten abhängig; c berücksichtigt ja nicht die Tatsache, daß *ganz gewisse Gegenstände bestimmte Eigenschaften* besitzen. Die Zusatzdaten $\Psi_1 \wedge \ldots \wedge \Psi_{i-1}$ sind jedoch *statistisch* bereits in Γ berücksichtigt und sagen deshalb *induktiv* nichts Neues aus.

Gewisse Schwierigkeiten bestehen bei der *Anwendung* von Korollar 12 und entsprechend von Theorem 22. Es wird ja sozusagen dabei vorausgesetzt, daß sich die Grenzwerte der relativen Häufigkeiten nicht ändern, wenn man k Gegenstände auf ihre Eigenschaften hin untersucht. Im Falle, daß die Grundgesamtheit unendlich ist, liegt es auf der Hand, daß die Entnahme einer endlichen Anzahl k von Dingen tatsächlich nichts an den Grenzwerten ändert. Im Falle des Münzwurfs oder des Würfelwurfs können wir zur Not von der Fiktion ausgehen, daß die Gegenstandsbereiche (die Gesamtheit der entspre-

chenden Würfe) unendlich sind. Beim Urnenbeispiel tun wir uns da schon schwerer. Man kann hier den unendlichen Bereich dadurch simulieren, daß man entweder m hinreichend groß gegenüber k (und sogar gegenüber k^2) wählt, so daß der damit gemachte Fehler vernachlässigt werden kann, oder daß man nach jedem Zug eine Kugel gleicher Farbe der Urne hinzufügt und die Kugeln der Urne neu mischt, oder daß man nach jedem Zug die Kugel, nachdem man sie auf ihre Farbe hin untersucht hat, wieder in die Urne zurücklegt und ebenfalls die Kugeln der Urne neu mischt.

Reguläre Bestätigungsmethoden c, die symmetrisch bezüglich der Gegenstandsausdrücke sind, haben im Zusammenhang mit dem direkten Schluß außerdem noch folgende wichtige Eigenschaften: c hat für die Hypothese, daß von den k Gegenständen der Stichprobe k_1 die Eigenschaft F'_1 haben, dann ein Maximum, wenn die relative Häufigkeit von F'_1 in der Stichprobe entweder gleich der relativen Häufigkeit von F'_1 in der Grundgesamtheit ist oder dieser letzteren doch soweit wie möglich nahekommt, falls die Grundgesamtheit endlich ist, und wenn die relative Häufigkeit von F'_1 in der Stichprobe gleich dem Limes der relativen Häufigkeiten von F'_1 in der Grundgesamtheit ist, falls diese abzählbar unendlich ist.[3] Wenn ferner ein Datum besagt, daß u_1 die relative Häufigkeit von F'_1 oder deren Limes in einer Grundgesamtheit ist, dann kann der c-Wert für die Hypothese, daß die relative Häufigkeit von F'_1 in der Stichprobe in einem beliebig kleinen Intervall $[u_1 - r, u_1 + r]$ liegt, beliebig der 1 angenähert werden, wenn man nur die Stichprobe genügend groß wählt (die Grundgesamtheit muß dabei hinreichend groß oder abzählbar unendlich sein).[4]

Die verschiedenen Formen des direkten Induktionsschlusses besitzen folgende Eigentümlichkeiten:

(1) Alle Gegenstände, über die die Hypothese etwas aussagt, werden bereits in der Prämisse erwähnt, d. h. die Stichprobe ist immer eine Teilklasse der Grundgesamtheit; daher können für diese Schlüsse c-Werte angegeben werden, ohne daß Annahmen über die Uniformität des Gegenstandsbereichs gemacht werden müssen.

---

[3] Vgl. Carnap [9], S. 495f.

[4] Dies ist das sogenannte Grenzwerttheorem von Bernoulli; zu diesen und verwandten Lehrsätzen vgl. Carnap [9], S. 504—510.

(2) Die c-Werte sind auch hier schon unabhängig von der Anzahl der Gegenstände des zugrundegelegten Universums (vorausgesetzt, daß diese nicht zu klein ist); es können ja auch Funktionen charakterisiert werden, die den Bedingungen (I) und (II) (a) sowie der Limeskonvention genügen und die für andere induktive Argumente von der Anzahl der Gegenstände des Universums abhängig sind.
(3) Es spielt bei diesen Schlüssen auch keine Rolle, wie groß die relativen Weiten der Prädikate sind und ob c linear von k_j abhängt oder nicht; die Prämisse sagt alles Wesentliche über die entsprechenden Eigenschaften aus, insbesondere die *faktischen* relativen Weiten der Prädikate, d. h. die relativen Häufigkeiten der Eigenschaften in der Grundgesamtheit, weshalb die a priori angenommenen *logischen* relativen Weiten irrelevant sind.
Oft interessieren uns gewisse relative Häufigkeiten, ohne daß wir sie kennen; in solchen Fällen müssen wir Schätzwerte dieser Häufigkeiten bilden. Im gewissen Sinn stellt der Wahrscheinlichkeitswert, so wie er durch Korollar 7 geliefert wird, ebenfalls einen Schätzwert dar: Es ist der Mittelwert aus den Wahrscheinlichkeiten, die sich einerseits durch rein aposteriorische Überlegungen und andererseits durch rein apriorische Betrachtungen ergeben. Es wäre dabei jedoch falsch, das arithmetische Mittel dieser beiden Wahrscheinlichkeiten, also $[(k_j/k) + (m/q)]/2$, zu bilden, da nur in wenigen Fällen der rein apriorischen Wahrscheinlichkeit und der rein aposteriorischen Wahrscheinlichkeit das gleiche Gewicht zukommen wird. Normalerweise wird man dem einen oder dem anderen Faktor ein größeres Gewicht beimessen; dann ist nicht durch 2 $(= 1 + 1)$, sondern durch die Summe der beiden Gewichte zu dividieren, also in diesem Fall durch $k + L$, und es ist die rein aposteriorische Wahrscheinlichkeit nicht mit 1, sondern mit k und entsprechend die rein apriorische Wahrscheinlichkeit ebenfalls nicht mit 1, sondern mit dem Gewicht L zu bewerten.
Im 3. Abschnitt des III. Kapitels ist der Gewinn aus einem System von Wetten geschätzt worden. Als Gewichte sind dort den einzelnen möglichen Gewinnen die Aprioriwahrscheinlichkeiten beigemessen worden, die jene Ausgänge haben, die zu diesen Gewinnen führen. Dividiert worden ist durch die Summe dieser Gewichte, also durch die Summe der Aprioriwahrscheinlichkeiten der verschiedenen Aus-

gänge; da diese jedoch logisch erschöpfend sind und sich wechselseitig logisch ausschließen, ist jene Summe mit 1 identisch.
Ähnlich soll nun vorgegangen werden, wenn die relative Häufigkeit einer Eigenschaft in einer Bezugsklasse im Hinblick auf ein Erfahrungsdatum geschätzt wird. Es wird die Summe der gewogenen relativen Häufigkeiten gebildet und durch die Summe der Gewichte dividiert. Da dabei *alle* relativen Häufigkeiten in Betracht gezogen werden und diese sich wechselseitig ausschließen und da als Gewichte zweckmäßigerweise die Wahrscheinlichkeitswerte der jeweiligen relativen Häufigkeit hinsichtlich eines Erfahrungsdatums (gemessen durch eine reguläre Bestätigungsmethode c) benützt werden, ist der Nenner wiederum 1 und kann vernachlässigt werden.
Es bleibt noch zu erwähnen, wie relative Häufigkeiten in der Objektsprache auszudrücken sind. Gegeben sei eine endliche Bezugsklasse G, die etwa 3 Elemente enthält. β_1, β_2 und β_3 seien deren Namen (bei der vorgegebenen Interpretation der Sprache), während Π die Eigenschaft F bezeichne. Der Satz $\neg (\beta_1 \varepsilon \Pi) \wedge \neg (\beta_2 \varepsilon \Pi) \wedge \neg (\beta_3 \varepsilon \Pi)$ besagt dann, daß 0 die relative Häufigkeit von F in G ist, und die Behauptung $\beta_1 \varepsilon \Pi \wedge \beta_2 \varepsilon \Pi \wedge \beta_3 \varepsilon \Pi$ sagt aus, daß 1 die relative Häufigkeit ist. Der Satz $\neg (\beta_1 \varepsilon \Pi) \wedge \beta_2 \varepsilon \Pi \wedge \beta_3 \varepsilon \Pi$ besagt zwar genauso, daß $^2/_3$ die relative Häufigkeit von F in G ist, er sagt jedoch weit mehr aus, da er noch zusätzlich angibt, *welche* Gegenstände F sind und welche nicht. Es muß deshalb von dieser individuellen Verteilung für β_1, β_2 und β_3 bezüglich Π zur entsprechenden statistischen Verteilung übergegangen werden, d. h. es muß der Satz
$[\neg (\beta_1 \varepsilon \Pi) \wedge \beta_2 \varepsilon \Pi \wedge \beta_3 \varepsilon \Pi] \vee [\beta_1 \varepsilon \Pi \wedge \neg (\beta_2 \varepsilon \Pi) \wedge \beta_3 \varepsilon \Pi] \vee$
$[\beta_1 \varepsilon \Pi \wedge \beta_2 \varepsilon \Pi \wedge \neg (\beta_3 \varepsilon \Pi)]$
gebildet werden. Dieser sagt tatsächlich nur aus, daß $^2/_3$ der Dinge aus G die Eigenschaft F haben. Negiert man in ihm alle bisher unnegierten Teilsätze und streicht die Negationszeichen aller bisher negierten Teilsätze, dann formuliert die so entstehende Behauptung, daß $^1/_3$ die relative Häufigkeit von F in G ist.
Da die Bezugsklasse G drei Gegenstände enthält, sind überhaupt nur vier relative Häufigkeiten möglich, nämlich 0, $^1/_3$, $^2/_3$ und 1. Analog kann man sich leicht klarmachen, daß für jede endliche und von der 0 verschiedene Anzahl m gilt: Wenn G m Dinge enthält, so sind $m + 1$ relative Häufigkeiten möglich, nämlich 0, 1/m, 2/m, ..., (m—1)/m

und 1. Jede rationale Zahl zwischen 0 und 1 kann demnach als relative Häufigkeit von F in G in Frage kommen, wenn man G nur genügend groß wählt. Falls G hingegen abzählbar unendlich ist, so kommt jeder Limes von konvergierenden Folgen von rationalen Zahlen zwischen 0 und 1 als Limes der relativen Häufigkeit in Frage, also jede reelle Zahl zwischen 0 und 1. In diesem Fall ergibt sich der Schätzwert als Grenzwert der Schätzwerte für endliche Teilklassen von G. Es kann daher (unter Benützung der Abkürzung „$\mathrm{srh_c}$" für „mittels c geschätzte relative Häufigkeit") definiert werden:[5]

D 31 B sei endlich oder abzählbar unendlich, und c sei entsprechend eine reguläre Bestätigungsmethode oder aber eine reguläre Bestätigungsmethode, die die Limeskonvention bezüglich einer passenden Folge von c-Funktionen erfüllt. Φ sei ein erfüllbarer Satz und $\Delta^{(i)}$ sei eine statistische Verteilung für die m Gegenstandsausdrücke $\beta_1, \ldots, \beta_m$ bezüglich des Eigenschaftsausdrucks Π, der die Eigenschaft F bezeichnet; $\Delta^{(i)}$ besage also, daß i der m durch jene Gegenstandsausdrücke bezeichneten Dinge die Eigenschaft F besitzen.

(a) Die Bezugsklasse G sei endlich und umfasse genau die durch die Ausdrücke $\beta_1, \ldots, \beta_m$ bezeichneten Dinge. Dann ist

$$\mathrm{srh_c}(F, G, \Phi) = \sum_{i=0}^{m} [(i/m) \cdot c(\Delta^{(i)}, \Phi)].$$

(b) Die Bezugsklasse G sei abzählbar unendlich und umfasse genau die durch die Ausdrücke $\beta_1, \beta_2, \ldots$ bezeichneten Ausdrücke. Dann ist

$$\mathrm{srh_c}(F, G, \Phi) = \lim_{m \to \infty} \sum_{i=0}^{m} [(i/m) \cdot c(\Delta^{(i)}, \Phi)],$$

falls dieser Limes existiert (ansonsten ist die Funktion nicht definiert).

---

[5] Vgl. Carnap [9], S. 521f. und 525. Carnap erörtert auf den Seiten 523f. auch noch andere Funktionen, die als Schätzungen dienen könnten, nämlich den c-Modus und den c-Median, und zeigt deren Nachteile auf. Zu den sogenannten Paradoxien der Schätzung vgl. seine Bemerkungen auf S. 530—533.

Der durch Mittelung berechnete *Schätzwert*[6] der relativen Häufigkeit muß nicht notwendigerweise eine der überhaupt möglichen relativen Häufigkeiten sein. Wenn etwa G nur drei Elemente enthält, so werden sich, wenn nicht alle Werte $c(\Delta^{(i)}, \Phi)$ miteinander identisch sind, auch andere rationale Werte als 0, $1/3$, $2/3$, 1 oder gar irrationale Schätzwerte ergeben. Wenn man allerdings G hinreichend groß wählt, ist jeder *rationale* Schätzwert auch eine mögliche relative Häufigkeit, und wenn G abzählbar unendlich ist, gilt dies auch für die irrationalen Werte.
Zwischen dem so definierten Begriff des Schätzwertes einer relativen Häufigkeit und dem logischen Wahrscheinlichkeitsbegriff bestehen folgende Beziehungen:

Theorem 23

Die Voraussetzungen von D 29 seien erfüllt; G sei eine Klasse von Gegenständen, die nicht durch Ausdrücke von Φ bezeichnet werden.
(a) G sei endlich; dann gilt:

(α) $\mathrm{srh}_c(F, G, \Phi) = \frac{1}{m} \cdot \sum_{j=1}^{m} c(\beta_j \varepsilon \Pi, \Phi)$;

(β) $\mathrm{srh}_c(F, G, \Phi) = c(\gamma \varepsilon \Pi, \Phi)$, falls γ nicht in Φ vorkommt und falls c die Bedingung (II) (a) erfüllt;

(γ) $\mathrm{srh}_c(F, G, \Phi) = \mathrm{srh}_c(F, G', \Phi)$, falls kein Element von G' durch Gegenstandsausdrücke von Φ bezeichnet wird (G' kann mit G elementfremd sein oder sich mit G überschneiden) und falls c die Bedingung (II) (a) erfüllt.

(b) G sei unendlich, und in Φ mögen die endlich vielen Gegenstandsausdrücke $\beta_1, \ldots, \beta_m$ vorkommen (die jedoch auch Dinge aus G bezeichnen können); c erfülle die Bedingung (II) (a); dann gilt:
(α) $\mathrm{srh}_c(F, G, \Phi) = c(\gamma \varepsilon \Pi, \Phi)$, falls γ verschieden von den Ausdrücken $\beta_1, \ldots, \beta_m$ ist;
(β) $\mathrm{srh}_c(F, G, \Phi) = \mathrm{srh}_c(F, H, \Phi)$, falls H eine unendliche Teilklasse des Gegenstandsbereichs ist, oder falls H eine endliche Teil-

[6] Dieser Schätzwert wird oft auch „Erwartungswert" oder „Mittelwert" genannt.

klasse des Gegenstandsbereichs ist und durch keinen in Φ vorkommenden Gegenstandsausdruck erwähnt wird.

Beweis

Die Voraussetzungen seien gegeben.

(a) G sei endlich.

(α) Nach D 29 (a) gilt, daß die c-Schätzung der relativen Häufigkeit von F in G auf Grund von Φ $\sum_{i=0}^{m} [i \cdot c(\Delta^{(i)}, \Phi) \cdot 1/m]$ und damit $(1/m) \cdot \sum_{i=0}^{m} [i \cdot c(\Delta^{(i)}, \Phi)]$ ist (der Wert $\sum_{i=0}^{m} [i \cdot c(\Delta^{(i)}, \Phi)]$ ist offensichtlich die c-Schätzung der *absoluten* Häufigkeit von F in G). Um die Teilbehauptung (α) zu beweisen, ist also zu zeigen, daß $\sum_{i=0}^{m} [i \cdot c(\Delta^{(i)}, \Phi)] = \sum_{j=1}^{m} c(\beta_j \in \Pi, \Phi)$.

Es sei M_j die Menge der Konjunktionen der Form $(\neg)(\beta_1 \in \Pi) \wedge \ldots \wedge (\neg)(\beta_{j-1} \in \Pi) \wedge \beta_j \in \Pi \wedge (\neg)(\beta_{j+1} \in \Pi) \wedge \ldots \wedge (\neg)(\beta_m \in \Pi)$, wobei also bis auf die Aussage $\beta_j \in \Pi$ die Konjunktionsglieder negiert oder unnegiert sind. Offensichtlich ist $\beta_j \in \Pi$ logisch äquivalent mit der Adjunktion der Sätze aus M_j. Da die Elemente aus M_j miteinander paarweise logisch unverträglich sind, gilt mit den Bedingungen (3) und (4) von D 4: $c(\beta_j \in \Pi, \Phi) = \sum_{\Psi_j \in M_j} c(\Psi_j, \Phi)$, und damit, wenn man über beide Seiten der Gleichung summiert:

$$\sum_{j=1}^{m} c(\beta_j \in \Pi, \Phi) = \sum_{j=1}^{m} \sum_{\Psi_j \in M_j} c(\Psi_j, \Phi).$$

Die Summe $\sum_{j=1}^{m} \sum_{\Psi_j \in M_j} c(\Psi_j, \Phi)$ kann nun folgendermaßen neu geordnet werden: Es sei $M^{(i)}$ die Menge aller Sätze der Form $(\neg)(\beta_1 \in \Pi) \wedge \ldots \wedge (\neg)(\beta_m \in \Pi)$, die i Konjunktionsglieder unnegiert und die übrigen negiert enthalten, und $\Psi^{(i)}$ sei ein Element von $M^{(i)}$. Da dann $\Psi^{(i)}$ i unnegierte Teilsätze enthält, kommt der Summand $c(\Psi^{(i)}, \Phi)$ in der doppelten Summe $\sum_{j=1}^{m} \sum_{\Psi_j \in M_j} c(\Psi_j, \Phi)$ genau i-mal vor. Man erhält deshalb die gleiche Summe, wenn man zunächst die i Glieder $c(\Psi^{(i)}, \Phi)$ zusammenfaßt, dann über alle Elemente aus $M^{(i)}$ und schließlich über alle derartigen Mengen summiert, d. h. es gilt:

$$\sum_{j=1}^{m} \sum_{\Psi_j \varepsilon M_j} c(\Psi_j, \Phi) = \sum_{i=0}^{m} \sum_{\Psi^{(i)} \varepsilon M^{(i)}} i \cdot c(\Psi^{(i)}, \Phi)$$

$$= \sum_{i=0}^{m} i \cdot \sum_{\Psi^{(i)} \varepsilon M^{(i)}} c(\Psi^{(i)}, \Phi)$$

Nun sind, wie man sich leicht klarmacht, die Elemente von $M^{(i)}$ paarweise logisch unverträglich, d. h. von zwei beliebigen Sätzen aus $M^{(i)}$ gilt, daß sie nicht simultan erfüllbar sind; da $\Delta^{(i)}$ als Konjunktionsglieder alle und nur die Aussagen aus $M^{(i)}$ enthält, folgt daraus durch mehrmalige Anwendung der Bedingung (4) von D 4:

$$\sum_{i=0}^{m} i \cdot \sum_{\Psi^{(i)} \varepsilon M^{(i)}} c(\Psi^{(i)}, \Phi) = \sum_{i=0}^{m} [i \cdot c(\Delta^{(i)}, \Phi)].$$

Aus allen diesen Gleichungen ergibt sich:

$$\sum_{i=0}^{m} [i \cdot c(\Delta^{(i)}, \Phi)] = \sum_{j=1}^{m} (\beta_j \varepsilon \Pi, \Phi),$$

was zu zeigen war.

(β) Mit Bedingung (II) (a) (vgl. D 17a) folgt diese Teilbehauptung aus (a) (α), da wegen der Symmetrie bezüglich Gegenstandsausdrükken gilt: $c(\beta_1 \varepsilon \Pi, \Phi) = \ldots = c(\beta_m \varepsilon \Pi, \Phi)$ und somit

$$\sum_{j=1}^{m} c(\beta_j \varepsilon \Pi, \Phi) = m \cdot c(\beta_1 \varepsilon \Pi, \Phi) = m \cdot c(\gamma \varepsilon \Pi, \Phi).$$

(γ) Durch zweimalige Anwendung von (a) (β) auf zwei Bezugsklassen G und G' der vorausgesetzten Art ergibt sich für einen nicht in Φ vorkommenden Gegenstandsausdruck γ:

$$\mathrm{srh}_c(F, G, \Phi) = c(\gamma \varepsilon \Pi, \Phi)$$
$$= \mathrm{srh}_c(F, G', \Phi).$$

(b) G sei unendlich, und H enthalte neben den Elementen von G eventuell auch noch einige oder alle der in Φ erwähnten Gegenstände (H kann aber auch mit G identisch sein); c erfülle ferner die Bedingung (II) (a).

(α) Da in Φ nur endlich viele Gegenstandsausdrücke vorkommen, während die Anzahl der Gegenstände aus G unendlich ist und da c wegen (II) (a) keinen Unterschied zwischen den durch Ausdrücke aus Φ benannten Dingen und den Gegenständen aus G macht, gleichen

sich die Werte der Summen $\sum_{i=0}^{m} [(i/m) \cdot c(\Delta^{(i)}, \Phi)]$, wobei die Ausdrücke aus $\Delta^{(i)}$ Gegenstände aus H bezeichnen, und $\sum_{i=0}^{m} [(i/m) \cdot c(\Delta^{(i)}, \Phi)]$, wobei die Ausdrücke aus $\Delta^{(i)}$ nur Gegenstände aus G bezeichnen, mit wachsendem m immer mehr an. Beide Summen streben also entweder keinem oder dem gleichen Grenzwert zu.

(β) Diese Teilbehauptung ergibt sich unmittelbar aus (a) (α) und (b) (α). Damit ist der Lehrsatz bewiesen.

Wenn in Theorem 23 (a) (α) als Bezugsklasse eine Klasse genommen wird, die nur den durch β_1 bezeichneten Gegenstand enthält, dann ist demnach die mittels der regulären Bestätigungsmethode c vorgenommene Schätzung der relativen Häufigkeit von F in G hinsichtlich Φ gleich der durch c gemessenen Wahrscheinlichkeit dafür, daß auf Grund von Φ dieser Gegenstand die Eigenschaft F besitzt. Ist c außerdem noch symmetrisch hinsichtlich der Gegenstandsausdrücke, behandelt sie also alle Gegenstände aus einer (dann umfangsreicheren) Klasse G gleich, so daß die Wahrscheinlichkeiten hinsichtlich des Datums, daß die einzelnen Dinge aus G F sind, gleich sind, dann ist darüber hinaus auch für größere Bezugsklassen G die mittels c und hinsichtlich Φ vorgenommene Schätzung der relativen Häufigkeit von F in G identisch mit dem singulären Voraussageschluß bezüglich eines beliebigen, noch nicht in Φ erwähnten Gegenstandes. Falls G dabei endlich und nicht zu groß hinsichtlich der Anzahl der in Φ erwähnten Gegenstände ist, darf G keines dieser Dinge enthalten; ist G groß hinsichtlich dieser Anzahl, dann ist der Fehler klein, den man macht, wenn G auch Dinge enthält, die in Φ erwähnt werden, und wenn G abzählbar unendlich ist, konvergiert der Fehler gegen 0. Wenn G hinreichend groß ist hinsichtlich dieser Anzahl, so kann man die Wahrscheinlichkeit demnach als Schätzung der relativen Häufigkeit nehmen; man kann sich im Hinblick darauf bei praktischen Situationen, etwa bei Wetten, einstellen und als Wettquotient die Wahrscheinlichkeit bzw. die geschätzte relative Häufigkeit nehmen. Nach Theorem 23 (a) (β) und (b) (α) spielt dabei die Größe und Beschaffenheit der

Bezugsklasse keine Rolle (bei kleinen Bezugsklassen muß noch vorausgesetzt werden, daß Φ über ihre Elemente nichts aussagt).
In vielen Fällen, etwa wenn c eine L-Methode ist, kann der Fall eintreten, daß der c-Schätzwert des Grenzwertes der relativen Häufigkeit von F hinsichtlich einer unendlichen Bezugsklasse (also die Schätzung der sogenannten *objektiven* Wahrscheinlichkeit) existiert, ohne daß man hierbei irgendeine Gewähr dafür hat, daß es diese objektive Wahrscheinlichkeit selbst (diesen Grenzwert) tatsächlich gibt. Es kann sogar effektiv der Fall eintreten, daß man zwar die objektiven Wahrscheinlichkeiten *schätzen* kann, daß diese jedoch gar nicht vorhanden sind, da, wie sich später herausstellt, die relativen Häufigkeiten nicht konvergieren. Ein Beispiel dafür wäre, wenn G die Gesamtheit der Erdjahre und F die Eigenschaft ist, eine relativ warme mittlere Jahrestemperatur zu haben, und wenn sich später erweist, daß etwa die Eiszeiten in immer längeren Perioden eintreten.

2. *Über statistische Verteilungen*

In diesem Abschnitt wird stets vorausgesetzt, daß die zugrundegelegte Sprache nur ein Grundprädikat Π und daher nur zwei V-Prädikate enthält; Π bezeichne unter der vorgegebenen Interpretation die Eigenschaft (Klasse) F. An Gegenstandsausdrücken enthalte sie zunächst lediglich α_1, d. h. der Bereich B_1, über dem sie interpretiert ist, bestehe nur aus einem Gegenstand. Dann gibt es lediglich zwei mögliche relative Häufigkeiten von F in B_1, nämlich 0 (= 0/1) und 1 (= 1/1). Fügt man B_1 ein neues Element hinzu, so daß B_2 entsteht (und addiert man der Sprache entsprechend den Gegenstandsausdruck α_2), dann werden drei relative Häufigkeiten möglich, nämlich 0, $^1/_2$ und 1. Allgemein gilt, daß, falls die Sprache über dem Bereich B_n mit n Elementen interpretiert ist, die relative Häufigkeit von F in B_n die $n+1$ Werte 0, 1/n, 2/n, ..., $(n-1)/n$ und 1 haben kann.
Es soll nun untersucht werden, wie groß die *logische* (Apriori-) *Wahrscheinlichkeit* für die einzelnen relativen Häufigkeiten (für die einzelnen *objektiven Wahrscheinlichkeiten*) ist, mit etwas vageren umgangssprachlichen Worten: wie groß die Chance ist, eine derartige objek-

tive Wahrscheinlichkeit anzutreffen (wenn man noch kein weiteres Wissen über die Eigenschaften der Dinge hat). Der Sicherheitsgrad, den ich den einzelnen objektiven Wahrscheinlichkeiten zumesse, hängt natürlich davon ab, wie groß man die Uniformität des Gegenstandsbereichs hinsichtlich der beiden V-Eigenschaften F und dessen Komplement einschätzt, und damit, welche L-Methode (eventuell einschließlich der beiden Grenzmethoden c^0 und c^∞) man wählt.

Um die logische Wahrscheinlichkeit für eine bestimmte objektive Wahrscheinlichkeit durch eine L-Methode messen zu können, ist es notwendig, daß die relativen Häufigkeiten in der vorgegebenen Objektsprache ausdrückbar sind. Dies ist jedoch der Fall, denn die statistischen Verteilungen (bezüglich der beiden V-Prädikate) sind die Aussagen, die diese möglichen relativen Häufigkeiten beschreiben. Für das Folgende sei $\Delta^{(i)}$ eine derartige statistische Beschreibung und besage, daß i der n Gegenstände des Objektbereichs die Eigenschaft F besitzen.

Für c^1, c^2 und c^4 etwa ergeben sich dann folgende Aprioriwerte für die statistischen Verteilungen:[7]

(1) Bei B_1 ist $\Delta^{(0)}$ der Satz $\neg\,(\alpha_1 \varepsilon \Pi)$ und $\Delta^{(1)}$ die Aussage $\alpha_1 \varepsilon \Pi$. Man erhält dann die Werte:

$c^1(\Delta^{(0)}, T) = 1/2$; $c^1(\Delta^{(1)}, T) = 1/2$.

$c^2(\Delta^{(0)}, T) = 1/2$; $c^2(\Delta^{(1)}, T) = 1/2$.

$c^4(\Delta^{(0)}, T) = 1/2$; $c^4(\Delta^{(1)}, T) = 1/2$.

(2) Bei B_2 steht $\Delta^{(0)}$ für $\neg\,(\alpha_1 \varepsilon \Pi) \wedge \neg\,(\alpha_2 \varepsilon \Pi)$,
$\Delta^{(1)}$ für $[\neg\,(\alpha_1 \varepsilon \Pi) \wedge \alpha_2 \varepsilon \Pi] \vee [\alpha_1 \varepsilon \Pi \wedge \neg\,(\alpha_2 \varepsilon \Pi)]$
und $\Delta^{(2)}$ für $\alpha_1 \varepsilon \Pi \wedge \alpha_2 \varepsilon \Pi$. Es ergeben sich dann die Werte:

$c^1(\Delta^{(0)}, T) = 3/8$; $c^1(\Delta^{(1)}, T) = 2/8$; $c^1(\Delta^{(2)}, T) = 3/8$.

$c^2(\Delta^{(0)}, T) = 1/3$; $c^2(\Delta^{(1)}, T) = 1/3$; $c^2(\Delta^{(2)}, T) = 1/3$.

$c^4(\Delta^{(0)}, T) = 3/10$; $c^2(\Delta^{(1)}, T) = 4/10$; $c^4(\Delta^{(2)}, T) = 3/10$.

---

[7] Die Werte ergeben sich dabei nach Theorem 15; vgl. auch das zu Beginn des 2. Abschnittes des Kapitels IV über die Aprioriwahrscheinlichkeit von Zustandsbeschreibungen Gesagte. Da statistische Verteilungen Adjunktionen von Zustandsbeschreibungen sind und da zwei verschiedene Zustandsbeschreibungen nicht simultan erfüllbar sind, ergeben sich die c^L-Werte für statistische Verteilungen nach Bedingung (4) von D 4. Dabei ist zu beachten, daß eine statistische Verteilung, die besagt, daß i der n Gegenstände die Eigenschaft F hat, $\binom{n}{i}$ Adjunktionsglieder hat. Die allgemeine Formel zur Berechnung dieser Werte findet der Leser bei Carnap-Stegmüller [1], S. 249f.

(3) Bei B_3 ist $\Delta^{(0)}$ der Satz $\neg(\alpha_1 \varepsilon \Pi) \wedge \neg(\alpha_2 \varepsilon \Pi) \wedge \neg(\alpha_3 \varepsilon \Pi)$, ist $\Delta^{(1)}$ der Satz $[\neg(\alpha_1 \varepsilon \Pi) \wedge \neg(\alpha_2 \varepsilon \Pi) \wedge \alpha_3 \varepsilon \Pi] \vee [\neg(\alpha_1 \varepsilon \Pi) \wedge \alpha_2 \varepsilon \Pi \wedge \neg(\alpha_3 \varepsilon \Pi)] \vee [\alpha_1 \varepsilon \Pi \wedge \neg(\alpha_2 \varepsilon \Pi) \wedge \neg(\alpha_3 \varepsilon \Pi)]$, ist $\Delta^{(2)}$ der Satz $[\neg(\alpha_1 \varepsilon \Pi) \wedge \alpha_2 \varepsilon \Pi \wedge \alpha_3 \varepsilon \Pi] \vee [\alpha_1 \varepsilon \Pi \wedge \neg(\alpha_2 \varepsilon \Pi) \wedge \alpha_3 \varepsilon \Pi] \vee [\alpha_1 \varepsilon \Pi \wedge \alpha_2 \varepsilon \Pi \wedge \neg(\alpha_3 \varepsilon \Pi)]$ und ist schließlich $\Delta^{(3)}$ der Satz $\alpha_1 \varepsilon \Pi \wedge \alpha_2 \varepsilon \Pi \wedge \alpha_3 \varepsilon \Pi$. An Aprioriwerten erhält man dann:

$c^1(\Delta^{(0)}, T) = {}^5/_{16}$; $c^1(\Delta^{(1)}, T) = {}^3/_{16}$;
$c^1(\Delta^{(2)}, T) = {}^3/_{16}$; $c^1(\Delta^{(3)}, T) = {}^5/_{16}$.
$c^2(\Delta^{(0)}, T) = {}^1/_4$; $c^2(\Delta^{(1)}, T) = {}^1/_4$;
$c^2(\Delta^{(2)}, T) = {}^1/_4$; $c^2(\Delta^{(3)}, T) = {}^1/_4$.
$c^4(\Delta^{(0)}, T) = {}^2/_{10}$; $c^4(\Delta^{(1)}, T) = {}^3/_{10}$;
$c^4(\Delta^{(2)}, T) = {}^3/_{10}$; $c^4(\Delta^{(3)}, T) = {}^2/_{10}$.

(4) Bei B_4 ist $\Delta^{(0)}$ der Satz $\neg(\alpha_1 \varepsilon \Pi) \wedge \ldots \wedge \neg(\alpha_4 \varepsilon \Pi)$, ist $\Delta^{(1)}$ der Satz $[\neg(\alpha_1 \varepsilon \Pi) \wedge \neg(\alpha_2 \varepsilon \Pi) \wedge \neg(\alpha_3 \varepsilon \Pi) \wedge \alpha_4 \varepsilon \Pi] \vee \ldots \vee [\alpha_1 \varepsilon \Pi \wedge \neg(\alpha_2 \varepsilon \Pi) \wedge \neg(\alpha_3 \varepsilon \Pi) \wedge \neg(\alpha_4 \varepsilon \Pi)]$, ist $\Delta^{(2)}$ der Satz $[\neg(\alpha_1 \varepsilon \Pi) \wedge \neg(\alpha_2 \varepsilon \Pi) \wedge \alpha_3 \varepsilon \Pi \wedge \alpha_4 \varepsilon \Pi] \vee \ldots \vee [\alpha_1 \varepsilon \Pi \wedge \alpha_2 \varepsilon \Pi \wedge \neg(\alpha_3 \varepsilon \Pi) \wedge \neg(\alpha_4 \varepsilon \Pi)]$, ist $\Delta^{(3)}$ der Satz $[\neg(\alpha_1 \varepsilon \Pi) \wedge \alpha_2 \varepsilon \Pi \wedge \alpha_3 \varepsilon \Pi \wedge \alpha_4 \varepsilon \Pi] \vee \ldots \vee [\alpha_1 \varepsilon \Pi \wedge \alpha_2 \varepsilon \Pi \wedge \alpha_3 \varepsilon \Pi \wedge \neg(\alpha_4 \varepsilon \Pi)]$ und ist schließlich $\Delta^{(4)}$ der Satz $\alpha_1 \varepsilon \Pi \wedge \alpha_2 \varepsilon \Pi \wedge \alpha_3 \varepsilon \Pi \wedge \alpha_4 \varepsilon \Pi$. Man erhält dann die Werte:

$c^1(\Delta^{(0)}, T) = {}^{35}/_{128}$; $c^1(\Delta^{(1)}, T) = {}^{20}/_{128}$; $c^1(\Sigma^{(2)}, T) = {}^{18}/_{128}$;
$c^1(\Delta^{(3)}, T) = {}^{20}/_{128}$; $c^1(\Delta^{(4)}, T) = {}^{35}/_{128}$.
$c^2(\Delta^{(0)}, T) = {}^1/_5$; $c^2(\Delta^{(1)}, T) = {}^1/_5$; $c^2(\Sigma^{(2)}, T) = {}^1/_5$;
$c^2(\Delta^{(3)}, T) = {}^1/_5$; $c^2(\Delta^{(4)}, T) = {}^1/_5$;
$c^4(\Delta^{(0)}, T) = {}^5/_{35}$; $c^4(\Delta^{(1)}, T) = {}^8/_{35}$; $c^4(\Sigma^{(2)}, T) = {}^9/_{35}$;
$c^4(\Delta^{(3)}, T) = {}^8/_{35}$; $c^4(\Delta^{(4)}, T) = {}^5/_{35}$.

Graphisch lassen sich diese Werte für c^1 folgendermaßen veranschaulichen:

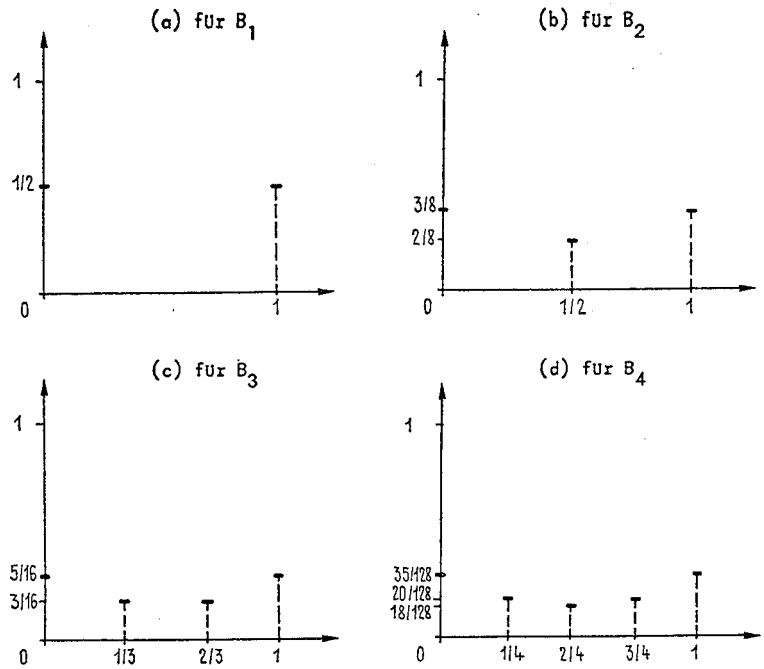


Fig. 5

Für c^2 sehen die geometrischen Darstellungen dieser Werte folgendermaßen aus:

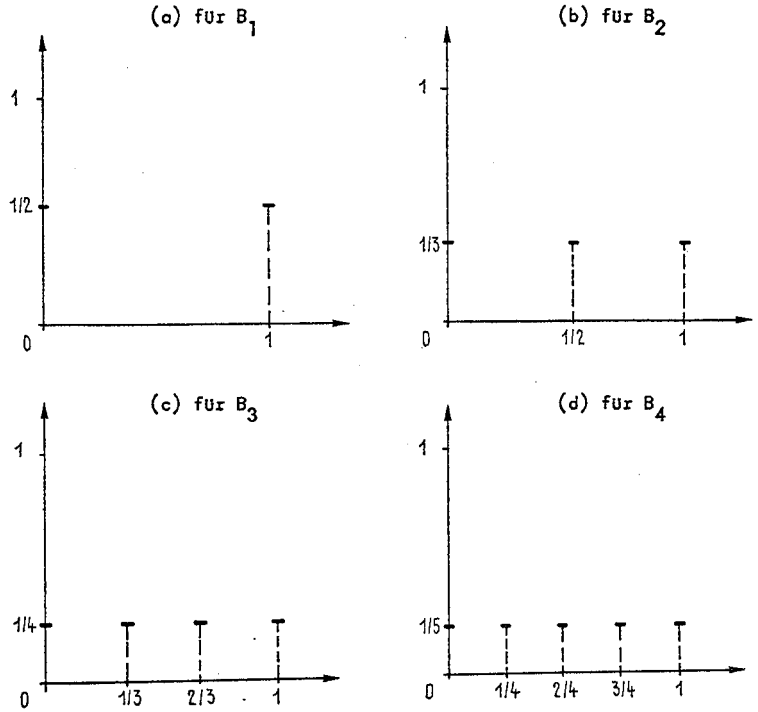


Fig. 6

Für c^4 schließlich ergeben sich folgende Abbildungen:

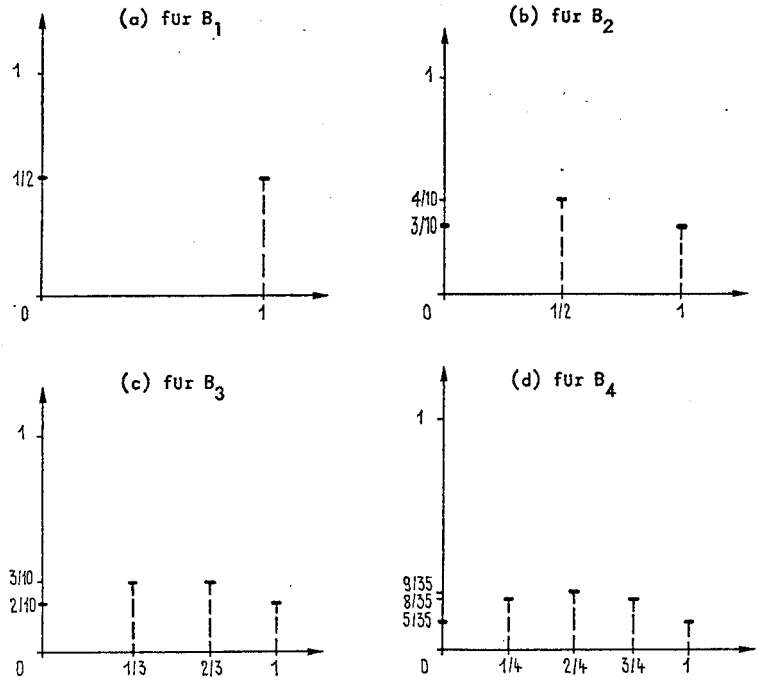


Fig. 7

Je größer der Gegenstandsbereich B also ist, um so mehr relative Häufigkeiten sind möglich und um so kleiner werden auf den horizontalen Achsen der Abbildungen die Intervalle 1/m. Die Punkte von 0 bis 1 auf dieser Achse charakterisieren dabei die möglichen relativen Häufigkeiten von F in B, genauer: wenn B von endlicher Anzahl n ist, dann kommen als relative Häufigkeiten nur die Werte 0, ..., i/n, ..., 1 in Frage; wenn B hingegen abzählbar unendlich ist, kommen nicht nur alle *rationalen* Zahlen von 0 bis 1, sondern auch alle Folgen von rationalen Zahlen und somit alle *reellen* Zahlen dieses Bereichs als Grenzwerte der relativen Häufigkeiten von F in B in Frage.
Die mit einer Funktion c^L gemessene Aprioriwahrscheinlichkeit (Apriorisicherheit, Apriorichance) für eine gewisse relative Häufigkeit kann beliebig der 0 angenähert werden, wenn man die Intervalle auf der horizontalen Achse hinreichend verkleinert, d. h. wenn man

die Anzahl der Elemente des Gegenstandsbereichs hinreichend vergrößert, so daß genügend zusätzliche relative Häufigkeiten möglich werden. Von zwei relativen Häufigkeiten ist durch Theorem 20 bereits gezeigt worden, daß sie gegen 0 konvergieren, nämlich von der relativen Häufigkeit 1, die für den endlichen Bereich B_n durch $\Delta^{(n)}$ beschrieben wird, womit gesagt ist, daß alle Dinge aus B_n die Eigenschaft F haben, und von der relativen Häufigkeit 0, die für B_n durch $\Delta^{(0)}$ beschrieben wird; $\Delta^{(0)}$ besagt, daß jedes Ding aus B_n die Eigenschaft nicht-F besitzt, d. h. daß *kein* Ding aus B_n in der Klasse F ist. Es bereitet keine prinzipiellen Schwierigkeiten, auch von den anderen objektiven Wahrscheinlichkeiten zu zeigen, daß ihre logische Wahrscheinlichkeit in einem abzählbar unendlichen Gegenstandsbereich im Hinblick auf die Tautologie 0 ist.

Da die L-Methoden, von c^2 abgesehen, bezüglich der einzelnen relativen Häufigkeiten nicht gleichmäßig gegen 0 konvergieren, ist es für unendliche Gegenstandsbereiche zweckmäßig, anders vorzugehen. Gegeben sei der endliche Abschnitt B_n von B_∞. Es wird nun nicht die Aprioriwahrscheinlichkeit für die statistische Verteilung $\Delta^{(i)}$, sondern die Summe aller Wahrscheinlichkeiten für die statistischen Verteilungen $\Delta^{(j)}$ mit $0 \leqslant j \leqslant i$ betrachtet. Da sich die Sätze $\Delta^{(0)}$, ..., $\Delta^{(i)}$ wechselseitig logisch ausschließen, gilt mit Bedingung (4) von D 4: $c^L(\Delta^{(0)}, T) + \ldots + c^L(\Delta^{(i)}, T) = c^L(\Delta^{(0)} \vee \ldots \vee \Delta^{(i)}, T)$. Es wird daher im folgenden nach der durch c^L gemessenen Aprioriwahrscheinlichkeit für die Hypothese gefragt, daß die relative Häufigkeit von F in B_n kleiner als oder identisch mit i/n ist, also nach dem Wert $c^L(\Delta^{(0)} \vee \ldots \vee \Delta^{(i)}, T)$. Für c^1 können diese Werte hinsichtlich der Bereiche $B_1, \ldots, B_4$ folgendermaßen veranschaulicht werden:

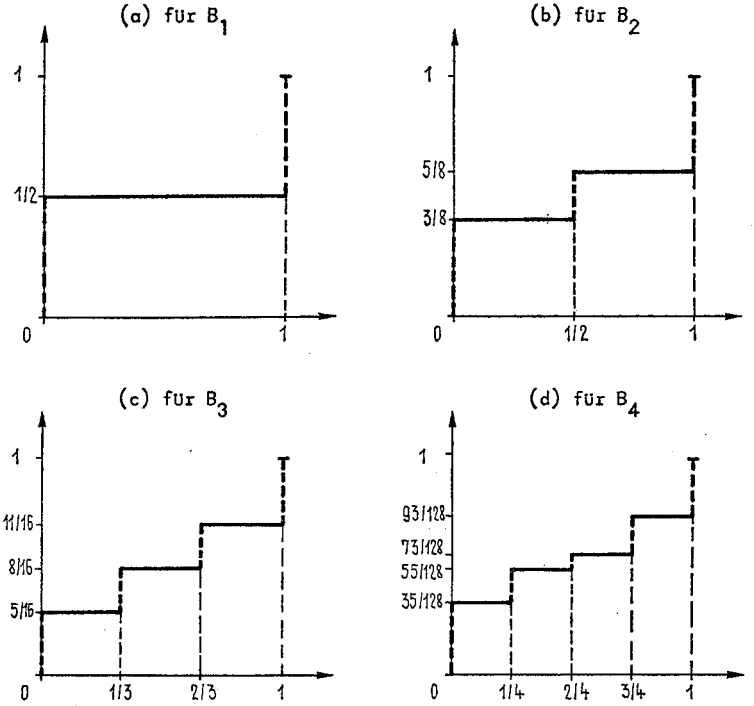


Fig. 8

Für c^2 hingegen nehmen die Abbildungen dieser Summen folgende Form an:

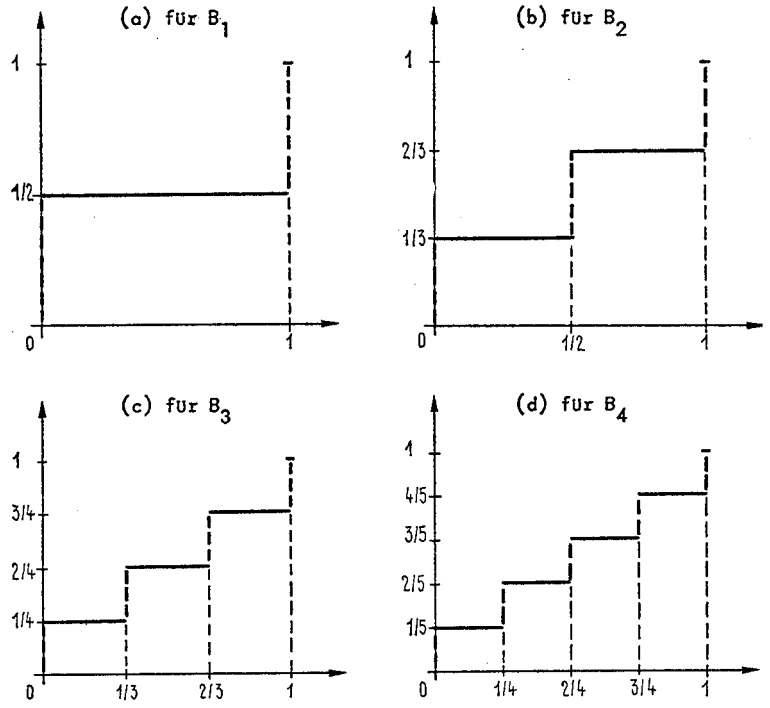


Fig. 9

Unter Verwendung von c^4 ergeben sich schließlich folgende Aprioriwerte dafür, daß die relative Häufigkeit von F in B_n kleiner oder gleich i/n ist:

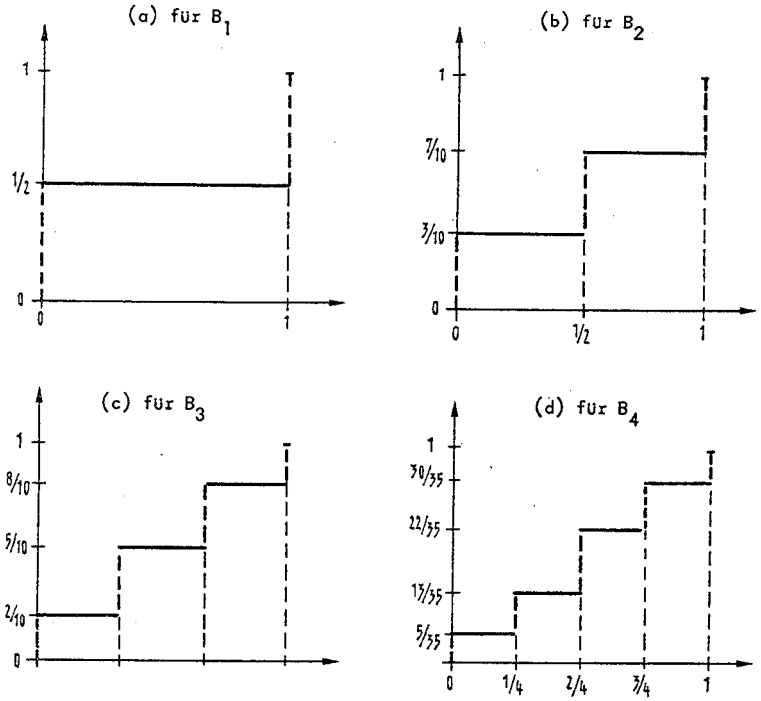


Fig. 10

Wenn man zu den Figuren 5 bis 10 weitere Abbildungen für B_5, B_6, B_7 usw. hinzufügt, so bemerkt man folgendes: c^1 hat seine höchsten Werte stets für die relativen Häufigkeiten 0 und 1 und nimmt gegen $1/2$ hin monoton ab; die relative Häufigkeit $1/2$ (oder, falls die Anzahl n der Gegenstände ungerade ist, die beiden an $1/2$ angrenzenden Werte) erhält bei allen endlichen Gegenstandsbereichen den kleinsten Wert. c^1 schätzt also a priori die objektive Wahrscheinlichkeit dafür, daß (ungefähr) gleich viele Dinge aus B_n F und nicht-F sind, am geringsten ein, während sie der Behauptung, daß B_n hinsichtlich F vollständig uniform ist, so daß also alle oder aber kein Gegenstand des Bereichs F ist, den relativ höchsten Glaubwürdigkeitsgrad gibt. c^4 verteilt die Aprioriwerte für die objektiven Wahrscheinlichkeiten umgekehrt, ordnet also der relativen Häufigkeit $1/2$ die relativ höchste und dem Wert 0 bzw. 1 die relativ geringste Glaubwürdigkeit zu.

c^2 hingegen verteilt die Aprioriwahrscheinlichkeiten gleichmäßig auf alle relativen Häufigkeiten.
Diese Verhältnisse können auch an den Treppenfunktionen der Figuren 8 (d), 9 (d) und 10 (d) sowie an deren Fortsetzungen für B_n mit $n > 4$ abgelesen werden: Die durch c^1 gemessene Aprioriwahrscheinlichkeit für die einzelnen möglichen relativen Häufigkeiten haben bei 0 und 1 die größten und bei $1/2$ die kleinsten Stufen, während es sich bei c^4 umgekehrt verhält; die Stufen von c^2 hingegen sind überall gleich groß. Läßt man n gegen ∞ gehen, so gleichen sich diese durch c^2 erzeugten Treppenfunktionen immer mehr der Geraden $f(x) = x$ an, während sich die durch c^4 erzeugten einer „S"-förmigen Kurve und die durch c^1 erzeugte einer Kurve angleicht, die einem umgekehrten „S" ähnelt.
Wählt man eine Folge von Funktionen c^L mit $1 < L < 2$ oder $2 < L < 4$, so flachen für B_∞ die Kurven um so mehr ab, je näher L bei 2 liegt, und wählt man eine Folge von Funktionen c^L mit $0 < L < 1$ oder $4 < L < \infty$, so wuchten sie sich immer mehr aus und ergeben als Limes (d. h. wenn L gegen 0 oder gegen ∞ geht) Stufen der folgenden Art:

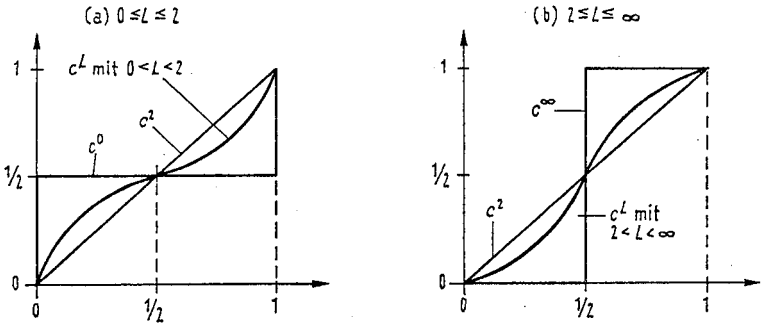


Fig. 11

Jede L-Methode erzeugt auf diese Weise genau eine derartige Funktion (die von Mathematikern üblicherweise „Verteilungsfunktion" genannt wird), und zwei verschiedene L-Methoden erzeugen zwei verschiedene Verteilungsfunktionen. Aus diesen können übrigens die L-Methoden zurückgewonnen werden, wie nun an einem Beispiel

gezeigt wird. Es handelt sich dabei um jene Verteilungsfunktion, die allen objektiven Wahrscheinlichkeiten a priori die gleichen logischen Wahrscheinlichkeiten zuordnet und die daher durch die Funktion $f(x) = x$ charakterisiert ist. Bei endlichen Bereichen B_n ist diese Forderung der Gleichverteilung identisch mit der von Carnap entwickelten Funktion c^*. Für abzählbar unendliche Gesamtheiten ist sie identisch mit der (von Venn so genannten) Sukzessionsregel von Laplace.

Theorem 24

Der Gegenstandsbereich B sei abzählbar unendlich, gegeben sei ein Grundprädikat Π_1 (so daß $q = 2$), und c sei eine reguläre Bestätigungsmethode, die die Limeskonvention für Gegenstandsausdrücke hinsichtlich einer passenden Folge von c-Funktionen erfüllt. Dann gilt: c erfüllt die Bedingung (II) (a) und gibt allen statistischen Verteilungen (hinsichtlich B und jedes endlichen Abschnittes B_n von B) die gleichen Aprioriwerte genau dann, wenn c eine adäquate induktive Methode mit $L = 2$ ist.

Beweis

Die Voraussetzungen seien gegeben.
(a) c sei die Funktion c^2; dann erfüllt c nach Theorem 15 die Bedingung (II) (a). B_n sei ein beliebiger endlicher Abschnitt von B, der aus den durch die Ausdrücke $\alpha_1, \ldots, \alpha_n$ bezeichneten Dingen besteht; zu zeigen ist, daß c^2 allen statistischen Verteilungen für die Gegenstandsausdrücke $\alpha_1, \ldots, \alpha_n$ bezüglich der beiden V-Prädikate die gleichen Aprioriwerte zuteilt, und zwar jeweils $1/(n + 1)$, da es für die B_n $n + 1$ statistische Verteilungen gibt. Gegeben sei eine statistische Verteilung Γ für die $\alpha_1, \ldots, \alpha_n$ bezüglich der beiden V-Prädikate. Γ ist eine Adjunktion einer gewissen Anzahl von individuellen Verteilungen, die hinsichtlich der relativen Häufigkeiten das gleiche besagen wie Γ, die sich jedoch hinsichtlich der einzelnen Gegenstände und deren Eigenschaften voneinander unterscheiden. Nach Lemma 6 (mit $d = 2$) besteht Γ aus $\binom{n}{m}$, d. h. aus $n!/[m! \cdot (n - m)!]$ Adjunktionsgliedern, wenn Γ besagt, daß m der n Gegenstände die Eigen-

schaft F_1 haben (es ist also $0 \leqslant m \leqslant n$). Der Aprioriwert einer solchen individuellen Verteilung Σ berechnet sich nun nach der Gleichung: $c^2(\Sigma, T) = c^2(\Psi_{i_1}, T) \cdot \ldots \cdot c^2(\Psi_{i_n}, \Phi_{n-1})$. Wegen Bedingung (3) von D 4 kann ohne Einschränkung der Allgemeinheit angenommen werden, daß die ersten n Konjunktionsglieder von Σ unnegiert und die letzten n — m negiert sind bzw. daß die ersten m das eine und die letzten n — m das andere V-Prädikat enthalten (nach (3) können die Konjunktionsglieder von Σ ja vertauscht werden, ohne daß sich der Wert von $c^2(\Sigma, T)$ ändert). Aus der obigen Gleichung erhält man dann durch Einsetzung nach Theorem 15:

$$\begin{aligned} c^2\ (\Sigma, T) &= \frac{1}{2} \cdot \frac{1+1}{1+2} \cdot \ldots \cdot \frac{(n-m-1)+1}{(n-m-1)+2} \\ &\quad \cdot \frac{1}{(n-m)+2} \cdot \frac{1+1}{(n-m+1)+2} \cdot \ldots \cdot \frac{(m-1)+1}{n+1} \\ &= \frac{(n-m)! \cdot m!}{(n+1)!}. \end{aligned}$$

Da c^2 die Bedingungen (II) (a) und (I) (3) erfüllt, gibt sie allen Adjunktionsgliedern von Γ die gleiche Aprioriwahrscheinlichkeit. Γ hat nun $\binom{n}{m}$ derartige Komponenten; daher gilt:

$$\begin{aligned} c^2\ (\Gamma, T) &= \frac{(n-m)! \cdot m!}{(n+1)!} \cdot \binom{n}{m} \\ &= \frac{(n-m)! \cdot m! \cdot n!}{(n+1)! \cdot m! \cdot (n-m)!} \\ &= \frac{(n-m)! \cdot m! \cdot n!}{(n+1) \cdot n! \cdot m! \cdot (n-m)!} \\ &= \frac{1}{n+1}. \end{aligned}$$

Da die unendliche Gesamtheit ${}_1c^2, {}_2c^2, {}_3c^2, {}_4c^2, \ldots$ von Bestätigungsmethoden bezüglich $B_1, B_2, B_3, B_4, \ldots$ eine passende Folge für ${}_\infty c^2$ ist, ergeben sich die ${}_\infty c^2$-Werte für statistische Verteilungen bezüglich B ($= B_\infty$) durch die Limeskonvention für Gegenstandsausdrücke hinsichtlich dieser passenden Folge (die statistischen Verteilungen sind

hier wieder Sätze von unendlicher Länge); offensichtlich bleibt bei diesem Grenzübergang die Gleichverteilung erhalten.

(b) c erfülle die Bedingung (II) (a) und gebe allen statistischen Verteilungen (hinsichtlich B und jedes endlichen Abschnittes B_n von B) die gleichen Aprioriwerte. Zu zeigen ist, daß daraus die Sukzessionsregel von Laplace folgt.[8] B_n sei ein endlicher Abschnitt von B, der die ersten n Elemente enthält. $\Delta^{(i)}$ sei eine statistische Verteilung, die besagt, daß i der n Dinge F sind (und daher $n - i$ die Eigenschaft nicht-F haben). Offensichtlich gibt es genau $n + 1$ derartige statistische Verteilungen. Φ_k sei eine individuelle Verteilung für die ersten k Gegenstandsausdrücke (es ist also $k \leqslant n$ zu wählen, und da später zum singulären Voraussageschluß übergegangen wird, sogar $k < n$; setzt man k als fest voraus, so muß n als hinreichend groß angenommen werden, so daß $k < n$). c ist nach Voraussetzung eine reguläre Bestätigungsmethode. Die Sätze $\Delta^{(0)}, \Delta^{(1)}, \ldots, \Delta^{(n)}$ sind miteinander paarweise logisch unverträglich und in ihrer Gesamtheit logisch erschöpfend. Durch Einsetzung in Theorem 3 (b) erhält man, wenn man „Δ_i", „Φ" und „Γ" durch „$\Delta^{(i)}$", „T" und „Φ_k" ersetzt:

$$c(\Delta^{(i)}, \Phi_k) = \frac{c(\Phi_k, \Delta^{(i)}) \cdot c(\Delta^{(i)}, T)}{\sum_{i=0}^{n} [c(\Phi_k, \Delta^{(i)}) \cdot c(\Delta^{(i)}, T)]}.$$

Weiter ist angenommen worden, daß c den statistischen Verteilungen $\Delta^{(i)}$ die gleichen Aprioriwerte zuordnet, so daß also für alle i gilt: $c(\Delta^{(i)}, T) = 1/(n + 1)$.

Beide Gleichungen implizieren:

$$c(\Delta^{(i)}, \Phi_k) = \frac{c(\Phi_k, \Delta^{(i)}) \cdot \frac{1}{n+1}}{\sum_{i=0}^{n} \left[c(\Phi_k, \Delta^{(i)}) \cdot \frac{1}{n+1} \right]}.$$

Es wird nun der Erwartungswert von F in B_n (hinsichtlich Φ_k) gebildet, d. h. es wird nach D 29 die relative Häufigkeit von F in B_n geschätzt:

---

[8] Vgl. Laplace [6], Bd. VIII S. 30f., Bd. X S. 325ff. und Bd. VII S. 16ff., sowie auch Wright [2], S. 102—112 und 213; für die Explikation der Deduktion dieser Regel in mathematischen Lehrbüchern der Wahrscheinlichkeitslehre bin ich Fräulein Dipl.-Math. Veronika Reiß zu Dank verpflichtet.

$$\mathrm{srh_c}(F, B_n, \Phi_k) = \sum_{i=0}^{n} \left[\frac{i}{n} \cdot c(\Delta^{(i)}, \Phi_k) \right]$$

$$= \sum_{i=0}^{n} \left[\frac{i}{n} \cdot \frac{c(\Phi_k, \Delta^{(i)}) \cdot \frac{1}{n+1}}{\sum_{i=0}^{n} \left[c(\Phi_k, \Delta^{(i)}) \cdot \frac{1}{n+1} \right]} \right]$$

$$= \frac{\sum_{i=0}^{n} \left[\frac{i}{n} \cdot c(\Phi_k, \Delta^{(i)}) \cdot \frac{1}{n+1} \right]}{\sum_{i=0}^{n} \left[c(\Phi_k, \Delta^{(i)}) \cdot \frac{1}{n+1} \right]}.$$

Die Forderung, daß für jeden Abschnitt B_n von B die Funktion ${}_nc$ den statistischen Verteilungen gleiche Aprioriwerte gibt und der Bedingung (II) (a) genügt, impliziert, daß die Gesamtheit der Funktionen ${}_1c$, ${}_2c$, ${}_3c$, ${}_4c$, ..., eine passende Folge im Sinne von D 28 darstellt. Dies kann folgendermaßen gezeigt werden: Es sei ${}_n\Sigma$ eine Zustandsbeschreibung aus der Teilsprache, die an Gegenstandsausdrücken $\alpha_1, \ldots, \alpha_n$ enthält; ${}_n\Sigma$ besage (unter anderem), daß m Dinge F und n — m Dinge nicht-F sind. In der Teilsprache, die an Gegenstandsausdrücken zusätzlich noch α_{n+1} enthält, ist ${}_n\Sigma$ keine Zustandsbeschreibung mehr, aber logisch äquivalent mit zwei Zustandsbeschreibungen ${}_{n+1}\Sigma'$ und ${}_{n+1}\Sigma''$, von denen die eine besagt, daß m der n + 1 Dinge F und n + 1 — m der Gegenstände nicht-F sind, und die andere, daß m + 1 der n + 1 Dinge F und die übrigen n — m nicht-F sind. Nun gilt:

$${}_nc({}_n\Sigma, T) = \frac{\frac{1}{n+1}}{\binom{n}{m}}$$

$$= \frac{m! \cdot (n-m)!}{(n+1)!},$$

da ${}_nc$ den statistischen Verteilungen für B_n die Werte $1/(n+1)$ gibt und diese wiederum wegen (II) (a) auf die jeweils $\binom{n}{m}$ Adjunktionsglieder gleichermaßen aufgeteilt werden.

$_{n+1}\Sigma'$ und $_{n+1}\Sigma''$ sind miteinander logisch unverträglich. Also kann das spezielle Additionstheorem angewandt werden, und man erhält:

$$
\begin{aligned}
{}_{n+1}c\,({}_{n}\Sigma, T) &= {}_{n+1}c\,({}_{n+1}\Sigma', T) + {}_{n+1}c\,({}_{n+1}\Sigma'', T) \\
&= \frac{1}{n+2} + \frac{1}{n+2} \\
&= \frac{m! \cdot (n+1-m)!}{(n+2) \cdot (n+1)!} + \frac{(m+1)! \cdot (n+1-m-1)!}{(n+2) \cdot (n+1)!} \\
&= \frac{m! \cdot (n-m)! \cdot [(n-m+1) + (m+1)]}{(n+1)! \cdot (n+2)} \\
&= \frac{m! \cdot (n-m)!}{(n+1)!}.
\end{aligned}
$$

$_{n}\Sigma$ erhält also durch $_{n+1}c$ den gleichen Aprioriwert wie durch $_{n}c$. Es kann deshalb für die obigen Gleichungen (die den Erwartungswert bzw. den Schätzwert der relativen Häufigkeit von F in B_n ausdrückt) die Limeskonvention für Gegenstandsausdrücke hinsichtlich einer solchen passenden Folge von c-Funktionen angewandt werden. Da die einzelnen Funktionen $_{n}c$ ferner symmetrisch sind, gilt deshalb mit Theorem 23 (b) (β) und Korollar 12 (a):

$$
\begin{aligned}
c\,(\gamma \varepsilon \Pi_1, \Phi_k) &= \mathrm{srh}_c\,(F, B, \Phi_k) \\
&= \lim_{n \to \infty} \mathrm{srh}_c\,(F, B_n, \Phi_k) \\
&= \lim_{n \to \infty} \frac{\sum_{i=0}^{n} \left[\frac{i}{n} \cdot c\,(\Phi_k, \Delta^{(i)}) \cdot \frac{1}{n+1} \right]}{\sum_{i=0}^{n} \left[c\,(\Phi_k, \Delta^{(i)}) \cdot \frac{1}{n+1} \right]} \\
&= \lim_{n \to \infty} \frac{\sum_{i=0}^{n} \left[\frac{i}{n} \cdot \left(\frac{i}{n}\right)^{k_1} \cdot \left(1 - \frac{i}{n}\right)^{k-k_1} \cdot \frac{1}{n+1} \right]}{\sum_{i=0}^{n} \left[\left(\frac{i}{n}\right)^{k_1} \cdot \left(1 - \frac{i}{n}\right)^{k-k_1} \cdot \frac{1}{n+1} \right]}
\end{aligned}
$$

$$= \frac{\lim\limits_{n\to\infty} \sum\limits_{i=0}^{n} \left[\left(\frac{i}{n}\right)^{k_1+1} \cdot \left(1-\frac{i}{n}\right)^{k-k_1} \cdot \frac{1}{n+1} \right]}{\lim\limits_{n\to\infty} \sum\limits_{i=0}^{n} \left[\left(\frac{i}{n}\right)^{k_1} \cdot \left(1-\frac{i}{n}\right)^{k-k_1} \cdot \frac{1}{n+1} \right]}.$$

Es sei $r_i = i/n$ und somit (für große n) $\Delta r_i = 1/n \approx 1/(n+1)$; dann gilt:

$$c(\gamma \varepsilon \Pi_1, \Phi_k) = \frac{\lim\limits_{n\to\infty} \sum\limits_{i=0}^{n} \left[r_i^{k_1+1} \cdot (1-r_i)^{k-k_1} \cdot \Delta r_i \right]}{\lim\limits_{n\to\infty} \sum\limits_{i=0}^{n} \left[r_i^{k_1} \cdot (1-r_i)^{k-k_1} \cdot \Delta r_i \right]}$$

$$= \frac{\int_0^1 r^{k_1+1} \cdot (1-r)^{k-k_1}\, dr}{\int_0^1 r^{k_1} \cdot (1-r)^{k-k_1}\, dr}.$$

Diese beiden Integrale sind Eulersche *B*-Funktionen, die unter Verwendung von Γ-Funktionen auf Grund der Tatsache, daß k und k_1 natürliche Zahlen sind, folgenden einfachen Gleichungen genügen:[9]

$$B(i+1, j+1) = \int_0^1 r^i \cdot (1-r)^j\, dr$$

$$B(i+1, j+1) = \frac{\Gamma(i+1) \cdot \Gamma(j+1)}{\Gamma(i+j+2)}$$

$\Gamma(i+1) = i!$ für alle natürlichen Zahlen i.
Unter Verwendung dieser Gleichungen gilt daher:

$$c(\gamma \varepsilon \Pi_1, \Phi_k) = \frac{\int_0^1 r^{k_1+1} \cdot (1-r)^{k-k_1}\, dr}{\int_0^1 r^{k_1} \cdot (1-r)^{k-k_1}\, dr}$$

---

[9] Vgl. W. I. Smirnow, Lehrgang der höheren Mathematik, [4]1963, Teil III, 2, S. 219—225. — Die Buchstaben Γ und *B* stehen hier natürlich nicht für Sätze der Objektsprache, sondern für mathematische Funktionen (und analog Δ für ein Zahlenintervall).

$$= \frac{B(k_1+2, k-k_1+1)}{B(k_1+1, k-k_1+1)}$$

$$= \frac{\Gamma(k_1+2)\cdot\Gamma(k-k_1+1)\cdot\Gamma(k_1+1+k-k_1+1)}{\Gamma(k_1+2+k-k_1+1)\cdot\Gamma(k_1+1)\cdot\Gamma(k-k_1+1)}$$

$$= \frac{\Gamma(k_1+2)\cdot\Gamma(k+2)}{\Gamma(k+3)\cdot\Gamma(k_1+1)}$$

$$= \frac{(k_1+1)!\cdot(k+1)!}{(k+2)!\cdot k_1!}$$

$$= \frac{(k_1+1)\cdot k_1!\cdot(k+1)!}{(k+2)\cdot(k+1)!\cdot k_1!}$$

$$= \frac{k_1+1}{k+2}.$$

Da nach Voraussetzung $q = 2$ ist, gilt somit:

$$c(\gamma \varepsilon \Pi_1, \Phi_k) = \frac{k_1+1}{k+2}$$

$$= \frac{k_1+(1/q)\cdot 2}{k+2}.$$

Folglich ist c mit c^2 und L mit 2 identisch. Damit ist der Lehrsatz bewiesen.

Unter der Voraussetzung, daß $p = 1$ und daß somit $q = 2$, ist also Carnaps Funktion c^* mit der Laplace-Verteilung identisch, mit einer Funktion c also, die eine reguläre Wahrscheinlichkeitsfunktion und symmetrisch bezüglich der Gegenstandsausdrücke ist und die allen relativen Häufigkeiten (allen statistischen Verteilungen) die gleichen Aprioriwerte zuordnet. Für $p = 1$ kann c^* damit als eine Explikation des Laplaceschen Wahrscheinlichkeitsbegriffs und für $p > 1$ als eine sinnvolle Verallgemeinerung desselben angesehen werden.
Wenn man (immer noch vorausgesetzt, daß $p = 1$) von Verteilungen ausgeht, die keine Laplace-Verteilung sind, von Funktionen c also, die reguläre Wahrscheinlichkeitsfunktionen und bezüglich der Gegenstandsausdrücke symmetrisch sind, die jedoch den statistischen

Verteilungen nicht gleiche Aprioriwerte zuordnen, so werden die dem Theorem 24 entsprechenden Beweise erheblich komplizierter. Es empfiehlt sich zum Zwecke der Integration dabei, eine Funktion f (i, n) anzugeben, so daß gilt: $c(\Delta^{(i)}, T) = f(i, n) \cdot 1/(n + 1)$.
Es kann dann gezeigt werden, daß die Bedingung (II) (b) gleichwertig damit ist, daß das Integral punktsymmetrisch zur Stelle $\langle 1/2, 1/2 \rangle$ im Koordinatensystem ist, d. h. daß die erste Ableitung dieser Funktion symmetrisch zur Achse $x = 1/2$ ist, und daß (III') besagt, daß die zweite Ableitung des Integrals (abgesehen von den Stellen $x = 0$ und $x = 1$) nur eine Nullstelle besitzt. Die Beweise dieser Behauptungen sind zum Teil kompliziert, doch sind die dabei auftretenden Schwierigkeiten rein *mathematischer* Natur und keine *philosophischen* Probleme. Während der *Philosoph* an diesem Punkt seine Arbeit (nämlich die *Grundlegung* des induktiven Schließens) beenden kann, fängt die des *Mathematikers* erst an: ihm ist hier ein weites und fruchtbares Tätigkeitsgebiet erschlossen. Selbstverständlich bleiben dem Philosophen *im Zusammenhang mit* der induktiven Logik noch genügend weitere Probleme zu lösen, von denen einige in den folgenden Abschnitten angedeutet werden.

3. Die mathematische Wahrscheinlichkeitslehre

Von *der* mathematischen Wahrscheinlichkeitslehre zu sprechen und damit vorauszusetzen, daß es genau eine mathematische Theorie gibt, die den Wahrscheinlichkeitsbegriff zu explizieren sucht, ist nur in folgender Hinsicht korrekt: So gut wie alle diesbezüglichen Werke gehen von den sogenannten Kolmogoroff-Axiomen aus und deduzieren daraus Lehrsätze, oder aber sie gehen von anderen Postulaten aus und gewinnen jene Axiome dann als Theoreme. Wenn man von der inhaltlichen Deutung jener Grundsätze absieht, herrscht also bei den Autoren, die derartige Lehrbücher verfassen, heute eine gewisse Einmütigkeit.
Kolmogoroff wollte in seinem Standardwerk[10] jene Axiome ebenfalls als uninterpretierte Sätze verstanden wissen, und insofern dürfen

[10] Vgl. Kolmogoroff [1].

sich auch andere Mathematiker berechtigt fühlen, auf detaillierte Überlegungen über deren Deutung zu verzichten. Damit ist allerdings nicht gesagt, daß die Mehrzahl der Autoren keine Deutungsversuche vornimmt. Kolmogoroff etwa hat ohne Zweifel an eine Interpretation seiner Axiome im Misesschen Sinne gedacht, und wenn dies auch von der großen Mehrzahl der heutigen Statistiker abgelehnt wird, so lassen deren außersystematischen Bemerkungen doch oft auf Interpretationen schließen, die der von Kolmogoroff intendierten in einem schwer präzisierbaren Sinn ähnlich sind. Eine Einigkeit hierüber liegt allerdings nicht vor, und so kann, wenn man auch die Interpretation jener Grundsätze mitberücksichtigt, von *der* mathematischen Wahrscheinlichkeitslehre keine Rede sein.

Den Ausgangspunkt der mathematischen Wahrscheinlichkeitslehre bildet eine Gesamtheit von Experimenten (manchmal auch „Zufallsexperimente" genannt). Bei einer genaueren Darstellung müßte man dabei eigentlich, was nicht oft geschieht, dreierlei unterscheiden:

(a) Die einzelnen konkreten Experimente, (b) die Gesamtheit aller solchen Experimente einer bestimmten Art, d. h. die Gesamtheit der Realisierungen einer Vorschrift zur Erzeugung solcher Experimente, und (c) die Vorschrift (das Verfahren) zur Erzeugung solcher Experimente. An ein solches Verfahren werden drei Forderungen gestellt:

(1) Es ist in allen relevanten Einzelheiten genau charakterisiert, d. h. die Vorschrift enthält alle relevanten Einzelheiten zur Realisierung der Experimente [Problem: was heißt „alle relevanten Einzelheiten"?].

(2) Bei jeder Realisierung des Verfahrens wird ein und nur ein präzise beschriebenes Ergebnis gewonnen, wobei die Gesamtheit der Ergebnisse von beliebiger Mächtigkeit sein kann (in allen praktischen Fällen wird diese Gesamtheit endlich sein).

(3) Das Verfahren ist beliebig wiederholbar [dies ist im Grunde genommen mit der Voraussetzung identisch, daß die Gesamtheit der Experimente nicht endlich ist].

Meist verlangt man von derartigen Verfahren noch folgendes:

(4) Die einzelnen Realisierungen dieser Vorschrift sind voneinander unabhängig, d. h. das Experiment ist unabhängig wiederholbar [Problem: was heißt das genau?].

Manchmal, vor allem, wenn eine Interpretation des Wahrscheinlich-

keitsbegriffs im Sinne eines Grenzwertes von relativen Häufigkeiten intendiert wird, fordert man noch:

(5) Wenn das Experiment genügend oft wiederholt wird, streben die relativen Häufigkeiten der einzelnen Ergebnisse in der Gesamtheit der gemachten Experimente einem Grenzwert zu [Problem: was heißt „genügend oft"?].

Die Mathematiker ahnen instinktiv, daß sie mit diesen Bestimmungen niemandem, der es nicht schon weiß, klarmachen können, was im folgenden unter den Begriffen „Experiment" und „Ergebnis" zu verstehen ist. Sie verdeutlichen deshalb ihre abstrakten (besser: ihre vagen) Formulierungen an Beispielen, sie geben demnach sogenannte *Definitionen durch Hinweis* an (die im strengen Sinn natürlich *keine* Definitionen sind). Die üblichen Beispiele lauten:

(1) Gegeben ist ein (endlicher oder abzählbar unendlicher) Bereich von Körpern, die alle kleiner als 1 m sind. Das Experiment besteht darin, daß mittels eines 1 m langen Maßstabs die Länge dieser einzelnen Körper gemessen wird. Der Maßstab habe etwa Millimetereinteilung. An Ergebnissen erhält man dann z. B., daß der Körper a_1 nicht kürzer als 372 mm und kürzer als 373 mm ist, daß a_2 nicht kürzer als 2 mm und kürzer als 3 mm ist usw., allgemein: daß der Körper x im halboffenen Intervall [i mm, i + 1 mm) liegt, wobei i eine natürliche Zahl zwischen 0 und 999 ist.

(2) Gegeben sei eine Gesamtheit n von Experimenten, von denen jedes darin besteht, daß sich ein Student einer gewissen Prüfung (etwa der Magisterprüfung in Philosophie) unterzieht. Als mögliche Ergebnisse hat man dann die einzelnen Noten, mit denen der Student die Prüfung abschließt.

(3) Die Gesamtheit der Experimente ist die Klasse der mit einem bestimmten Würfel gemachten Würfe. Jeder dieser Würfe kann als Ergebnis haben, daß bei ihm die Ziffer „1", daß bei ihm die Ziffer „2", ..., oder daß bei ihm die Ziffer „6" oben liegt (analog im Fall der Münze).

(4) Gegeben ist eine Urne mit Kugeln, die alle einfarbig sind und deren Farben die einzelnen Spektralfarben sind. Das Experiment besteht darin, Kugeln aus der Urne zu ziehen und sie auf ihre Farbe hin zu beobachten. An Ergebnissen erhält man dann, daß der erste Zug die Farbe *rot* ergibt (d. h. daß die Kugel a_1 rot ist), usw.

(5) Gegeben ist eine Urne mit verschiedenfarbigen Kugeln, wobei wiederum nur Spektralfarben in Frage kommen. Das Experiment besteht darin, Kugeln aus dieser Urne zu ziehen und auf ihre Einzelfarben hin zu untersuchen. An Ergebnissen erhält man diesmal, daß der erste Zug die Farben *rot und orange und ... und violett* ergibt, usw.

Streng genommen müßte man stets zwischen Gegenständen und den Experimenten mit diesen Dingen unterscheiden. Diese Unterscheidung ist jedoch für die Zwecke der Wahrscheinlichkeitsrechnung unergiebig, d. h. sie äußert sich nicht in verschiedenen Wahrscheinlichkeitswerten, weshalb der Mathematiker mit einem sicheren Instinkt für das Richtige Experimente mit den einzelnen Gegenständen, an denen sie vorgenommen werden, identifiziert.[11] Er benützt den Begriff „Experiment" daher in einem weiten, allumfassenden Sinn: Nicht nur Menschen oder Maschinen können Experimentatoren sein, sondern auch die Natur kommt hierbei in Frage: Das Experiment besteht dann einfach darin, zu beobachten, welche der zugrundegelegten Eigenschaften das Objekt besitzt; folglich ist jeder Gegenstand im Prinzip auch Gegenstand eines Experiments.

Der Mathematiker geht normalerweise nicht von einem einzigen Versuch, sondern von einer Folge von Experimenten (von einer *Versuchskette*) aus, und zwar von einer endlichen Versuchskette: $\overrightarrow{a_1, \ldots, a_n}$ oder von einer abzählbar unendlichen: $\overrightarrow{a_1, a_2, \ldots}$. Diese Folge identifiziert er mit einem Vektor $\mathfrak{a}$. Jedes Experiment dieser Versuchskette hat eines der möglichen Ergebnisse (z. B. beim Münzwurf: *beim Wurf k liegt der Kopf oben* und *beim Wurf k liegt das Wappen oben*, kurz: *Kopf* und *Wappen*), und die ganze Versuchskette (der Vektor) hat dann ein daraus zusammengesetztes Gesamtergebnis (im endlichen Fall etwa: *beim 1. Wurf liegt der Kopf oben und beim 2. Wurf liegt das Wappen oben und ... und beim n-ten Wurf liegt das Wappen oben*, kurz: *Kopf—Wappen—...—Wappen*, und analog im abzählbar unendlichen Fall). Die Begriffe „Ergebnis", „Ereignis", „Mengenkörper" und „Wahrscheinlichkeitsfeld" können dann für jedes Experiment extra

[11] Genaugenommen wird hier ein Prinzip der Identität des Ununterscheidbaren angewandt: Hinsichtlich der Fragen der Wahrscheinlichkeitslehre sind Gegenstand und Untersuchung des Gegenstands (Experiment) in allen relevanten Eigenschaften gleich; folglich ist der Gegenstand mit dem entsprechenden Experiment identisch.

(mittels eines fest, aber beliebig angenommenen Experiments a_k) oder auch für die gesamte Versuchskette α festgelegt werden.
Für ein einzelnes Experiment a_k werden die Begriffe „Ergebnis" und „Ereignis" folgendermaßen bestimmt: Ein mögliches Ergebnis von a_k ist der Sachverhalt, daß a_k das Ergebnis F^*_i hat, kurz: daß $F^*_i|a_k$ (noch kürzer, wenn a_k vorgegeben ist: F^*_i). Sodann wird der zu einem derartigen Experiment a_k gehörende Ereigniskörper entwickelt: Es wird zunächst die Gesamtheit M_k der zu diesem (fest, aber beliebig vorgegebenen) Experiment gehörenden Ergebnisse gebildet und hierauf festgesetzt, daß A genau dann ein (zu diesem Experiment gehörendes) Ereignis ist, wenn A eine Teilklasse von M_k ist; atomare Ereignisse sind jene Teilmengen von M_k, die als einziges Element ein Ergebnis enthalten. M_k ist dann das notwendige Ereignis und die leere Klasse (das Komplement von M_k) das unmögliche Ereignis.
Die Ergebnisse der Versuchskette α bilden den Vektor

$$\overrightarrow{F^*_{i_1}|a_1, \ldots, F^*_{i_n}|a_n} \quad \text{bzw.} \quad \overrightarrow{F^*_{i_1}|a_1, F^*_{i_2}|a_2, \ldots}.$$

M ist die Gesamtheit aller dieser Ergebnisse, und K ist ein Ereignis genau dann, wenn K Teilklasse von M ist. Ein atomares Ereignis (hinsichtlich dieser Versuchskette, dieser Koppelung von Experimenten) ist ein Ereignis, das als Element genau ein Ergebnis (von α) enthält. Wenn es q Ergebnisse zu jedem Experiment a_k gibt, so ist die Anzahl der Ergebnisse des Vektors α offenbar n^q (mit $1 \leqslant q \leqslant \infty$).
Im folgenden wird, wenn nicht ausdrücklich anders erwähnt, stets angenommen, daß eine Versuchskette α, also eine Koppelung von einzelnen Experimenten, vorliegt.
Allgemein wird das Tripel $\langle M, \mathfrak{K}, P\rangle$ ein Wahrscheinlichkeitsfeld (und die mittels M erzeugte Klasse $\mathfrak{K}$ ein Mengenkörper für das Wahrscheinlichkeitsmaß P) genannt, wenn (und nur wenn) folgende Bedingungen erfüllt sind:[12]

(1) $\mathfrak{K}$ enthält M als Element.
(2) Zu jedem Element K enthält $\mathfrak{K}$ auch dessen Komplement $\llcorner K$, d. h. die Klasse der Ergebnisse aus M, die nicht in K enthalten sind.
(3) Zu je zwei Elementen K_1 und K_2 aus $\mathfrak{K}$ enthält $\mathfrak{K}$ auch deren Vereinigung $K_1 \cup K_2$, d. h. die Klasse der Ergebnisse aus M, die in

[12] Vgl. z. B. Kolmogoroff [1], S. 2f.; die folgenden Bedingungen decken sich inhaltlich mit den Axiomen Kolmogoroffs.

K_1 oder (auch) in K_2 enthalten sind, und folglich auch deren Durchschnitt $K_1 \cap K_2$, d. h. die Klasse der Elemente aus M, die sowohl in K_1 als auch in K_2 enthalten sind (wenn $\mathfrak{K}$ unendlich ist, wird $\mathfrak{K}$ ein σ-Körper genannt, wenn zusätzlich für jede abzählbar unendliche Folge $K_1, K_2, \ldots$ von Elementen von $\mathfrak{K}$ auch deren Vereinigung $\bigcup_{i=1}^{\infty} K_i$ und somit auch deren Durchschnitt $\bigcap_{i=1}^{\infty} K_i$ Element von $\mathfrak{K}$ ist).

(4) Die auf $\mathfrak{K}$ definierte Funktion P erfüllt folgende Bedingungen:
(a) $0 \leqslant P(K) \leqslant 1$ für jedes Element K aus $\mathfrak{K}$,
(b) $P(M) = 1$,
(c) $P(K_1 \cup K_2) = P(K_1) + P(K_2)$ für zwei elementfremde Klassen K_1 und K_2 aus $\mathfrak{K}$ (bzw. für σ-Körper $\mathfrak{K}$: $P(\bigcup_{i=0}^{\infty} K_i) = \sum_{i=0}^{\infty} P(K_i)$ für jede abzählbar unendliche Folge von Ereignissen $K_1, K_2, \ldots$ aus $\mathfrak{K}$, die paarweise elementfremd sind).

Die Gleichung „$P(K) = r$" wird dann üblicherweise gelesen als „die Wahrscheinlichkeit des Ereignisses K ist r", d. h. der Ausdruck „P" wird als eine (absolute, unbedingte) Wahrscheinlichkeitsfunktion interpretiert. Ein zweistelliger Funktionsausdruck „P", der dann die entsprechende bedingte Wahrscheinlichkeitsfunktion bezeichnet, wird folgendermaßen definiert:[13] $P(K_2, K_1) = P(K_2 \cap K_1)/P(K_1)$, falls $P(K_1) \neq 0$. Es wird dann „$P(K_2, K_1) = r$" gelesen als „die Wahrscheinlichkeit des Ereignisses K_2 auf Grund des Ereignisses K_1 ist r". Offensichtlich gilt hierbei: $P(K) = P(K, M)$.

Es fällt nicht schwer, für ein Wahrscheinlichkeitsfeld $\langle M, \mathfrak{K}, P \rangle$ Lehrsätze zu beweisen, die hinsichtlich des zweistelligen Funktionsausdrucks „P" von gleicher Struktur sind wie die Bedingungen (1) bis (5) von D 4; genauso leicht ist es, einzusehen, daß sich Bedingung (6) nicht herleiten läßt. Daß die Mathematiker keine mit ihr gleichwertige Bedingung aufstellen, ist einerseits leicht verständlich, da diese ja auf endliche Gegenstandsbereiche beschränkt ist und von unendlichen Bereichen nicht gefordert werden darf und da die Mathematiker sich von Anfang an auf unendliche Bereiche konzentrieren (außerdem wollen sie sich meist auch der Häufigkeitsinterpretation

---

[13] Vgl. Kolmogoroff [1], S. 6.

nicht verschließen und sind dann sogar genötigt, auf eine derartige Bedingung zu verzichten). Andererseits ist das Fehlen einer mit (6) gleichwertigen Bedingung jedoch erstaunlich, da die mathematische Wahrscheinlichkeitslehre doch überwiegend in vielleicht sehr großem, aber immerhin endlichen Bereichen angewandt wird und da (6) ja mit der Bedingung der *strengen* Kohärenz gleichwertig ist. Es scheint, daß der Mathematiker bereit ist, kohärente, aber nicht streng kohärente Funktionen P zu akzeptieren, die (bei Wetten oder sonstwo) zwar einen Gewinn, nicht jedoch einen Verlust ausschließen. Den Einwand, daß die Bedingungen (1) bis (4) der Definition des Begriffs „Wahrscheinlichkeitsfeld" von Mathematikern und Statistikern der Praxis mit großem Geschick angewandt werden, und zwar so, daß derartige Fälle nicht eintreten, lasse ich nicht gelten: An den intuitiven Fähigkeiten dieser Forscher habe ich nie gezweifelt; als Logiker darf ich jedoch verlangen, daß sie etwas, was sie grundsätzlich nicht tun, auch in ihren Bestimmungen ausdrücklich ausschließen.

Man könnte versuchen, dieses Versäumnis der Mathematiker damit zu entschuldigen, daß die Definition des Begriffs des Wahrscheinlichkeitsfeldes den Begriff „P" hinreichend festlege, und zwar so, daß er derartige Wettsysteme ausschließe. Dieser Ausweg ist jedoch nicht gangbar. Ein Blick auf das Theorem 13 wird den Leser davon überzeugen, daß dieses auch hier gültig ist, ja daß hier die Zahl r sogar die Werte 0 und 1 annehmen kann. Auch der Mathematiker steht also vor dem Problem, die Grundbedingungen durch weitere Forderungen zu ergänzen.

Eine zusätzliche Bedingung, die der Statistiker meist stillschweigend und der Mathematiker an allen wichtigen Stellen explizit voraussetzt, ist die Bedingung, daß ein Experiment unabhängig wiederholbar sein muß (diese Bedingung ist, wie später noch gezeigt wird, mit (II) (a) gleichwertig). Diese Forderung ist zwar schon zu Beginn dieses Abschnitts formuliert worden, doch hat es sich hier nur um heuristische und außersystematische Bemerkungen gehandelt; in die Definition des Begriffs des Wahrscheinlichkeitsfeldes ist sie nicht eingegangen. Wenn der Statistiker sie nicht voraussetzen darf, hat er kein Recht, von den vielen Einzelfällen zu relativen Häufigkeiten überzugehen, und der Mathematiker kann ohne sie keinen interessanten Lehrsatz der Wahrscheinlichkeitslehre beweisen. Über ihre Rechtfertigung

machen sich beide keine großen Sorgen: Sie wenden die durch sie gewonnenen Lehrsätze eben nur in solchen Fällen an, in denen (empirisch) gesichert ist, daß diese Voraussetzung tatsächlich erfüllt ist.[14] Allein auch diese Zusatzbedingung führt noch zu keinen interessanten Wahrscheinlichkeitswerten. Solche gewinnt der Mathematiker vielmehr, indem er von gewissen Aprioribewertungen ausgeht, die oft mit der Laplace-Verteilung übereinstimmen, gelegentlich jedoch eine andere Form haben, und der Statistiker, indem er seinen Berechnungen die tatsächlichen relativen Häufigkeiten der einzelnen Eigenschaften in der Grundgesamtheit zugrundelegt. Fragt man den Mathematiker, wie er diese Aprioribewertungen rechtfertigt, so wird dieser antworten, daß diese weitgehend irrational vorgenommen werden und sich einer rationalen Kontrolle entziehen; rational ist vielmehr erst der Übergang von einer gegebenen Bewertung zu einer neuen, wenn zusätzliche Erfahrungen vorliegen.[15] Mit anderen Worten: Auch auf die Frage, von welchen Verteilungen und Verteilungsfunktionen man ausgehen soll (mittels derer man dann über den Erwartung wert Wahrscheinlichkeitsgrade berechnen kann, vgl. Theorem 24)[16] und

---

[14] Vgl. z. B. Basler [1], S. 39 und 42f.

[15] Vgl. Richter [2], S. 67ff. Es braucht nach allem wohl nicht ausdrücklich betont zu werden, daß der Wahrscheinlichkeitsbegriff „p" der direkten Theorie Richters eine axiomatische Charakterisierung des Begriffs der relativen Häufigkeit auf lange Sicht darstellen soll, während der Zentralbegriff „χ" („Chance") seiner indirekten Theorie, der sich auf derartige Aprioribewertungen stützt, zur Schätzung von objektiven Wahrscheinlichkeiten verwendet wird; diesem Begriff der Chance entspricht der Begriff der logischen Wahrscheinlichkeit Carnaps, vgl. Humburg [1], S. 27—37. Nun läßt sich jedoch der objektive Wahrscheinlichkeitsbegriff als Hilfsbegriff in einem System der subjektiven und damit auch der logischen Wahrscheinlichkeit einführen, wie Finetti in [1] andeutungsweise gezeigt hat; außerdem ist es fraglich, ob wir überhaupt in irgendeinem Einzelfall berechtigt sind, aus unseren Erfahrungen auf den Limes einer relativen Häufigkeit zu schließen, da uns ja nur ein enger räumlicher und zeitlicher Ausschnitt des Universums bekannt ist, der kaum repräsentativ sein dürfte (vgl. auch das in Kapitel I., 4. Abschnitt Gesagte). Es wäre deshalb wohl zweckmäßig gewesen, wenn Richter sein Hauptgewicht auf den Begriff der Chance gelegt hätte.

[16] Aus der Unkenntnis dessen, daß die Carnapschen Wahrscheinlichkeiten mit solchen Erwartungswerten identisch sind, hat Vetter in [1] eine einseitige Darstellung des Carnapschen Ansatzes gegeben und diesen dann lächerlich gemacht. Dabei hat er Carnaps Ansatz zur Grundlegung einer Theorie der logischen Wahrscheinlichkeit an den entscheidenden Stellen mißverstanden. Es ist ja Carnap (wie er immer wieder betont) gerade *nicht* darum gegangen, eine Explikation dessen zu liefern, was die Statistiker heutzutage üblicherweise „Wahrschein-

welche Voraussetzungen wir mit derartigen Verteilungen machen, gibt der Mathematiker keine Auskunft.

Der Statistiker andererseits erschließt mit irgendwelchen Methoden aus gewissen Erfahrungsaussagen Behauptungen, die besagen, daß die relative Häufigkeit einer Eigenschaft F in einer Bezugsklasse G (auf lange Sicht) so und so groß ist, und rechnet dann mit den sich aus der Voraussetzung, daß es sich um unabhängig wiederholbare Experimente handelt und ⟨M, $\mathfrak{R}$, P⟩ ein Wahrscheinlichkeitsfeld ist, ergebenden Theoremen weiter.[17]

Mich interessiert hier nicht die Frage, *auf welchem Weg* er von seinen Erfahrungen (die im allgemeinen Stichproben und nach Möglichkeit durch Zufall ausgewählte, vielleicht sogar repräsentative Stichproben aus der Grundgesamtheit sind) zu seinen statistischen Hypothesen gelangt; ich will vielmehr wissen, wie groß die *Sicherheit* dieser von ihm gefundenen Hypothesen hinsichtlich jenes Erfahrungsdatums ist, welchen Grad an *logischer Wahrscheinlichkeit* ihr also durch dieses Datum zukommt. Ich interessieren mich, mit anderen Worten, nicht für den *Entdeckungs-*, sondern für den *Begründungszusammenhang* (aus der Tatsache, daß der Statistiker vielleicht interessante Entdeckungsmethoden besitzt, folgt noch nichts darüber, wie gut oder schlecht begründet die statistische Hypothese durch das Erfahrungsdatum ist: Mit diesen Methoden kann er vielleicht gelegentlich oder sogar relativ häufig die richtige Hypothese treffen, und trotzdem kann diese durch das Datum schlecht begründet sein).

Zur Abschätzung der Sicherheit wird man eine streng kohärente Methode und damit eine reguläre Bestätigungsmethode benützen müssen. Diese Methode wird der Bedingung (II) (a) genügen, gleich-

---

lichkeit" nennen (und womit in seiner Terminologie *relative Häufigkeit*$_a$ *relative Häufigkeit auf lange Sicht* oder etwas Verwandtes gemeint ist); vielmehr deckt sich sein Begriff der logischen Wahrscheinlichkeit teilweise oder ganz mit dem, was in der mathematischen Wahrscheinlichkeitslehre meist „Sicherheit" genannt wird, und wofür etwa Richter den Begriff „Chance" verwendet. — Im übrigen fällt auf, daß Vetter gelegentlich Argumente gegen Carnap vorbringt, die von Carnap selbst stammen, ohne daß er darauf hinweist, wie etwa, daß bei großem q die Funktion c^* kaum von c^+ zu unterscheiden ist (vgl. Vetter [1], S. 68f., und Vetter [2], S. 62). — Mit der Polemik Vetters hat sich bis jetzt lediglich Hintikka in [5] kritisch auseinandergesetzt.

[17] In der Sprache der induktiven Logik sind diese zum Teil im 1. Abschnitt dieses Kapitels dargestellt worden.

gültig, ob der Statistiker die Bedingung der unabhängigen Wiederholbarkeit eines Experiments mit Recht vorausgesetzt hat oder nicht: Wenn er es mit Recht vorausgesetzt hat, ist (II) (a) ohne Einschränkungen anzuwenden; ansonsten hat man jenen Zusammenhang zwischen den einzelnen Experimenten, die den Grund für die Nichtwiederholbarkeit darstellen, zu formulieren und sie der Aussage über die Beobachtungen konjunktiv hinzuzufügen. Aus den gleichen Erwägungen heraus wird man verlangen, daß diese Methode bei *mehreren* Grundeigenschaften (Grundprädikaten) die Bedingung (III) erfüllt; liegt nur *eine* Grundeigenschaft vor, so wird man die Erfüllung von (III′) verlangen: Man kann ja, wie schon erwähnt, jedes beliebige Paar von Eigenschaften, wie z. B. *intelligent* und *nicht intelligent*, durch Hinzufügung eines beliebigen anderen Paares von Eigenschaften, das mit diesem gar nichts zu tun zu haben braucht, wie etwa *reich* und *nicht reich*, aufteilen; dann erhält man die Eigenschaften *intelligent und reich*, *intelligent und nicht reich*, *nicht intelligent und reich* und *nicht intelligent und nicht reich*, für die die Gültigkeit von (III) und somit (III′) vorausgesetzt wird. Dann gilt (III′) jedoch auch hinsichtlich des Paares von Eigenschaften (*intelligent und reich*) *oder* (*intelligent und nicht reich*) und (*nicht intelligent und reich*) *oder* (*nicht intelligent und nicht reich*), d. h. hinsichtlich *intelligent* und *nicht intelligent*. Wenn man ferner keine hinreichenden Gründe dafür hat, eine der V-Eigenschaften als a priori weiter anzusehen als die anderen, wird man verlangen, daß jene Methode auch der Bedingung (II) (b) genügt, und man wird im anderen Fall versuchen, (II) (b) durch eine Bedingung zu ersetzen, die den wirklichen Weiten näherkommt (oder aber man behält (II) (b) bei und fügt dieses Zusatzwissen, auch wenn es kein Wissen im strengen Sinn ist, dem Erfahrungsdatum konjunktiv hinzu). Schließlich wird man aus den im V. Kapitel angeführten Gründen auch verlangen, daß man mit einer solchen Methode aus der Erfahrung lernen kann, daß sie also die Bedingung (IV) erfüllt. *Die Methode, mit der wir die Sicherheit statistischer Hypothesen hinsichtlich gewisser Stichproben bewerten, wird also eine adäquate induktive Methode im definierten Sinn sein.*

Hat man auf Grund einer Annahme über die Homogenität des Gegenstandsbereichs eine derartige Methode gewählt, so kann man genaue Grade der Sicherheit angeben. Liegen diese nahe bei 1 oder nahe bei 0, so kann man das Prinzip von Cournot anwenden, welches

besagt, daß die Hypothese dann (praktisch) sicher bzw. (praktisch) unmöglich ist. Genaugenommen macht man dann also einen Fehler, da man eine zwar kleine, aber von der 0 verschiedene reelle Zahl r mit der 0 und entsprechend 1 — r mit der 1 identifiziert; je kleiner r allerdings ist, um so kleiner wird dieser Fehler sein (in der Psychologie und in der Biologie fordert man üblicherweise, daß $r < 0{,}05$, und in Bereichen der Physik, daß $r < 0{,}005$ sein muß; eine Begründung für diese Festsetzungen, die von der Behauptung, diese Schranken hätten sich in der Praxis bewährt, verschieden ist, konnte ich nirgends finden). Wenn dann die Sicherheit einer Hypothese auf Grund eines Erfahrungsdatums unter Zuhilfenahme des Cournotschen Prinzips den Grad 1 erreicht, so kann man wegen Theorem 2 (a) den *modus ponens* anwenden und aus diesem Datum logisch auf die Hypothese schließen; man setzt hier also, wie in der deduktiven Logik, voraus, daß zusätzliche Prämissen an der Gültigkeit der Hypothese nichts mehr ändern können. Hat die Sicherheit hingegen den Grad 0, so folgt die Negation der Hypothese aus dem Datum.

Erweist sich die Hypothese später (d. h. auf Grund neuer Beobachtungen) aber als falsch (man denke etwa an die Wahlen zum westdeutschen Bundestag von 1965 und an die vorher formulierten Prognosen), so hat man neben der Möglichkeit, das Erfahrungsdatum anzuzweifeln, folgende Ausflüchte: (1) Man kann argumentieren, daß die Hypothese im Hinblick auf das Erfahrungsdatum zwar extrem wahrscheinlich ist, daß damit aber selbstverständlich ihre Falschheit nicht unverträglich ist; man bringt also die reelle Zahl r, die man vernachlässigt hat, wieder in Erinnerung. Der Statistiker kann aber auch zugeben, daß er größere Fehler gemacht hat und etwa ausführen: (2) Das Erfahrungsdatum (die Stichprobe, die statistische Erhebung) war hinsichtlich der gesamten Population nicht repräsentativ, so daß ein größeres Maß an Vorsicht, d. h. eine Methode c^L mit größerem L, am Platze gewesen wäre; durch c^L hätte die Hypothese im Hinblick auf jenes Erfahrungsdatum einen Wert 1 — s erhalten, der nicht nahe genug bei 1 gelegen hätte und daher nicht mit 1 identifiziert hätte werden können. Der Statistiker gesteht damit ferner zu, daß seine *Entdeckungsmethode* zu verbessern ist. Er kann aber auch prinzipiellere Erwägungen vorbringen und etwa sagen: (3) Es hat sich damit (vielleicht wieder einmal) gezeigt, daß die Schranke r zu groß gewählt

worden ist und durch eine kleinere (vielleicht durch r/10) ersetzt werden muß; hieraus ergibt sich dann als praktische Konsequenz, daß bei den nächsten statistischen Erhebungen größere Stichproben gewählt werden müssen. — Das Operieren mit solchen Schranken r scheint eine Kunstfertigkeit zu sein, die die Statistiker mit Geschick (manchmal allerdings auch mit wechselndem Geschick) handhaben und die sich einer exakten Analyse und damit auch einer Begründung weitgehend oder ganz entzieht.

4. Über den mathematischen und den logischen Ansatz

Es wird nun die Sprechweise des Mathematikers mit der des Logikers verglichen. Dabei ist zu berücksichtigen, daß ein Vergleich nicht genauer sein kann als das Ungenaueste seiner Glieder. Haben alle Glieder den Genauigkeitsgrad, der heute von der modernen Logik bestimmt wird, so nimmt der Vergleich die Form eines Theorems an; ansonsten muß man sich mit mehr oder weniger überzeugenden Plausibilitätsbetrachtungen begnügen. Das letztere liegt hier vor. Zwar lassen die Kolmogoroff-Axiome, die die Bedingungen des Begriffs „Wahrscheinlichkeitsfeld" ausmachen, an Genauigkeit und Bestimmtheit nichts zu wünschen übrig, aber als Axiome stellen sie zunächst einen uninterpretierten Kalkül dar, der der verschiedensten Deutungen fähig ist, auch solcher, die auf keinen Fall intendiert sind; die sogenannte „Normalinterpretation", die die Mathematiker im Auge haben, wird von ihnen jedoch meist nur mit einigen vagen Andeutungen umrissen. Es stehen sich also diese sogenannte „Normalinterpretation" des Mathematikers und die vom Logiker benützte Objektsprache einschließlich deren semantische Interpretation gegenüber sowie die Kolmogoroff-Axiome zuzüglich einiger weiterer Bestimmungen des mathematischen Wahrscheinlichkeitsbegriffs und die Adäquatheitsbedingungen der induktiven Logik.
Zur Verdeutlichung des Zusammenhangs der Normalinterpretation der Kolmogoroff-Axiome mit der Objektsprache der induktiven Logik kann als Beispiel der Begriff der logischen Folgerung aus der deduktiven Logik herangezogen werden. Logische Folgerungen sind

ja Spezialfälle von induktiven Argumenten, nämlich solche, die in endlichen Gegenstandsbereichen den c-Wert 1 ergeben. Im alltäglichen Argumentieren benützen wir den Begriff der logischen Folgerung so, daß er eine Relation zwischen Sachverhalten ist oder zumindest zu sein scheint: daraus, daß dies und jenes gilt, wird ja gefolgert, daß auch das und das der Fall ist.[18] Ein erster Schritt zur Angabe einer umkehrbar eindeutigen Abbildung in der Objektsprache auf jene Normalinterpretation besteht also darin, eine derartige Beziehung zwischen Sätzen und Sachverhalten anzugeben. In einigen einfachen Fällen besteht diese Abbildung darin, die Anführungszeichen wegzulassen, in denen die Sätze der Objektsprache stehen. Nun kann aber im allgemeinen ein Ereignis durch verschiedene Sätze beschrieben werden; das sichere (oder notwendige) Ereignis etwa kann sowohl durch einen Satz $\Phi \vee \neg \Phi$ als auch durch $\neg (\Phi \wedge \neg \Phi)$ in der Objektsprache wiedergegeben werden. Es ist daher zweckmäßig, die Sätze der Objektsprache zunächst in sogenannte Äquivalenzklassen zusammenzufassen, und zwar so, daß für zwei Sätze Φ und Ψ einer Äquivalenzklasse gilt, daß $\Phi \leftrightarrow \Psi$ aus der Gesamtheit der Axiome der Sprache logisch folgt (und damit, falls diese Klasse leer ist, logisch wahr ist, d. h. daß Φ dann mit Ψ logisch äquivalent ist) und daß, wenn Φ und Ψ nicht zur gleichen Äquivalenzklasse gehören, $\Phi \leftrightarrow \Psi$ nicht aus jenen Axiomen logisch folgt. Die Interpretation J der Objektsprache, so wie diese im III. Kapitel eingeführt worden ist, liefert dann eine umkehrbar eindeutige Abbildung der Sachverhalte auf diese Äquivalenzklasse von Sätzen.

Zu zeigen ist noch, welche Beziehung zwischen diesen Sachverhalten und dem durch M erzeugten Mengenkörper bestehen. Die Interpretation J ist so gewählt worden, daß die Gegenstandsausdrücke α_1, α_2, ... bei ihr gerade die Experimente (oder Gegenstände) a_1, a_2, ... bezeichnen (die Anzahl der Gegenstandsausdrücke und entsprechend die Anzahl der Experimente kann dabei endlich oder abzählbar unendlich sein). Ferner sind einerseits alle Sätze in α_k logisch äquivalent

[18] Ich habe allerdings in Essler [1], S. 14f., an einem Beispiel zu zeigen versucht, daß dies nur eine kürzere und elegantere Formulierung dafür ist, daß aus dem *Satz*, der dies und jenes besagt, der *Satz* folgt, daß das und das gilt. Sachverhalte bestehen oder bestehen nicht, man kann sie nicht voraussetzen oder aus anderen folgern; wohl aber kann man Behauptungen voraussetzen und Sätze aus anderen Aussagen deduzieren.

mit einer Adjunktion von V-Sätzen in α_k und somit aus diesen durch Adjunktion darstellbar, und es sind andererseits sämtliche Teilklassen M_k mit Vereinigungen von gewissen atomaren Ereignissen identisch und somit als derartige Vereinigungen darstellbar. Es liegt daher nahe, anzunehmen, daß die V-Sätze in α_k unter J gerade die Ergebnisse des k-ten Experiments beschreiben.[19] Erhärtet wird diese Annahme durch die eingangs dieses Abschnittes aufgestellte Forderung, daß jede Realisierung einer Versuchsvorschrift ein und nur ein Ergebnis hat, daß die Gesamtheit der Ergebnisse also vollständig ist und daß sich zwei verschiedene Ergebnisse (logisch oder auf Grund der Versuchsvorschrift, d. h. auf Grund gewisser Axiome) ausschließen.

Die umkehrbar eindeutige Abbildung der Sachverhalte auf den Ereigniskörper ergibt sich dann also dadurch, daß durch J die Ausdrücke $\alpha_1, \alpha_2, \ldots$ die Gegenstände $a_1, a_2, \ldots$ bezeichnen und die einzelnen V-Sätze in α_k (für jedes k) die einzelnen Ergebnisse beschreiben. Den Sachverhalten, die die V-Sätze beschreiben, werden dann als atomare Ereignisse von M_k Einerklassen zugeordnet, deren einziges Element sie sind, und einer beliebigfachen Adjunktion von V-Sätzen in α_k die Vereinigung der entsprechenden atomaren Ereignisse von M_k. Den Sachverhalten, die die Zustandsbeschreibungen ausdrücken, entsprechen dann die Ergebnisse des Vektors $\mathfrak{a}$ und somit die atomaren Ereignisse von $\mathfrak{K}$, d. h. jene Teilklassen von M, die genau ein Element enthalten. Die Gesamtheit der Sätze der Sprache (bzw. die Gesamtheit der in Äquivalenzklassen aufgeteilten Sätze der Sprache) bildet also einen Booleschen Verband, der isomorph zum Mengenkörper $\mathfrak{K}$ ist. Jeder beliebigen Funktion P, die über einem derartigen Mengenkörper $\mathfrak{K}$ definiert ist, entspricht also umkehrbar eindeutig eine Funktion c, die über Sätzen der Sprache definiert ist (wobei sie Sätzen der gleichen Äquivalenzklasse gleiche Werte zuordnet), und zwar der einstelligen Funktion P die Funktion c, deren zweites Argument die

[19] Carnap hat nur ganz bestimmte Sätze von unendlicher Länge in Betracht gezogen, nämlich solche, die mit All- oder Existenzbehauptungen äquivalent sind; hier werden beliebige Sätze von unendlicher Länge zugelassen, so daß auch im Fall abzählbar unendlich vieler Grundprädikate die umkehrbar eindeutige Entsprechung bestehen bleibt; in diesem Fall ist die Anzahl der V-Prädikate in α_k normalerweise überabzählbar unendlich, es sei denn, daß alle bis auf abzählbar unendlich viele von ihnen den Axiomen der Sprache widersprechen (etwa bei einem fiktiven Würfel mit abzählbar unendlichen vielen Seiten).

Tautologie ist, und der zweistelligen Funktion P die Funktion c ohne diese Einschränkung.[20]

Zu zeigen ist ferner, welche Bestimmungen, die den Begriff „P" näher festlegen, unter Zugrundelegung dieser Übereinstimmung von Normalinterpretation und Objektsprache den Adäquatheitsbedingungen der induktiven Logik entsprechen. Abgesehen von (II) (a) und der Bedingung der unabhängigen Wiederholbarkeit eines Experiments genügt es dabei, auf früher gemachte Äußerungen hinzuweisen bzw. diese kurz zu wiederholen. Daß die Bedingung (4) in der Definition des Begriffs „Wahrscheinlichkeitsfeld" mit dem Begriff der Bestätigungsmethode (allerdings ohne Regularität) logisch äquivalent ist, ist bereits gesagt worden. Wichtig ist nun, daß sich die Forderung der unabhängigen Wiederholbarkeit eines Experiments (eingangs als Forderung 3 aufgestellt) und die Bedingung (II) (a) in der gleichen Weise entsprechen.[21] Auch dies kann nicht klarer als die unklarste der beiden Bestimmungen gezeigt werden, und auch hier ist die von den Mathematikern benützte die unpräzisere. Zwei Gründe legen es nahe, die beiden Forderungen als äquivalent anzusehen:

Ein formaler Grund dafür ist, daß die Mathematiker aus den Kolmogoroff-Axiomen und der Voraussetzung der unabhängigen Wiederholbarkeit eines Experiments die gleichen Theoreme gewinnen, die sich auch aus den Bedingungen (I) und (II) (a) ergeben (von denen einige im ersten Abschnitt dieses Kapitels dargestellt worden sind).

Ein mehr intuitiver Grund dafür ist die folgende Überlegung: Wenn ein Experiment nicht unabhängig wiederholbar ist, so wird man zweckmäßigerweise die Funktion P so wählen, daß nicht allgemein

---

[20] Werden die atomaren Ereignisse nicht als Einerklassen von *Sachverhalten*, sondern als Einerklassen von *Sätzen* konstituiert, was aus philosophischen Überlegungen heraus als zweckmäßig erscheint, so vereinfachen sich die geschilderten Abbildungen entsprechend. Ich möchte noch erwähnen, daß speziell für das System von Richter erstmals Jürgen Humburg in [1], S. 28—34, einen derartigen Nachweis der Gleichwertigkeit von Normalinterpretation des mathematischen Apparats und der Objektsprache, die der Funktion c zugrundegelegt wird, erbracht hat.

[21] Wichtig ist es in diesem Zusammenhang, daß die Bedingung der *unabhängigen Wiederholbarkeit von Experimenten* scharf von der Behauptung der *Unabhängigkeit von Ereignissen* (die mit der Aussage der *Irrelevanz von Prämissen* gleichwertig ist) unterschieden wird, was bei Mathematikern nicht immer geschieht; vgl. z. B. Basler [1], S. 38f.

gilt: $P(F^{\star}_i|a_3, F^{\star}_j|a_1) = P(F^{\star}_i|a_2, F^{\star}_j|a_1)$; denn der Wert von $P(F^{\star}_i|a_3, F^{\star}_j|a_1)$ hängt ja unter anderem auch davon ab, wie das zweite Experiment ausgegangen ist, während der Wert von $P(F^{\star}_i|a_2, F^{\star}_j|a_1)$ davon unabhängig ist. In diesem Fall wird man jedoch auch eine Funktion c als (im intuitiven Sinn) adäquat ansehen, die (II) (a) verletzt und die c-Werte besitzt, die den P-Werten entsprechen.[22] Wenn man umgekehrt eine Bestätigungsmethode c als (im intuitiven Sinn) adäquat ansieht, die (II) (a) verletzt, so wird man Gründe dafür haben, die besagen, daß es Zusammenhänge zwischen den durch die Gegenstandskonstanten bezeichneten Gegenstände gibt, die eine Funktion, die (II) (a) erfüllt, nicht ohne weiteres berücksichtigt. Die Existenz solcher Zusammenhänge ist jedoch gleichwertig damit, daß das betreffende Experiment nicht unabhängig wiederholbar ist.

Die Frage, was den Bedingungen (I) (6), (II) (b), (III) bzw. (III′) und (IV) in der mathematischen Wahrscheinlichkeitslehre entspricht, ist ebenfalls bereits beantwortet worden, so daß ich mich hier auf ein paar kurze Andeutungen beschränken kann. Zunächst scheint diesen Adäquatheitsbedingungen in der Wahrscheinlichkeitslehre nichts zu entsprechen, da nirgendwo analoge Forderungen an die Funktion P gestellt werden. Einige Statistiker glauben daher, daß die Frage der relativen Weite der Prädikate der Sprache (und deren Eigenschaften, die sie bezeichnen) in der Statistik nirgendwo gestellt wird.[23] Eine derartige Behauptung ist widerlegt, wenn man ein Gegenbeispiel angibt. Hierzu genügt es, auf Theorem 24 zu verweisen. Bestimmte Wahrscheinlichkeitsverteilungen auf die einzelnen relativen Häufigkeiten (auf die statistischen Verteilungen) führen zu Funktionen, die jenen Bedingungen genügen.

Natürlich gibt es andere Wahrscheinlichkeitsbewertungen jener objektiven Wahrscheinlichkeiten, die zu c-Funktionen führen, die einige der Bedingungen (I) (6), (III) (b), (III) bzw. (III′) und (IV) verletzen. Auch hierzu braucht nur noch wiederholt zu werden, was

---

22 Das ändert natürlich nichts an der Überlegung, daß die c-Werte mit denen einer L-Methode, die als Zusatzprämisse das Wissen um die genaue Abhängigkeit enthält, identisch sind.

23 Vetter ist z. B. ein solcher Statistiker, der dies glaubt, vgl. Vetter [2], S. 56 bis 61.

bereits mehrfach gesagt worden ist: Falls man in einer gegebenen Situation eine solche Bewertung bzw. Verteilung als vernünftig ansieht, dann muß man (als rationaler Mensch) in der Lage sein, Gründe dafür anzugeben. Wenn man dazu in der Lage ist, dann wäre es vom *theoretischen* Standpunkt aus zweckmäßiger gewesen, eine andere Verteilung zu wählen, die zu einer L-Methode führt, und jene Gründe der Prämisse konjunktiv hinzuzufügen (vom *praktischen* Standpunkt aus wäre ein solches Verfahren allerdings zugegebenermaßen etwas umständlich).

Wenn zwei Systeme, wie etwa die induktive Logik und die mathematische Wahrscheinlichkeitslehre, miteinander äquivalent sind, so hat es nicht viel Sinn zu fragen, welches von beiden das bessere ist. Wohl aber hat es Sinn zu fragen, welches von beiden für gewisse Zwecke das geeignetere ist. Für die Zwecke der Deduktion weiterer Lehrsätze sowie der Anwendung der Theorie scheint die Wahrscheinlichkeitslehre geeigneter zu sein, vor allem deshalb, weil diese von Mathematikern und Statistikern betrieben wird, die zwar die Sprache der Mathematik, aber selten oder nie die Sprache der Logik gebrauchen. Allerdings ist dies zum Teil auch eine Frage der Zeit; in einigen Jahrzehnten kann die Lage hier vielleicht ganz anders aussehen.

Es muß jedoch zugegeben werden, daß die Behandlung der Materie durch den Logiker viel umständlicher und aufwendiger geschieht als durch den Mathematiker. Dies ist eine unvermeidliche Folge seines Präzisionsfanatismus und seiner Pedanterie, mit der er zu Werke geht. Allerdings will der Logiker im allgemeinen auch den Gegenstand nicht so darstellen, daß er aus irgendwelchen Grundsätzen alle denkbaren und für irgendwelche Zwecke wichtigen Lehrsätze deduziert, sondern er will eben diese Grundsätze *begründen*.[24] Der Mathematiker braucht also nicht zu befürchten, daß ein Logiker ein ihm unverständliches, weil in einer ihm ungewohnten Sprache geschriebenes Textbuch über ein Gebiet schreibt, das er Wahrscheinlichkeitslehre nennt (obwohl ich mir nicht vorstellen kann, daß ein derartiges Werk der Mathematik schaden würde).

Es ist demnach eine Tatsache, daß Logiker und Mathematiker auf diesem Gebiet verschiedene Interessen haben. Deshalb wird der Logiker

[24] So ist auch die Leistung des Logikers (und Mathematikers) Gottlob Frege bei der Grundlegung der Theorie der natürlichen Zahlen zu bewerten.

selten mit Borelschen Erweiterungen einer Menge und noch seltener mit Lebesgueschen Integralen operieren, und der Mathematiker wird kaum auf die Forderung des Gesamtdatums oder auf das Problem der zerrütteten Prädikate stoßen. Selbstverständlich sind gewisse Probleme in einer Theorie, die Sachverhalte statt Sätze zugrundelegt, nicht einfach zu behandeln;[25] doch kann der mathematische Ansatz genauso wie der logische auf Relationen erweitert werden (und an einer Erweiterung auf Sprachen vom Ausdrucksreichtum der einfachen Typentheorie ist der Mathematiker vielleicht mangels Anwendungsmöglichkeiten gar nicht interessiert). Wem daran liegt, versteckte Probleme offenkundig zu machen, wird vielleicht am logischen Ansatz Gefallen finden. Einige dieser Probleme sollen nun erörtert werden.

Im III. Kapitel ist die Untersuchung auf Objektsprachen vom Ausdrucksreichtum der engeren Quantorenlogik eingeschränkt worden (die zusätzliche Einschränkung auf endliche Gegenstandsbereiche ist später durch die Limeskonvention teilweise wieder aufgehoben worden), und im IV. Kapitel ist zusätzlich vorausgesetzt worden, daß die Sprache an Prädikaten nur Eigenschaftsausdrücke enthält. Es wird nun an Beispielen angedeutet, wie man sich von diesen Einschränkungen befreien bzw. wie man den bisher entwickelten Ansatz auf ausdrucksreichere Sprachen erweitern und verallgemeinern kann.

Gegeben sei eine über dem Bereich der Menschen interpretierte Objektsprache, die neben den vier Gegenstandsausdrücken „a", ... „d" nur die Ausdrücke „Q" und „R" enthält, die für zweistellige Beziehungen stehen, wie etwa „verliebt in" und „verheiratet mit". Ein Gegenstand*paar* ⟨a, b⟩ (und nicht, wie im Fall der Eigenschaftsausdrücke, d. h. der *einstelligen* Beziehungsausdrücke, ein *Gegenstand*) ist dann (hinsichtlich der Ausdrucksmöglichkeiten der Sprache) vollständig beschrieben, wenn gesagt wird, daß a in b verliebt und mit b verheiratet ist, daß a in b verliebt, aber nicht mit b verheiratet ist, daß a nicht in b verliebt, aber mit b verheiratet ist oder schließlich, daß a weder in b verliebt noch mit b verheiratet ist. Die 4 ($= 2^2$) V-Prädikate „Q_1", ..., „Q_4" können dann folgendermaßen definiert werden:

[25] Man stelle sich etwa eine Theorie der Identitätslogik vor, bei der der Folgerungsbegriff nicht über Sätze, sondern über Sachverhalte definiert ist!

„$\Lambda\, x\, \Lambda\, y\, [\langle x, y\rangle \varepsilon Q_1 \leftrightarrow \langle x, y\rangle \varepsilon Q \wedge \langle x, y\rangle \varepsilon R]$“
„$\Lambda\, x\, \Lambda\, y\, [\langle x, y\rangle \varepsilon Q_2 \leftrightarrow \langle x, y\rangle \varepsilon Q \wedge \neg (\langle x, y\rangle \varepsilon R)]$“
„$\Lambda\, x\, \Lambda\, y\, [\langle x, y\rangle \varepsilon Q_3 \leftrightarrow \neg (\langle x, y\rangle \varepsilon Q) \wedge \langle x, y\rangle \varepsilon R]$“
„$\Lambda\, x\, \Lambda\, y\, [\langle x, y\rangle \varepsilon Q_4 \leftrightarrow \neg (\langle x, y\rangle \varepsilon Q) \wedge \neg (\langle x, y\rangle \varepsilon R)]$“

Der Gegenstandsbereich B bestehe aus den Menschen a (= Hans Meier), ..., d (= Lieschen Müller), die unter der vorgegebenen Interpretation durch die Ausdrücke „a“, ..., „d“ bezeichnet werden. Es wird dann von B zu einem neuen Bereich B′ von Pseudoobjekten übergegangen, die ihrerseits geordnete Paare von Elementen von B sind; B′ besteht also aus:

$\langle a, a\rangle$, $\langle a, b\rangle$, $\langle a, c\rangle$, $\langle a, d\rangle$,
$\langle b, a\rangle$, $\langle b, b\rangle$, $\langle b, c\rangle$, $\langle b, d\rangle$,
$\langle c, a\rangle$, $\langle c, b\rangle$, $\langle c, c\rangle$, $\langle c, d\rangle$,
$\langle d, a\rangle$, $\langle d, b\rangle$, $\langle d, c\rangle$, $\langle d, d\rangle$.

Die neuen V-Prädikate „Q_1“, ..., „Q_4“ werden nun als Pseudoeigenschaftsausdrücke aufgefaßt, die unter der vorgegebenen Interpretation Pseudoeigenschaften (nämlich Teilklassen des Bereichs B′ von Pseudogegenständen) bezeichnen. Eine Zustandsbeschreibung (Standardkonjunktion einer maximalkonsistenten Menge von Sätzen) ist dann ein Satz, der jedem der Pseudoobjekte (in der Reihenfolge, in der sie oben angeführt sind) eine der vier Pseudoeigenschaften zuschreibt.

Wenn nicht alle Prädikate von gleicher Stellenzahl sind, ist hier allerdings eine gewisse Vorsicht am Platze, da dann nicht jeder V-Satz erfüllbar ist und damit eine Zustandsbeschreibung nur durch konjunktive Verknüpfung von *erfüllbaren* V-Sätzen gebildet werden kann. Die Objektsprache enthalte an Grundprädikaten z. B. den Eigenschaftsausdruck „F“ und den zweistelligen Beziehungsausdruck „Q“. Die zweistelligen V-Prädikate „Q_1“, ..., „Q_8“ können dann folgendermaßen definiert werden:

„$\Lambda\, x\, \Lambda\, y\, [\langle x, y\rangle \varepsilon Q_1 \leftrightarrow \langle x, y\rangle \varepsilon Q \wedge x \varepsilon F \wedge y \varepsilon F]$“
„$\Lambda\, x\, \Lambda\, y\, [\langle x, y\rangle \varepsilon Q_2 \leftrightarrow \langle x, y\rangle \varepsilon Q \wedge x \varepsilon F \wedge \neg (y \varepsilon F)]$“
„$\Lambda\, x\, \Lambda\, y\, [\langle x, y\rangle \varepsilon Q_3 \leftrightarrow \langle x, y\rangle \varepsilon Q \wedge \neg (x \varepsilon F) \wedge y \varepsilon F]$“
„$\Lambda\, x\, \Lambda\, y\, [\langle x, y\rangle \varepsilon Q_4 \leftrightarrow \langle x, y\rangle \varepsilon Q \wedge \neg (x \varepsilon F) \wedge \neg (y \varepsilon F)]$“
„$\Lambda\, x\, \Lambda\, y\, [\langle x, y\rangle \varepsilon Q_5 \leftrightarrow \neg (\langle x, y\rangle \varepsilon Q) \wedge x \varepsilon F \wedge y \varepsilon F]$“
„$\Lambda\, x\, \Lambda\, y\, [\langle x, y\rangle \varepsilon Q_6 \leftrightarrow \neg (\langle x, y\rangle \varepsilon Q) \wedge x \varepsilon F \wedge \neg (y \varepsilon F)]$“
„$\Lambda\, x\, \Lambda\, y\, [\langle x, y\rangle \varepsilon Q_7 \leftrightarrow \neg (\langle x, y\rangle \varepsilon Q) \wedge \neg (x \varepsilon F) \wedge y \varepsilon F]$“
„$\Lambda\, x\, \Lambda\, y\, [\langle x, y\rangle \varepsilon Q_8 \leftrightarrow \neg (\langle x, y\rangle \varepsilon Q) \wedge \neg (x \varepsilon F) \wedge \neg (y \varepsilon F)]$“

Aus diesen V-Prädikaten kann dann „Q" folgendermaßen konstituiert werden:
„$\Lambda\, x\, \Lambda\, y\, [\langle x, y\rangle \varepsilon\, Q \leftrightarrow \langle x, y\rangle \varepsilon\, Q^{\star}_1\, v\, \langle x, y\rangle \varepsilon\, Q^{\star}_2\, v\, \langle x, y\rangle \varepsilon\, Q^{\star}_3\, v\, \langle x, y\rangle \varepsilon\, Q^{\star}_4]$".
Der entsprechende Satz für „F" lautet:
„$\Lambda\, x\, [x \varepsilon F \leftrightarrow \langle x, x\rangle \varepsilon\, Q^{\star}_1\, v\, \langle x, x\rangle \varepsilon\, Q^{\star}_2\, v\, \langle x, x\rangle \varepsilon\, Q^{\star}_5\, v\, \langle x, x\rangle \varepsilon\, Q^{\star}_6]$"
Da die Teilbehauptungen „$\langle x, x\rangle \varepsilon\, Q^{\star}_2$" und „$\langle x, x\rangle \varepsilon\, Q^{\star}_6$" jedoch kontradiktorisch sind, ist jener Teilsatz logisch äquivalent mit
„$\Lambda\, x\, [x \varepsilon F \leftrightarrow \langle x, x\rangle \varepsilon\, Q^{\star}_1\, v\, \langle x, x\rangle \varepsilon\, Q^{\star}_5]$".
Gegeben sei eine beliebig große natürliche Zahl (und damit endliche Anzahl) m und eine Objektsprache, die an Grundprädikaten nur Beziehungsausdrücke enthält, deren Stellenzahl zwischen 1 und m liegt (die Grenzen eingeschlossen). Dann kann in der geschilderten Weise von den Gegenständen des Bereichs B zu den Pseudogegenständen (den geordneten m-Tupeln von Gegenständen), zu den V-Prädikaten und Pseudoeigenschaften sowie zu den Zustandsbeschreibungen übergegangen werden, und es lassen sich auf diese Weise alle Lehrsätze der Kapitel IV bis VI auf diese ausdrucksreiche Sprache übertragen.
Prinzipielle Schwierigkeiten entstehen, wenn man nicht eine derartige endliche obere Schranke m voraussetzen, sondern Prädikate von beliebig großer Stellenzahl betrachten will; wenn der Gegenstandsbereich B abzählbar unendlich ist, erhält der Bereich der Pseudoobjekte (der ∞-Tupel) dann nämlich eine Anzahl, die nicht mehr abzählbar unendlich (und selbstverständlich auch nicht endlich), sondern überabzählbar unendlich ist. Möglicherweise lassen sich diese Schwierigkeiten dadurch überwinden, daß man zunächst stets nur den endlichen Fall im Auge hat und daß es sich beim Grenzübergang erweist, daß man nicht alle dieser überabzählbar unendlich vielen Pseudoobjekte benötigt. Der Fall, daß Relationsausdrücke von beliebig hoher Stellenzahl vorliegen, ist ja ohnehin nur von rein akademischem Interesse, da in keinem mir bekannten Fall eine derartige ausdrucksreiche Sprache tatsächlich benötigt wird. Dies ist natürlich kein Grund dafür, die Schwierigkeiten vom Tisch zu wischen, da es immerhin möglich ist, daß sich irgendwann die Verwendung derartiger Sprachen in den Wissenschaften als zweckmäßig erweist.
Für einen speziellen Fall kann das Problem als gelöst gelten, und zwar dann, wenn man a priori voraussetzt, daß für eine Hypothese, die

(neben den Gegenstandsausdrücken) nur m-stellige Prädikate enthält, nur jener Teil der Prämisse relevant ist, der (neben den Gegenstandsausdrücken) ebenfalls nur m-stellige Beziehungsausdrücke enthält. Dann kann man nämlich folgendermaßen vorgehen: Man betrachtet zunächst für jede von der 0 verschiedenen natürlichen Zahl m jene Teilsprache, die neben den Gegenstandsausdrücken nur m-stellige Prädikate enthält, bildet die Zustandsbeschreibungen dieser Teilsprachen und berechnet für sie mittels einer vorgegebenen regulären Bestätigungsmethode bzw. mittels einer vorgegebenen L-Methode die Aprioriwerte für diese Sätze. Eine Zustandsbeschreibung Σ der Gesamtsprache ist dann logisch äquivalent mit einer unendlichen Konjunktion $\Sigma_1 \wedge \Sigma_2 \wedge \Sigma_3 \wedge \ldots$, wobei jedes Σ_m eine Zustandsbeschreibung für die Teilsprache mit den m-stelligen Prädikaten ist. Die Aprioriwahrscheinlichkeit für Σ ergibt sich dann durch ein unendliches Produkt mittels der Gleichung:

$c(\Sigma, T) = c(\Sigma_1, T) \cdot c(\Sigma_2, T) \cdot c(\Sigma_3, T) \cdot \ldots$

Mit Bedingung (5) von D 4 müßte diese Gleichung an sich lauten:

$c(\Sigma, T) = c(\Sigma_1, T) \cdot c(\Sigma_2, \Sigma_1) \cdot c(\Sigma_3, \Sigma_1 \wedge \Sigma_2) \cdot \ldots$

Es wird hier also vorausgesetzt, daß $c(\Sigma_{m+1}, \Sigma_1 \wedge \ldots \wedge \Sigma_m) = c(\Sigma_{m+1}, T)$, d. h. daß $\Sigma_1 \wedge \ldots \wedge \Sigma_m$ für Σ_{m+1} *a priori irrelevant* ist; diese Irrelevanz überträgt sich, wie man sich leicht klarmachen kann, auch auf alle Aposterioriwahrscheinlichkeiten.

Die Voraussetzung einer derartigen Irrelevanz (einer derartigen Unabhängigkeit der entsprechenden Ereignisse) kann aber in bestimmten Situationen inadäquat sein. Man denke etwa an eine Objektsprache, deren Gegenstandsausdrücke über dem Bereich der Menschen interpretiert sind, die zwei Eigenschaftsausdrücke „F" und „G" und den Beziehungsausdruck „Q" enthält, die bei dieser Interpretation die Eigenschaften *intelligent* und *reich* sowie die Beziehung *Ratgeber von* bezeichnen (es ist dabei natürlich an *bezahlte* Ratgeber gedacht). Es ist dann unvernünftig, als Bestätigungsmethode eine Funktion c zu wählen, die einen induktiven Zusammenhang zwischen den beiden Eigenschaften und der Beziehung ausschließt, da sich möglicherweise reiche Leute intelligente Ratgeber halten (irgendwie müssen sie oder ihre Vorfahren ja reich geworden sein) und da intelligente Menschen eventuell vorwiegend in abhängigen Stellungen beschäftigt sind.

Vergleichsweise unproblematisch ist auch die Erweiterung des Be-

griffs der induktiven Methode auf Objektsprachen der einfachen (endlichen) Typentheorie, falls dabei vorausgesetzt wird, daß zwischen den Teilzustandsbeschreibungen der verschiedenen Typen kein induktiver Zusammenhang besteht. Da in der deduktiven Logik von jedem Typ von Entitäten höchstens abzählbar unendlich viele benötigt werden,[26] auch wenn nach den Gesetzen der Mengenlehre bei unendlichem Grundbereich verschiedene Teilbereiche (Bereiche von Entitäten eines bestimmten Typs) überabzählbar unendlich sind, wird zusätzlich angenommen, daß dies auch beim induktiven Argumentieren zutrifft.[27] Es wird deshalb angenommen, daß entweder sowohl der Grundbereich als auch jeder der einzelnen Teilbereiche der Gesamtheit der Entitäten, die durch einen bestimmten Typ charakterisiert sind, endlich, oder aber, daß alle abzählbar unendlich sind. Durch eine Interpretation wird eine umkehrbar eindeutige Abbildung dieses Bereichs der Entitäten (dieser Erweiterung des Grundbereichs entsprechend der Typentheorie) auf die deskriptiven Ausdrücke einer Sprache angegeben. Sodann wird die Sprache entsprechend den Typen in einzelne Teilsprachen aufgeteilt; es wird für die einzelnen Teilsprachen festgelegt, welche Sätze Zustandsbeschreibungen sind, es werden für diese Teilsprachen und den ihnen entsprechenden Teilbereichen Adäquatheitsbedingungen für induktive Methoden formuliert, und es wird schließlich, nach der Wahl einer adäquaten induktiven Methode (normalerweise einer L-Methode), der Aprioriwert einer Zustandsbeschreibung für die Gesamtsprache durch das Produkt der Aprioriwerte der Zustandsbeschreibungen für die Teilsprachen festgelegt, d. h. es wird, wenn Σ die Standardkonjunktion der Ge-

---

[26] Vgl. Henkins Erfüllbarkeitslemma in: Leon Henkin, Completeness in the Theory of Types, in: JSL 15 (1950) S. 85. Die dort nicht berücksichtigten logischen Grundsätze (das Extensionalitätsprinzip, das Abstraktionsprinzip und das Auswahlprinzip) können ja wie außerlogische Axiome behandelt und der Theorie oder gar jedem einzelnen deduktiven und induktiven Argument als Prämisse vorangestellt werden.

[27] Diese Annahme ist nicht ganz unproblematisch; denn ein deduktiver Schluß ist ja ein Grenzfall einer induktiven Argumentation, und es ist möglich, daß die Überabzählbarkeit für induktive Argumente im allgemeinen relevant, für die Grenzfälle jedoch irrelevant ist. Allerdings besteht die begründete Vermutung, daß sich das Theorem 2 von Gaifman ([1], S. 7) auch auf Sprachen der einfachen Typentheorie erweitern läßt, womit die Annahme bewiesen wäre.

samtheit aller Sätze Σ^i ist, die aus jeder Teilsprache genau eine Zustandsbeschreibung enthält, festgesetzt, daß
$c(\Sigma, T) = c(\Sigma^1, T) \cdot c(\Sigma^2, T) \cdot \ldots \cdot c(\Sigma^i, T) \cdot \ldots .$
Mit diesen Bemerkungen ist natürlich nur angedeutet, in welcher Richtung für einen speziellen Fall (nämlich den der wechselseitigen Irrelevanz der Zustandsbeschreibungen verschiedenen Typs) eine Lösung des Problems der Definition des Begriffs der adäquaten induktiven Methode gefunden werden kann; als Lösung des Problems hingegen können sie nicht angesehen werden. Wie die Erweiterung des Ansatzes auf die einfache Typentheorie zu geschehen hat, wenn diese Voraussetzung der Irrelevanz nicht angenommen werden darf, ist im Moment schwer vorstellbar.
Die Ergebnisse des Kapitels IV können auch in einer anderen Weise auf mehrstellige Prädikate übertragen werden: Es wird hierbei nicht der Gegenstandsbereich durch einen neuen Bereich von Pseudoobjekten ersetzt, sondern es werden zusätzlich zu den vorgegebenen einstelligen Grundprädikaten neue einstellige *Pseudogrundprädikate* betrachtet, die mittels der vorgegebenen mehrstelligen Grundprädikate und der Gegenstandsausdrücke definierbar sind. Wenn die Sprache etwa die Gegenstandsausdrücke „a" und „b", die einstelligen Begriffe „F" und „G" sowie den zweistelligen Relationsausdruck „R" enthält, so wird „R" aus ihr entfernt und die so entstehende Restsprache durch die folgendermaßen definierbaren Eigenschaftsausdrücke „F_1", ..., „F_4" ergänzt:
„$\Lambda x [x \varepsilon F_1 \leftrightarrow \langle x, a\rangle \varepsilon R]$"
„$\Lambda x [x \varepsilon F_2 \leftrightarrow \langle x, b\rangle \varepsilon R]$"
„$\Lambda x [x \varepsilon F_3 \leftrightarrow \langle a, x\rangle \varepsilon R]$"
„$\Lambda x [x \varepsilon F_4 \leftrightarrow \langle b, x\rangle \varepsilon R]$"
Dieses Verfahren funktioniert nur für Prädikate der engeren Quantorenlogik, da auf diese Weise nur mehrstellige Relationsausdrücke der ersten Stufe durch einstellige Begriffe der ersten Stufe ersetzt werden können. Ob es in diesem Rahmen mit der oben ausgeführten Methode gleichwertig ist (d. h. hier: für die Zwecke der induktiven Logik zu den gleichen Ergebnissen führt) ist bislang ein offenes Problem, desgleichen die Frage, welchem der beiden Verfahren der Vorzug zu geben ist, falls sie nicht in diesem Sinne gleichwertig sind.
Bei den bisherigen Überlegungen und insbesondere auch beim Voll-

ständigkeitsbeweis im V. Kapitel ist stets davon ausgegangen worden, daß man den Uniformitätsgrad des Gegenstandsbereichs hinsichtlich *aller Eigenschaften zusammen* und nicht *für jede extra* berechnen muß, so daß konsequenterweise auch hinsichtlich *aller Eigenschaften zusammen* (bzw. hinsichtlich der V-Eigenschaften *aller* in Erwägung gezogener Eigenschaften) die optimale induktive Methode zu ermitteln ist. Carnap hat nun Argumente vorgebracht, die diese stillschweigende Voraussetzung in Frage stellen.[28]

Gegeben sei ein endlicher Gegenstandsbereich B, der hinsichtlich der beiden Eigenschaften F und G betrachtet wird. Es sei B etwa die Gesamtheit der bei Siemens beschäftigten Menschen (einschließlich jener Personen, die sich nicht direkt mit Forschungsaufgaben befassen, wie etwa Putzfrauen und Direktoren), F sei die Eigenschaft *intelligent* und G die Eigenschaft *sterblich*. Der Bereich B habe eine gerade Anzahl, und der klassifikatorische Begriff „intelligent" sei (durch Zufall oder absichtlich) so festgelegt, daß genau der Hälfte der Mitglieder von B die Eigenschaft F zukommt. Es darf vorausgesetzt werden, daß alle bei Siemens Beschäftigten sterblich sind (die Sage berichtet zwar, daß der Mensch Elias nicht gestorben ist, doch war dieser nicht bei Siemens beschäftigt). Den V-Eigenschaften *intelligent und sterblich* sowie *nicht intelligent, aber sterblich* kommen dann in der Klasse B die relativen Häufigkeiten $1/2$ und den V-Eigenschaften *intelligent, aber nicht sterblich* sowie *weder intelligent noch sterblich* in der Bezugsklasse B die relativen Häufigkeiten 0 zu.

Der Uniformitätsgrad w ist nun nach Definition die Summe der Quadrate dieser relativen Häufigkeiten, d. h. es gilt:

$w = (1/2)^2 + 0^2 + (1/2)^2 + 0^2 = 1/2.$

Die optimale Methode c^L ist dann bestimmt durch:

$L = (1 - w)/(w - 1/q) = 2.$

Läßt man nun die Eigenschaft *intelligent* außer Betracht, so erhält man an V-Prädikaten lediglich *sterblich* und *nicht sterblich* sowie an relativen Häufigkeiten die Zahlen 0 und 1. Es gilt dann: $w_1 = 1^2 + 0^2 = 1$, so daß die optimale Methode c^L festgelegt ist durch $L_1 = 0$. Betrachtet man hingegen lediglich die Eigenschaft *intelligent*, so daß *intelligent* sowie *nicht intelligent* die einzigen V-Eigenschaften sind, so ergeben

---

[28] Vgl. Carnap-Stegmüller [1], S. 251f., und Carnap [9], S. XXIf.

sich als relative Häufigkeiten nach Voraussetzung jeweils $^1/_2$; es gilt dann, daß $w_2 = (^1/_2)^2 + (^1/_2)^2 = ^1/_2$ und für die optimale Methode c^L, daß $L_2 = \infty$.

Aus diesem Beispiel ergeben sich daher (für solche Situationen) andere Konsequenzen als die, die Carnap gezogen hat. Nach Carnap zeigen derartige Beispiele, daß es in solchen Fällen notwendig ist, den Analogieeinfluß abzuschwächen oder zu beseitigen. Beseitigt ist der Analogieeinfluß, wenn zunächst durch c^L die Aprioriwerte der Zustandsbeschreibungen der Teilsprache berechnet werden, die nur das Grundprädikat „F" enthält, sodann durch c^L die Aprioriwerte für die Zustandsbeschreibungen jener Teilsprache berechnet werden, deren einziges Grundprädikat „G" ist, und schließlich für jede Zustandsbeschreibung Σ der Gesamtsprache festgesetzt wird, daß

$c^L{}_{1;2}(\Sigma, T) = c^L(\Sigma^1, T) \cdot c^L(\Sigma^2, T)$,

wobei Σ logisch äquivalent ist mit $\Sigma^1 \wedge \Sigma^2$ und Σ^1 eine Zustandsbeschreibung für die erste und Σ^2 eine für die zweite Teilsprache ist.[29] Abgeschwächt wird der Analogieeinfluß, wenn man den direkt berechneten Wert $c^L(\Sigma, T)$ weder mit 0 noch, wie bisher, mit 1, sondern mit einem neuen, von L unabhängigen Parameter E mit $0 < E < 1$ und entsprechend $c^L{}_{1;2}(\Sigma, T)$ mit $(1 - E)$ bewertet; man kann dann eine neue Funktion $c^{L;E}$ definieren, indem man festsetzt, daß für beliebige Zustandsbeschreibungen Σ gilt:[30]

$c^{L;E}(\Sigma, T) = E \cdot c^L(\Sigma, T) + (1 - E) \cdot c^L{}_{1;2}(\Sigma, T)$.

Auf diese Weise können induktive Analogieschlüsse von einer Eigenschaft auf eine andere mit beliebig kleinen Sicherheitsgraden belegt werden.

Mir will jedoch, wie schon bemerkt, scheinen, daß das obige Beispiel andere und weiterreichende Konsequenzen hat als die, die Carnap vorgeschwebt haben. Zunächst ist zu sagen, daß in der Gesamtsprache verschiedene Prädikate (und damit verschiedene Eigenschaften) *zusammengeworfen* und durch die V-Prädikate damit Homogenitäten erzeugt werden, die, sozusagen *in Wirklichkeit*, gar nicht existieren:

---

[29] Wenn man die Werte $c^L(\Sigma, T)$ nach der durch die Theoreme 14 und 15 aufgezeigten Methode berechnet, sieht man, daß nur für $L = 0$ und $L = \infty$ gilt: $c^L(\Sigma, T) = c^L{}_{1;2}(\Sigma, T)$.

[30] Diese Definition ist gleichwertig mit der von Carnap in Carnap-Stegmüller [1] auf S. 252 gegebenen.

die beiden Teilsprachen haben durchgehende Uniformität bzw. durchgehende Disuniformität, die Gesamtsprache hingegen das Mittel davon. Es wäre dann doch, statt den Analogieinfluß abzuschwächen, vernünftiger, die einzelnen Grundprädikate extra zu behandeln, d. h. den Uniformitätsgrad des Gegenstandsbereichs hinsichtlich jeder Grundeigenschaft und deren Komplemente extra zu bestimmen, auf diese Weise für jede der Teilsprachen die optimale induktive Methode zu ermitteln und auf den Analogieeinfluß dann ganz zu verzichten, d. h. die Aprioriwerte der einzelnen Teilzustandsbeschreibungen miteinander zu multiplizieren.[31]

Es seien $\Sigma^1_1, \ldots, \Sigma^1_m$ und $\Sigma^2_1, \ldots, \Sigma^2_m$ die Zustandsbeschreibungen der beiden Teilsprachen des obigen Beispiels, und es sei

$c^{L_1}(\Sigma^1_1, T) = r_1, \ldots, c^{L_1}(\Sigma^1_m, T) = r_m$, und

$c^{L_2}(\Sigma^2_1, T) = s_1, \ldots, c^{L_2}(\Sigma^2_m, T) = s_m$.

Dann ist $r_1 + \ldots + r_m = 1$ und $s_1 + \ldots + s_m = 1$. Es gilt in diesem Fall:

$$
\begin{aligned}
1 &= r_1 + \ldots + r_m \\
&= r_1 \cdot 1 + \ldots + r_m \cdot 1 \\
&= r_1 \cdot (s_1 + \ldots + s_m) + \ldots + r_m \cdot (s_1 + \ldots + s_m) \\
&= r_1 \cdot s_1 + \ldots + r_1 \cdot s_m + \ldots + r_m \cdot s_1 + \ldots + r_m \cdot s_m.
\end{aligned}
$$

Somit charakterisiert diese Vertei ung von m^2 Aprioriwerten $r_i \cdot s_j$ auf die m^2 Zustandsbeschreibungen der Gesamtsprache eine Bestätigungsmethode $c^{L_1|L_2}$, die, falls keiner der Werte mit 0 identisch ist (d. h. falls $L_1 \neq 0 \neq L_2$), regulär ist. $c^{L_1|L_2}$ ist also selbst keine L-Methode, sie ist jedoch durch solche determiniert.

Nicht immer wird der Uniformitätsgrad des Bereichs hinsichtlich der einzelnen Eigenschaften verschieden sein. Wenn er etwa aus einer

---

[31] Es kann sich darüber hinaus gelegentlich sogar als notwendig erweisen, auf derartige Multiplikationen ganz oder teilweise zu verzichten. Gewisse Schwierigkeiten in der Quantenmechanik (daß nämlich zwar Verteilungen für den Ort von Partikeln wie auch für deren Impuls, aber nicht gemeinsame Verteilungen für Ort *und* Impuls existieren) können, falls man nicht Hand an die klassische Logik und an die unter ihrer Hilfe entwickelte Mathematik legen will, vermutlich so behoben werden, daß man für die einzelnen Größen (Ort und Impuls) verschiedene Wahrscheinlichkeitskörper (in unserer Terminologie: verschiedene Sprachsysteme) aufbaut und für beide getrennt die Verteilungen bestimmt, ohne in allen Fällen gemeinsame Verteilungen anzunehmen (für beide Sprachsysteme getrennt Wahrscheinlichkeitsfunktionen charakterisiert, ohne sie generell in mathematischen Zusammenhang zu bringen).

Urne voll Kugeln besteht und die einzelnen Eigenschaften die Spektralfarben sind, so wird man mit Recht annehmen, daß die einzelnen Uniformitätsgrade miteinander identisch sind.[32]

In solchen Fällen sehe ich jedoch keinen Grund zur Abschwächung des Analogieeinflusses. Dieser sollte hingegen — und dann aber radikal — abgebaut werden, wenn diese Grade verschieden sind. Sollte sich aber ein gewisser Analogieeinfluß dennoch als vorteilhaft erweisen, so könnte man eine Funktion $c^{L_1|L_2;E}$ in Erwägung ziehen, die zusätzlich von einem Parameter E mit $0 < E < 1$ abhängt und die folgendermaßen festgelegt ist:

$$c^{L_1|L_2;E}(\Sigma, T) = E \cdot c^{L}(\Sigma, T) + (1 - E) \cdot c^{L_1|L_2}(\Sigma, T),$$

wobei c^{L} die für den Gegenstandsbereich hinsichtlich der beiden (*zusammengeworfenen*) Grundeigenschaften faire induktive Methode ist. Damit ist jedoch nicht ausgeschlossen, daß ein Analogieeinfluß nicht vielleicht auf andere Weise besser gemessen werden könnte.

---

[32] Ich ziehe es vor, hier nicht a priori von einer Familie von Prädikaten auszugehen, sondern eher im nachhinein zu verlangen, daß verschiedene V-Eigenschaften leer sind. Es scheint mir eine Tatsache zu sein, daß man es gelegentlich mit verschiedenfarbigen Kugeln zu tun hat.

VII. Über die Voraussetzungen der Induktion

1. Das Problem der zerrütteten Prädikate

Es ist gezeigt worden, daß man bei Kenntnis des Uniformitätsgrades des Gegenstandsbereichs hinsichtlich der in Erwägung gezogenen Eigenschaften (oder, was das gleiche besagt: des Homogenitätsgrades dieser Eigenschaften bezüglich des Gegenstandsbereichs) in der Lage ist, jene L-Methode c^L zu ermitteln, die für diese Situation (und für die Sprache, die sie repräsentiert) fair ist, die also bei der Anwendung einen größeren Erfolg verspricht als alle anderen adäquaten induktiven Methoden. Wenn uns hingegen der tatsächliche Uniformitätsgrad unbekannt ist, was der Normalfall sein dürfte, so müssen wir trotzdem, um eine adäquate induktive Methode anwenden zu können, von einer bestimmten *Annahme* über den Grad der Uniformität ausgehen; je näher dieser angenommene Grad beim tatsächlichen liegt, desto größer wird der Erfolg bei der Anwendung der entsprechenden induktiven Methode sein.

Wenn der Gegenstandsbereich unendlich oder praktisch unendlich (d. h. sehr groß im Hinblick auf das verfügbare Erfahrungsdatum) ist, dann sind alle Uniformitätsgrade w mit $1/q \leqslant w < 1$ deduktiv und induktiv mit dem Erfahrungsdatum verträglich (die Wahl $w = 1$ nur, wenn alle beobachteten Gegenstände die gleichen Eigenschaften besitzen); es gibt in diesem Fall also keine *theoretischen* (wenn auch vielleicht manche praktische) Gründe dafür, eine bestimmte Wahl w *nicht* zu treffen. Denn je mehr das Beobachtungsdatum den Gegenstandsbereich (falls dieser endlich ist) ausschöpft, desto kleiner wird das Intervall der in Frage kommenden Werte w, das um den Wert w′ liegt, der entsteht, wenn man den Gegenstandsbereich auf die Gesamtheit der *beobachteten* Dinge beschränkt. Sind alle Dinge beobachtet worden, so enthält dieses Intervall nur noch w′. Der einzige Weg zur endgültigen und unbezweifelbaren Festlegung des Uniformitätsgrades besteht also darin, alle Gegenstände des Bereichs hinsichtlich ihrer Eigenschaften zu untersuchen.

Jede induktive Methode c, ob sie nun im Kontinuum liegt oder nicht, ist durch gewisse Aprioribewertungen der Zustandsbeschreibungen charakterisiert; wenn sie darüber hinaus auch (II) (a) erfüllt, ist sie bereits durch die Aprioribewertungen der einzelnen statistischen Verteilungen (die objektive Wahrscheinlichkeiten beschreiben) eindeutig festgelegt. Es gilt demnach im Hinblick auf die deduktive Logik:
Der erste Unterschied zwischen der deduktiven und der induktiven Logik besteht darin, daß diese genötigt ist, solche Aprioribewertungen vorzunehmen, jene hingegen nicht. Jede adäquate induktive Methode ist darüber hinaus mit der Annahme eines Uniformitätsgrades gleichwertig, während die deduktive Logik ohne derartige Annahmen auskommt.
Dieser Unterschied ist nicht so zu verstehen, daß es demnach verkehrt war, die induktive Logik als eine Logik der partiellen Implikation zu verstehen, die als Grenzfall die deduktive Logik enthält.[1] Vielmehr kann man erst durch eine derartige Deutung des induktiven Arguments sehen, warum die deduktive Logik ohne solche Aprioribewertungen (und a fortiori ohne Annahmen über die Uniformität) auskommt: Wenn Ψ aus Φ logisch folgt, dann enthält der Spielraum von Ψ als Teilklasse den von Φ und dann ist der Spielraum von $\Psi \wedge \Phi$ mit dem von Φ identisch; nach D 3 ist dann 1 der c-Wert, gleichgültig, wie die Zustandsbeschreibungen bewertet werden, d. h. welche Aprioriwerte c den Elementen dieser Spielräume zuordnet. Analog verhält es sich bei logischer Unverträglichkeit, die den c-Wert 0 zur Folge hat. Dieser *vollständige Einschluß* des einen Spielraums im anderen bzw. dieser *vollständige Ausschluß* dieser beiden Spielräume bewirkt also, daß die tatsächlich vorgenommene Bewertung der Zustandsbeschreibungen uninteressant und ohne Bedeutung für die Berechnung des Grades der partiellen Implikation ist; notwendig ist sie allerdings, wenn der Spielraum der Prämisse Φ nicht Teilklasse des Spielraums der Hypothese Ψ ist: Dann hängt der c-Wert von eben diesen Aprioribewertungen ab. Wenn im folgenden von

[1] Normalerweise ist ja ein induktives Argument mit dem Wert 1 gleichwertig mit einer logischen Folgerung und eines mit dem Wert 0 gleichwertig mit logischer Unverträglichkeit; Ausnahmen davon erhält man, wenn in diesem induktiven Argument unendlich viele Gegenstände erwähnt werden oder wenn die Hypothese gewisse Aussagen über die Anzahl des Gegenstandsbereichs macht (ist diese Aussage wahr, erhält das Argument den Wert 1, und sonst den Wert 0).

Unterschieden zwischen deduktiver und induktiver Logik gesprochen wird, so ist dabei ebenfalls stets, so wie hier, nicht von einer Zurücknahme der Ausgangsbasis die Rede, sondern lediglich von deren Verdeutlichung.

Wenn die Anzahl der Gegenstands- und der Eigenschaftsausdrücke abzählbar unendlich ist (so wie dies auch bei Vollständigkeitsbeweisen für die einfache Typentheorie vorausgesetzt werden muß), dann gibt es eine umkehrbar eindeutige Entsprechung zwischen den einzelnen möglichen Uniformitätsgraden (den reellen Zahlen zwischen 0 und 1) und den einzelnen L-Methoden. Die Gesamtheit der L-Methoden ihrerseits ist durch die Bedingungen (I) bis (IV) und durch die Limeskonventionen charakterisiert. Einige dieser Forderungen, nämlich die Bedingungen (II) und (III), stehen und fallen mit der Forderung des Gesamtdatums, einer außersystematischen Regel also, die besagt, daß ein induktives Argument von einer Person x im allgemeinen nur angewandt werden darf bzw. von ihr im allgemeinen nur korrekt angewandt wird, wenn die Prämisse das gesamte Wissen formuliert, das x zur Verfügung steht bzw. das x akzeptiert (es sei denn, es steht schon fest, daß ein Teil dieses Wissens für die Hypothese im Hinblick auf den Rest irrelevant ist). Diese Forderung des Gesamtdatums ist nun allerdings nicht aufgestellt worden, um (II) und (III) zu rechtfertigen, sondern um Paradoxien von der Art des im 2. Abschnittes des II. Kapitels behandelten Beispiels auszuschließen; ist sie verletzt, so führt die Verwendung einer beliebigen regulären Bestätigungsmethode, die symmetrisch bezüglich der Gegenstandsausdrücke ist, zu eben diesen induktiven Widersprüchen; wenn man zugibt, daß bei induktiven Argumenten ein Zusatzwissen den Wahrscheinlichkeitsgrad im allgemeinen ändert, so wird man die Forderung des Gesamtdatums oder eine damit im wesentlichen gleichwertigen Bedingung als vernünftig ansehen müssen.

Der zweite Unterschied zwischen der deduktiven und der induktiven Logik besteht also darin, daß die Anwendung der induktiven Argumente jener Forderung des Gesamtdatums zu genügen hat, was für die Anwendung der deduktiven Schlüsse nicht gilt.

Für den Fall, daß Ψ von Φ logisch impliziert wird und damit gilt, daß $c(\Psi, \Phi) = 1$, ist ja der Spielraum von Φ in dem von Ψ eingeschlossen. Ein zusätzliches Wissen Γ kann den Spielraum der Prämisse

nur noch weiter einschränken, so daß also der Spielraum von $\Phi \wedge \Gamma$ echte oder unechte Teilklasse des Spielraums von Φ und damit wiederum Teilklasse des Spielraums von Ψ ist; also gilt dann weiterhin: $c(\Psi, \Phi \wedge \Gamma) = 1$. Ist hingegen Ψ mit Φ logisch unverträglich, so daß deren Spielräume elementfremd sind und $c(\Psi, \Phi) = 0$, so sind auch die von Ψ und $\Phi \wedge \Gamma$ elementfremd, da der Spielraum von Φ ja den von $\Phi \wedge \Gamma$ enthält; es gilt dann also ebenfalls: $c(\Psi, \Phi \wedge \Gamma) = 0$.
Ich habe bisher stets vorausgesetzt, daß die Objektsprache in einer gewissen Weise interpretiert ist, und habe gelegentlich auch Beispiele für derartige Deutungen angegeben. Hingegen bin ich nicht darauf eingegangen, ob diese Interpretationen gewissen Bedingungen zu genügen haben, und wenn ja, welchen. In der deduktiven Logik spielen derartige Fragen keine Rolle. Die Gültigkeit eines deduktiven Schlusses ist unabhängig davon, wie vernünftig oder wie verrückt die in den Sätzen Ψ und Φ vorkommenden Begriffe interpretiert sind. Die folgenden Beispiele zeigen, daß die Frage, was die Bedeutung der objektsprachlichen Begriffe ist, durchaus ein Problem der induktiven Logik ist.
Zur Bestimmung der faktischen relativen Weiten (also nicht der durch (II) (b) vorgenommenen apriorischen relativen Weiten) der Farbprädikate ist es wichtig, zu wissen, wie diese Begriffe definiert oder zumindest näher festgelegt sind. Definiert man den Begriff „rot" etwa unter Bezugnahme auf die Frequenz der elektromagnetischen Schwingungen, so ergibt sich für seine relative Weite vermutlich ein anderer Wert, als wenn diese durch den Quotient der Anzahl meiner Sehnerven, die *rot* wahrnehmen können, zur Anzahl meiner Sehnerven, die *Farben* wahrnehmen können, festgelegt wird, und wiederum andere relative Weiten erhält man sicherlich, wenn dieser Quotient bei anderen Personen ermittelt wird.
Als naheliegender und durchaus befriedigender Ausweg bietet sich hier der Vorschlag an, bei der Bestimmung der optimalen induktiven Methode auf die a priori gegebenen relativen Weiten der Begriffe zurückzugreifen und die Theorie über die faktischen relativen Weiten als Zusatzwissen den übrigen Erfahrungsdaten hinzuzufügen, also das Problem von der induktiven Logik auf die Einzelwissenschaften zu verschieben (wohin es auch tatsächlich gehört).
Doch auch die korrekte Bestimmung der apriorischen relativen Wei-

ten der Prädikate scheint gelegentlich schwierig zu sein. Eine Person x etwa sei *rot-grün-blind*, d. h. sie habe die gleichen Empfindungen, wenn Frequenzen, die für *rot*, und solche, die für *grün* charakteristisch sind, auf ihre Sehnerven treffen. Es erscheint dann auf den ersten Blick als adäquat, wenn x von den fünf Grundprädikaten „rot-grün", „orange", „gelb", „blau" und „violett" ausgeht. Die Wahrscheinlichkeitswerte, die x für eine Hypothese Ψ hinsichtlich eines Erfahrungsdatums Φ mit einer Methode c^L ermittelt, werden sich dann von denen einer Person y unterscheiden, die nicht rot-grün-blind ist und die von sechs Grundprädikaten ausgeht, und zwar auch hinsichtlich solcher Hypothesen, die an Prädikaten nur „orange", „gelb", „blau" und „violett" enthalten. Dieser Unterschied kann durch ein hinreichend großes Erfahrungsdatum überspielt und beliebig klein gemacht werden, doch ist damit das theoretische Problem nicht gelöst. Seine Lösung erreicht man, wenn man sich vergegenwärtigt, wie y die Situation des x betrachtet: Für x sind die beiden Eigenschaften (bzw. Klassen) *rot* und *grün* ja identisch; wenn y dieser Identitätsbehauptung seinen sonstigen Axiomen (bzw. den Prämissen aller induktiven Argumente) hinzufügt, so werden seine c^L-Werte mit denen des x identisch.

Größere Schwierigkeiten entstehen der induktiven Logik durch Eigenschaftsausdrücke und deren Interpretationen, die noch stärker als „rot-grün" von denen abweichen, die wir im Alltag wie auch in den Wissenschaften verwenden. Auf dieses Problem hat mit aller Deutlichkeit erstmals Nelson Goodman hingewiesen.[2] Gegeben sei etwa eine Urne mit 20 Kugeln, die alle einfarbig sind, und zwar entweder rot oder grün; „grün" ist in diesem speziellen Fall also durch „nicht rot" definierbar, so daß man „rot" als das einzige Grundprädikat ansehen kann. Die Sprache, die diese Situation beschreibt, enthalte die außerlogischen Ausdrücke „a_1", ..., „a_{20}" sowie „rot". Dann ist der Begriff „grot" definierbar als „rot und verschieden von a_{20}, oder nicht rot und mit a_{20} identisch". Die Identität wird dabei als logische Relation angesehen. Durch Postulate möge behauptet werden, daß die 20 Dinge voneinander verschieden sind, d. h. es liege eine sogenannte *strenge Identität* vor. Nun ist „$a_1 \in$ rot $\wedge \ldots \wedge a_{19} \in$ rot" per

---

[2] Vgl. Goodman [1], [2] und [3].

definitionem äquivalent mit „$a_1 \in \text{grot} \wedge \ldots \wedge a_{19} \in \text{grot}$“, während „$a_{20} \in \text{rot}$“ per definitionem äquivalent ist mit „$\neg (a_{20} \in \text{grot})$“. Dann gilt für jedes c^L:

$c^L(\text{„}a_{20} \in \text{rot}\text{“}, \text{„}a_1 \in \text{rot} \wedge \ldots \wedge a_{19} \in \text{rot}\text{“}) =$
$= c^L(\text{„}\neg (a_{20} \in \text{grot})\text{“}, \text{„}a_1 \in \text{grot} \wedge \ldots \wedge a_{19} \in \text{grot}\text{“})$
$= 1 - c^L(\text{„}a_{20} \in \text{grot}\text{“}, \text{„}a_1 \in \text{grot} \wedge \ldots \wedge a_{19} \in \text{grot}\text{“})$
$= 1 - (L/2)/(19 + L)$
$= (19 + L/2)/(19 + L).$

Es gilt damit für das definierte Prädikat „grot“:

$c^L(\text{„}a_{20} \in \text{grot}\text{“}, \text{„}a_1 \in \text{grot} \wedge \ldots \wedge a_{19} \in \text{grot}\text{“}) = (L/2)/(19 + L).$

Wählt man andererseits „grot“ als Grundprädikat, so ist ja „rot“ definierbar als „grot und verschieden von a_{20}, oder nicht grot und mit a_{20} identisch“. Unter Verwendung der gleichen Funktion c^L erhält man dann:

$c^L(\text{„}a_{20} \in \text{grot}\text{“}, \text{„}a_1 \in \text{grot} \wedge \ldots \wedge a_{19} \in \text{grot}\text{“}) = (19 + L/2)/(19 + L),$

und entsprechend für das hier definierbare Prädikat „rot“:

$c^L(\text{„}a_{20} \in \text{rot}\text{“}, \text{„}a_1 \in \text{rot} \wedge \ldots \wedge a_{19} \in \text{rot}\text{“}) = (L/2)/(19 + L).$

Nun ist für jede positive reelle Zahl L jedoch $(L/2)/(19 + L) \neq (19 + L/2)/(19 + L)$; lediglich für $L = \infty$ stimmen die Zahlenwerte überein. Man sieht unmittelbar, daß dieses von Goodman entdeckte Paradoxon keine spezifische Eigenheit der L-Methoden ist, sondern daß es für beliebige Funktionen c gilt, die in wenigstens einem Fall einen Wahrscheinlichkeitswert ergeben, der von der Aprioriwahrscheinlichkeit abweicht, daß es also alle von c^∞ verschiedenen Funktionen c betrifft. Zur Verdeutlichung sei der Begriff des Lernens aus der Erfahrung nochmals und etwas allgemeiner definiert, so daß die Analyse nicht nur für Funktionen, die auch (I), (II) und (III) erfüllen, adäquat ist, sondern allgemein gilt:

D 19+ c ermöglicht es, aus der Erfahrung zu lernen, genau dann, wenn für jedes V-Prädikat Π^*, für beliebige Gegenstandsausdrücke $\beta_1, \ldots, \beta_k$ (wobei $k + 1 < n$), für jede individuelle Verteilung Φ für die Gegenstandsausdrücke $\beta_1, \ldots, \beta_k$ bezüglich Π^* und für alle nicht in Φ vorkommenden Gegenstandsausdrücke α und β gilt:
(a) $c(\beta \in \Pi^*, \Phi \wedge \alpha \in \Pi^*) \not< c(\beta \in \Pi^*, \Phi)$, sowie
(b) $c(\beta \in \Pi^*, \Phi \wedge \alpha \in \Pi^*) \neq c(\beta \in \Pi^*, \Phi)$.

Wie das obige Beispiel zeigt, ergibt sich das Paradoxon der Abhängigkeit der Wahrscheinlichkeitswerte von der Wahl der Grundbegriffe, wenn Bedingung D 19+ durch c erfüllt ist, unabhängig davon, ob auch (a) gilt oder nicht gilt. Von einem deduktiven Widerspruch kann allerdings noch nicht gesprochen werden, da es ja denkbar, wenn auch gegen alle Intuition ist, daß der Wahrscheinlichkeitswert nicht von dem abhängt, was die Sätze aussagen, sondern davon, welche Begriffe Grundbegriffe und welche definiert sind.

Es ist jedoch nicht sinnvoll, sich mit diesem Sachverhalt resignierend zufriedenzugeben. Es ergeben sich nämlich keine derartigen Paradoxa, wenn man an Stelle von „rot" als Grundbegriffe „blau", „grün", „nicht rot", „rot oder blau" oder irgendeinen anderen der üblichen einfachen oder auch zusammengesetzten Begriffe wählt. Lediglich die Ersetzung von „rot" durch „grot" und ähnliche Begriffe, die man als „zerrüttet" bezeichnen kann, schaffen jene paradoxe Situation.

Das von Goodman aufgeworfene Problem dieser zerrütteten Prädikate ist nicht, wie es zunächst erscheinen mag, ein spezifisches Problem der induktiven Logik, sondern ein allgemeines und grundlegendes Problem der Erkenntnistheorie, das durch die Fragen „Wie erlernen wir erstmals eine Sprache?" und „Worüber sprechen wir mit der Sprache?" umrissen werden kann. Wenn ich eine Sprache, etwa die englische, nicht dadurch erlerne, daß mir Übersetzungsregeln in eine andere, mir schon bekannte, gegeben werden, so wird dies im allgemeinen so vor sich gehen, daß mir jemand die einzelnen Begriffe durch Hinweise näher zu bringen sucht, indem er z. B. auf bestimmte Gegenstände zeigt, und dabei jedesmal „red" sagt und auf eine Reihe von anderen Dingen hinweist und hierbei die Worte „not red" ausspricht; aus seinen Gebärden möge mir klargeworden sein, daß „not" für die Negation bzw. für das Komplement steht.[3] Er macht mich also, mit anderen Worten, mit einem Teil der *Extension des Begriffs* „red" und mit einem Teil der *Extension des Begriffs* „not red" bekannt und hofft, daß ich nun gelernt habe, *wie er dieses Begriffspaar gebraucht*,

[3] Ich muß also, um den korrekten Gebrauch von „red" zu erlernen, bereits die Sprache der Gebärden meines Mitmenschen hinreichend gut verstehen, nicht nur, um logische Partikel wie die Negation von deskriptiven unterscheiden zu können, sondern auch, um erkennen zu können, *worauf* er hinweist, daß er dabei nicht z. B. seine Fingerspitze oder die Berührung eines Gegenstandes oder die ihn dabei durchwallenden Gefühle meint.

daß ich nun weiß, was er mit den Begriffen „red" und „not red" *intendiert*, daß ich also die *Intension dieser Begriffe* erfaßt habe.[4]

Daß seine Hoffnung nicht sehr begründet ist, kann man folgendermaßen sehen: Es widerspricht keinem einzigen seiner Hinweise, wenn ich annehme, daß er mit „red" die Gesamtheit jener Gegenstände meint, mit der ich bestimmte Politiker ärgern kann, oder aber die Gesamtheit der Dinge, die in meinem Partner ein angenehmes Gefühl (etwa das der kommenden Weltrevolution) hervorrufen, oder schließlich die Gesamtheit der Objekte, die entweder vor dem Jahr 2324 auf ihre Farbe hin überprüft worden sind und sich als rot erwiesen haben oder aber, die vor diesem Zeitpunkt nicht auf ihre Farbe hin untersucht worden sind und die grün sind. Verschiedene dieser Intensionen können vielleicht durch weitere Hinweise ausgeschaltet werden, aber es ist wohl offensichtlich, daß damit immer noch eine unendliche Mannigfaltigkeit von möglichen Fehldeutungen offen bleibt. Damit ist aber auch entschieden, daß dieses Problem der Ausschaltung nicht intendierter Deutungen nicht mit den Mitteln der induktiven Logik gelöst werden kann, so wie sie in diesem Buch dargestellt worden ist. Andererseits stellt es zwar eine Schwierigkeit dar, die die induktive Logik *berührt*, aber *nicht* zur induktiven Logik *gehört*, und Carnap hat daher richtig gehandelt, als er sein System der induktiven Logik entwickelt hat, ohne die Lösung dieses Problems abzuwarten.

Der einfachste und naheliegendste Lösungsvorschlag, nämlich daß „rot" das Grundprädikat und „grot" der definierte Ausdruck ist, ist gleichzeitig auch der am einfachsten zu widerlegende: Wie schon erwähnt worden ist, kann man ja auch den Begriff „rot" durch den Ausdruck „grot" definieren, der in einer geeigneten Theorie das Grundprädikat ist. Die Axiomensysteme der Mathematik und der Naturwissenschaften haben die Unhaltbarkeit der Aristotelischen Vorstellung gezeigt, daß es in einer Disziplin *die* Grundsätze und *die* Grundbegriffe gibt; ein Axiomensystem kann immer durch ein mit ihm logisch äquivalentes ersetzt werden, das andere Grundsätze und

[4] Der Begriff „Bedeutung" wird üblicherweise mehrdeutig verwendet: Manchmal im Sinne von „Extension", gelegentlich auch im Sinne von „Intension" und meist synonym mit „Extension oder Intension"; ich möchte ihn daher in diesen Zusammenhängen nach Möglichkeit vermeiden.

andere Grundbegriffe enthält. Der Mythos, daß es Grundeigenschaften und zusammengesetzte Eigenschaften gibt und daß damit aus der Gesamtheit jener miteinander logisch äquivalenten Axiomensysteme eines ausgezeichnet ist, nämlich jenes, in dem die Grundbegriffe Grundeigenschaften und die definierten Begriffe zusammengesetzte Eigenschaften bezeichnen, ist aus dem gleichen Grund nicht haltbar, da man dann ja die Eigenschaft, *Grundeigenschaft* bzw. *zusammengesetzte Eigenschaft* zu sein, im System irgendwie ausdrücken müßte und da man auch zu diesem Axiomensystem ein äquivalentes finden könnte, in dem diese Begriffe nicht als Grundbegriffe auftreten.[5]

Die Forderung, keine Prädikate zuzulassen, die mittels Gegenstandsausdrücken definierbar sind, erweist sich in verschiedener Hinsicht als unhaltbar: Erstens ist ja auch der Begriff „rot" mittels „grot" und Gegenstandsausdrücken definierbar, so daß er also bei induktiven Argumenten nicht verwendet werden dürfte, was absurd ist; da auf ähnliche Weise jedes Prädikat definierbar ist, müßten danach alle Prädikate ausgeschlossen werden. Zweitens ist die Frage, welche Ausdrücke Grundbegriffe und welche definierte Begriffe sind, systemabhängig: Zu jedem Axiomensystem kann man ohne Schwierigkeiten ein dazu logisch äquivalentes konstruieren, das andere Grundbegriffe enthält.[6] Schließlich gibt es, wie Goodman hervorgehoben hat,[7] eine Reihe von Begriffen, die üblicherweise unter Bezugnahme auf gewisse Gegenstände, z. B. Raum-Zeit-Regionen, definiert werden, wie etwa „arktisch", „mittelalterlich", usw., die aber trotzdem ohne Zweifel in Sätzen vorkommen dürfen, die Prämissen oder Hypothesen induktiver Argumente darstellen. Die Lösung des von Goodman aufgezeigten Problems ist also offenbar nicht so einfach, wie man zunächst meinen könnte.

Goodmans eigener Lösungsvorschlag[8] geht davon aus, daß man beim Induktionsproblem nicht einen tabula-rasa-Standpunkt voraussetzen darf: Wir dürfen nicht davon abstrahieren, daß wir bei der Formulierung von Naturgesetzen und bei Schlüssen von Bekanntem auf Unbe-

[5] Carnap hat diesen Mythos in Carnap [7] vertreten und ihn in Carnap [8] auf Grund der Kritik in Goodman [2] wieder zurückgezogen.

[6] Vgl. hierzu auch die Bemerkungen von Goodman in [3], S. 78f., sowie Stegmüller [2], S. 308f.

[7] Vgl. Goodman [3], S. 150.

[8] Vgl. Goodman [3], S. 84ff.

kanntes tatsächlich gewisse Ausdrücke benützen oder andere tunlichst vermeiden; wir sagen dann, daß die einen Begriffe in der Sprache *gut* und die anderen in ihr *schlecht verankert* sind. Goodman formuliert dann Eliminationsregeln, die Aussagen mit schlecht verankerten Prädikaten, wie etwa „Alle Rubine sind grot", von induktiven Argumenten ausschließen, wenn sie (auf Grund der Tatsache, daß es Rubine gibt) Behauptungen mit gut verankerten Prädikaten widersprechen, wie etwa „Alle Rubine sind rot". — Man braucht den Goodmanschen Lösungsvorschlag nicht weiter zu entwickeln, um erkennen zu können, daß er auf die uns bekannten Sprachen relativiert ist: Er funktioniert nicht mehr, wenn einmal eine Sprache entwickelt wird, die zerrüttete Prädikate als Grundbegriffe enthält und in der mittels dieser Ausdrücke Gesetzesaussagen und induktive Argumente formuliert werden. Es ist nicht *logisch* ausgeschlossen, daß irgendwo Lebewesen mit einer solchen Sprache mehr oder weniger erfolgreich operieren; vielleicht kann es als *empirisch* gesichert gelten, daß dies nicht der Fall ist, aber *die Tatsache allein, daß die Adäquatheit des Goodmanschen Lösungsvorschlags von einer empirischen Hypothese abhängt, spricht gegen diesen Vorschlag.*[9]

Ähnlich muß man auch Carnaps Lösungsvorschlag bewerten.[10] Er setzt dabei voraus, daß c^* die adäquate Explikation des logischen Wahrscheinlichkeitsbegriffs ist. Es sei $\Delta_{j_{k+1}}$ der Satz $\alpha_{k+1} \varepsilon \Pi$, wobei Π die relative Weite j besitzt, und Γ_k die Aussage $\alpha_1 \varepsilon \Pi \wedge \ldots \wedge \alpha_k \varepsilon \Pi$. Dann ist Π in bezug auf c^* ein nicht zerrüttetes Prädikat und für induktive Argumente tauglich genau dann, wenn für alle natürlichen Zahlen k gilt: $c^* (\Delta_{j_{k+2}}, \Phi_{k+1}) > c^* (\Delta_{j_{k+1}}, \Phi_k)$.

Nun ist Carnap jedoch seit 1942 von jener Voraussetzung, daß c^* die einzige adäquate Explikation des logischen Wahrscheinlichkeitsbegriffs ist, abgerückt und hat stattdessen ein Kontinuum von adäquaten induktiven Methoden in Betracht gezogen. Es ist daher eingewendet worden, daß vermutlich auch die obige Festlegung des (metasprachlichen) Prädikats „nicht zerrüttetes Prädikat" inadäquat sei, da nach ihr ja ein Prädikat vielleicht in bezug auf c^{L_1} zerrüttet, in

---

[9] Vgl. das analoge Argument bei Stegmüller in [2], S. 305.

[10] Vgl. Carnap [7], S. 146f.; Carnap ist an dieser Stelle allerdings nicht ganz klar (vor allem ist nicht klar, was „in all cases of this kind" bedeutet), und so kann ich nur hoffen, ihn hier richtig zu interpretieren.

bezug auf c^{L_2} hingegen nicht zerrüttet ist.[11] Dieser Fall kann für L-Methoden jedoch nicht eintreten: Ein Blick auf die Theoreme 14 und 15 genügt, um erkennen zu können, daß sich alle Funktionen c^L mit $0 < L < \infty$ in dieser Hinsicht genauso verhalten wie c^*. Trotzdem ist Carnaps Vorschlag inadäquat, da sich dagegen im wesentlichen das gleiche sagen läßt wie gegen Goodmans Theorie: Wenn die zugrundegelegte Objektsprache Ausdrücke wie „grot" als Grundbegriffe enthält, so folgt aus der obigen Definition, daß „grot" in bezug auf c^* (und in bezug auf jede andere Funktion c^L) nicht zerrüttet ist. Dieser Vorschlag ist also ebenfalls auf die uns bekannten Sprachen relativiert und setzt für seine Gültigkeit die (sicher falsche) empirische Hypothese voraus, daß keine Sprache entwickelt werden kann, deren Grundbegriffe zerrüttet sind.

Daß das Problem der zerrütteten Prädikate die Adäquatheit der Voraussetzung der unabhängigen Wiederholbarkeit eines Versuchs und damit die der Bedingung (II) (a) berührt, hat meines Wissens erstmals Jürgen Humburg gesehen.[12] Wenn nämlich die Ergebnisse eines Experiments mittels zerrütteter Prädikate beschrieben werden, ist es nicht mehr in allen Fällen adäquat, von der unabhängigen Wiederholung eines Versuchs zu sprechen. Unrichtig ist aber die von ihm intendierte Verallgemeinerung, daß genau für Sprachen mit nichtzerrütteten Prädikaten die Bedingung (II) (a) adäquat ist (bei vorausgesetzter Adäquatheit der Bedingungen (I), (II) (b) und (III) bzw. (III')): (1) Es gibt Sprachen, die über einem Gegenstandsbereich interpretiert sind, der hinsichtlich seiner Eigenschaften durchgehend disuniform ist, so daß c^∞ adäquat ist; auch wenn die Grundprädikate zerrüttet sind, ist c^∞ optimal, wie schon gezeigt worden ist; da c^∞ die Bedingung (II) (a) erfüllt, ist diese also zumindest gelegentlich in Fällen adäquat, bei der die Grundprädikate der Sprache zerrüttet sind. Ferner ist, wie noch gezeigt wird, (II) (a) bei zerrütteten Prädikaten adäquat, wenn die Sprache sich auf einen Gegenstandsbereich bezieht, der nicht mehr als zwei Elemente enthält. (2) Es gibt nicht nur Sprachen mit zerrütteten Prädikaten, hinsichtlich derer (II) (a) adäquat ist, sondern es gibt umgekehrt auch Sprachen mit nichtzerrütteten Prädikaten, hinsichtlich derer (II) (a) nicht adäquat ist; um dies einzu-

11 Vgl. Stegmüller [2], S. 311.
12 Vgl. Humburg [1], S. 59.

sehen, genügt es, Sprachen in Betracht zu ziehen, die so ausdrucksarm sind, daß in ihnen nicht alle für induktive Argumente relevanten Zusammenhänge zwischen den Individuen des Gegenstandsbereichs formulierbar sind. Ein solcher Fall liegt z. B. vor, wenn die Sprache neben den Gegenstandsausdrücken und logischen Konstanten nur die einstelligen Grundprädikate $\Pi_1, \ldots, \Pi_6$ besitzt, wenn sie über dem Bereich der Würfe mit einem bestimmten Würfel interpretiert ist und dieser Würfel einen Mechanismus enthält, der den Schwerpunkt von Wurf zu Wurf verändert und zwar so, daß diese Änderung nur in der Sprache der Differential- und Integralgleichungen formuliert werden kann.

Obwohl ich der Ansicht bin, daß die Lösung des Problems der zerrütteten Prädikate einfach ist, ist sie doch nicht so einfach, wie man sich das gelegentlich vorgestellt hat.

2. Auf dem Wege zur Lösung dieses Problems

Wenn man bereit ist, aus der Erfahrung zu lernen, und wenn man sich vergegenwärtigt, daß alle bisherigen Lösungsversuche erstens von sehr allgemeinen Annahmen oder Prinzipien ausgegangen sind und zweitens als gescheitert betrachtet werden können, so wird man auch einem weiteren Lösungsvorschlag dieser Art keine große Chance einräumen. Ich ziehe es daher vor, am konkreten Beispiel den Weg zu einer Lösung zu suchen, und zwar an einem möglichst einfachen Beispiel, das nicht von Zusatzproblemen belastet ist.

Gegeben sei ein Gegenstandsbereich, der aus zwei Elementen a_1 und a_2 besteht, sowie eine Eigenschaft dieser Dinge. Zur Beschreibung dieser Miniaturwelt genügt eine Sprache, die an außerlogischen Begriffen nur α_1, α_2 und Π enthält. In ihr sind die folgenden Zustandsbeschreibungen formulierbar:

$\alpha_1 \varepsilon \Pi \wedge \alpha_2 \varepsilon \Pi$ [Abkürzung: $_2\Sigma_1$]
$\alpha_1 \varepsilon \Pi \wedge \neg (\alpha_2 \varepsilon \Pi)$ [Abkürzung: $_2\Sigma_2$]
$\neg (\alpha_1 \varepsilon \Pi) \wedge \alpha_2 \varepsilon \Pi$ [Abkürzung: $_2\Sigma_3$]
$\neg (\alpha_1 \varepsilon \Pi) \wedge \neg (\alpha_2 \varepsilon \Pi)$ [Abkürzung: $_2\Sigma_4$]

Es wird aus irgendwelchen Überlegungen heraus geschätzt, daß $^3/_4$

der Uniformitätsgrad dieses Bereichs hinsichtlich seiner Eigenschaft ist (die Schätzung ist eine gewogene Mittelung der einzelnen möglichen Grade und braucht deshalb nicht mit einem von diesen identisch zu sein). Als optimale induktive Methode empfiehlt sich dann $_2c^1$. Diese Methode gibt den Zustandsbeschreibungen folgende Aprioriwerte:

$_2c^1\,(_2\Sigma_1, T) = {}_2c^1\,(_2\Sigma_4, T) = {}^3/_8$,

$_2c^1\,(_2\Sigma_2, T) = {}_2c^1\,(_2\Sigma_3, T) = {}^1/_8$.

Bei der Ermittlung des Uniformitätsgrades und der fairen induktiven Methode ist stillschweigend davon ausgegangen worden, daß Π unter der Interpretation J eine nichtzerrüttete Eigenschaft, etwa *rot* [kurz: *F*], bezeichnet; dann beschreiben $_2\Sigma_1$ und $_2\Sigma_4$ einen jeweils uniformen Gegenstandsbereich hinsichtlich F. J′ sei eine Interpretation der Sprache, die sich von J nur hinsichtlich Π unterscheidet und die Π die Eigenschaft (*rot und von a_2 verschieden*) *oder* (*nicht rot und mit a_2 identisch*) [kurz: *F′*] zuordnet; dann beschreiben $_2\Sigma_1$ und $_2\Sigma_4$ hinsichtlich F sowie $_2\Sigma_2$ und $_2\Sigma_3$ hinsichtlich F′ einen uniformen Gegenstandsbereich. Da wir annehmen, daß die Uniformität primär hinsichtlich F und nicht hinsichtlich F′ zu ermitteln ist, sehen wir, mit $_2c^1$ in Gedanken, für diesen Fall eine Funktion $_2c$ als optimal an, die durch die folgende Bewertung der Zustandsbeschreibungen charakterisiert ist:

$_2c\,(_2\Sigma_1, T) = {}_2c\,(_2\Sigma_4, T) = {}^1/_8$,

$_2c\,(_2\Sigma_2, T) = {}_2c\,(_2\Sigma_3, T) = {}^3/_8$.

Von Funktionen dieser Art ist im IV. Kapitel gesagt worden, daß sie *anti-induktiv* sind: Es ist ja $_2c\,(\alpha_2 \in \Pi, \alpha_1 \in \Pi) = {}^1/_4 < {}^1/_2 = {}_2c\,(\alpha_2 \in \Pi, T)$, so daß also ein positives Zusatzdatum $\alpha_1 \in \Pi$ den Wahrscheinlichkeitsgrad für die Hypothese $\alpha_2 \in \Pi$ herabsetzt. Diese Funktion $_2c$ verletzt also die Forderung (IV) (a), während sie die übrigen Forderungen [und insbesondere die Bedingung (II) (a)] erfüllt.

Im Zusammenhang mit dem Problem der zerrütteten Prädikate sind, was bisher nirgendwo geschehen ist, drei Fälle auseinanderzuhalten:

(1) Die fraglichen Eigenschaften weisen im Hinblick auf den Gegenstandsbereich keine oder nur eine geringe Uniformität auf. Wenn der Gegenstandsbereich etwa aus den zu einer bestimmten Zeit im Auditorium Maximum der Universität München versammelten (z. B. 500) Studierenden besteht und die Eigenschaft *Erstgeborener oder*

Erstgeborene betrachtet wird, so dürfte diese hinsichtlich des zugrundegelegten Bereichs ein Minimum an Homogenität aufweisen, so daß wir uns bei einer Stichprobe von etwa 50 oder auch 100 Personen dieser Gruppe, die sich alle (bzw. fast alle) als Erstgeborene erweisen, nicht zu dem Schluß verleiten lassen, daß alle (bzw. fast alle) der 500 Hörer Erstgeborene sind.

(2) Die der Sprache zugrundegelegten Prädikate (und die Eigenschaften, für die sie stehen) sind zerrüttet, und zwar so, daß eine effektive Regel zur Erzeugung des zerrütteten Prädikates aus unzerrütteten Begriffen sowie auch ein Verfahren zur Behebung der Zerrüttung angegeben werden kann. „grot" ist ein derartiges Prädikat, aber auch der durch den Satz „(x ist rot und von den Dingen a_2, a_5 und a_9 verschieden) oder (x ist grün und mit einem der Dinge a_2 und a_5 identisch) oder (x ist blau und mit a_9 identisch)" definierbare Eigenschaftsausdruck oder gar der durch „(a_k ist rot und k ist keine Primzahl) oder (a_k ist nicht rot und k ist eine Primzahl)" definierbare Begriff.

(3) Die der Sprache zugrundegelegten Prädikate (und die Eigenschaften, für die sie stehen), sind irreparabel zerrüttet, da zwar ein Verfahren zur Erzeugung der Zerrüttung existiert, da es jedoch nicht im strengen mathematischen Sinn effektiv ist. Gegeben sei ein Relais-Experiment, das darin besteht, daß gleichzeitig aus einer bestimmten Urne eine Kugel gezogen und eine Münze geworfen wird; die Urne enthalte 10000 Kugeln, und die Münze werde genauso oft geworfen. Die Kugeln der Urne sind rot oder schwarz (= nicht rot), beim Münzwurf kann Kopf oder Wappen oben liegen. Es wird nur der aus den Kugeln der Urne bestehende Gegenstandsbereich und die Eigenschaft betrachtet, die durch den Satz „beim k-ten Zug ist die Farbe der Kugel rot und beim k-ten Wurf der Münze liegt Wappen oben, oder beim k-ten Zug ist die Farbe der Kugel nicht rot und beim k-ten Wurf der Münzen liegt nicht Wappen oben" zum Ausdruck kommt.

Die Lösung für den Fall 1 ist bereits gegeben worden: Da die Uniformität hier minimal oder fast minimal ist, ist auch c^{∞} oder eine Funktion c^{L} mit einer sehr großen reellen Zahl L fair. Hier ist jedoch, wie immer bei der Verwendung einer induktiven Methode, Vorsicht am Platze. Gegeben sei ein Gegenstandsbereich, der aus den 20 in einem Zimmer sitzenden Personen besteht. Meine Untersuchungen an 8

dieser Leute haben das Ergebnis, daß alle von ihnen augenkrank sind. Wenn es sich bei den Elementen des Gegenstandsbereichs offensichtlich um Hörer einer bestimmten Geographievorlesung handelt, so werde ich den Ergebnissen meiner Untersuchung keine große Relevanz beimessen, d. h. ich werde c^{∞} oder eine Funktion c^{L} mit großem L als optimale induktive Methode ansehen. Wenn sich hingegen mein Verdacht bestärkt, daß ich mich im Wartezimmer eines Augenarztes befinde, so werde ich sofort einen anderen Uniformitätsgrad schätzen, etwa 1 oder einen Wert, der nur wenig darunter liegt, und entsprechend zur Methode c^{0} oder zu einer Funktion c^{L} mit kleinem L übergehen. Wenn es sich hingegen schließlich herausstellt, daß der Gegenstandsbereich aus den zu einer Kabinettsitzung versammelten Mitgliedern der Bundesregierung besteht, so werde ich mich vielleicht für einen mehr dazwischenliegenden Uniformitätsgrad und entsprechend etwa für die Methode c^{1} oder für c^{15} oder auch für $c^{243,7}$ entscheiden.

Vergleichsweise einfach ist die Lösung für den Fall 3: Eine eventuell vorhandene Homogenität der Eigenschaft (im Beispiel: der Eigenschaft *rot*) ist vollständig zerstört worden. Der Gegenstandsbereich besitzt demnach hinsichtlich der neuen Eigenschaft ein Minimum an Uniformität, so daß auch in diesem Fall die induktive Methode c^{∞} (die genau genommen ja *a-induktiv* ist, da sie keinen Schluß von Bekanntem auf Unbekanntes gestattet) optimal ist.

Schwieriger ist die Lösung des Problems im Fall 2. Hier hat man, ausgehend von Eigenschaften mit einem gewissen Homogenitätsgrad (und entsprechend von Bereichen mit dem gleichen Uniformitätsgrad), durch gewisse Operationen neue Eigenschaften erhalten, die sozusagen *anti-homogen* sind (so daß der Gegenstandsbereich im Hinblick auf sie *anti-uniform* ist); in solchen Situationen erweisen sich daher *anti-induktive* Bestätigungsmethoden als optimal. Zu klären bleibt die Frage, warum wir im einen Fall *induktive* Methoden (vielleicht sogar adäquate induktive Methoden, d. h. L-Methoden) und im anderen Fall *anti-induktive* Methoden als optimal ansehen.

Dazu ist es vielleicht ganz nützlich, sich zu vergegenwärtigen, was unter dem Begriff „Begriff" üblicherweise verstanden wird. Ein Begriff ist zunächst einmal ein sprachlicher Ausdruck; gelegentlich, vor allem in Untersuchungen zur deduktiven Logik, werden die Ausdrücke „Begriff" und „Ausdruck" synonym verwendet. Hin und

wieder versteht man unter „Begriff" auch das gleiche, wie „das, was ein Ausdruck bedeutet (wofür er unter der gegebenen Interpretation steht, was er bei ihr bezeichnet)" oder das gleiche wie unter „Ausdruck und seine Bedeutung". Das, wofür ein Ausdruck einer Sprache (bei einer vorgegebenen Interpretation) steht, ist nun nicht der sogenannte Begriffsinhalt, sondern ist seine Extension (der Gegenstandsbegriff „Abendstern" etwa bezeichnet bei der vorgegebenen Interpretation der deutschen Sprache nicht irgendeine Intension, sondern einen ganz bestimmten Himmelskörper).

Die von Goodman entdeckten Paradoxa zeigen nun, daß es bei der Anwendung der induktiven Logik, d. h. bei der Auswahl einer bestimmten Bestätigungsmethode in einer vorgegebenen Situation, nicht genügt, dem Begriff der partiellen Implikation eine Objektsprache ohne zusätzliche Bedingungen zugrundezulegen, d. h. die Prädikate schlicht als gewisse graphische Zeichen und als sonst nichts anzusehen. Wenn man hingegen die durch die Interpretation gegebenen Extensionen der Ausdrücke mit in Betracht zieht, erhält man einerseits zuviel und andererseits immer noch zu wenig: zuviel, weil die Anwendung einer induktiven Methode sinnlos wird, wenn man über den Wahrheitswert der Hypothese bereits Bescheid weiß, zu wenig aber, weil damit für die Lösung des Problems nichts gewonnen ist: Der Wert von c („$a_{20} \in F$", „$a_1 \in F \wedge \ldots \wedge a_{19} \in F$") ist unabhängig davon, ob „$a_{20} \in F$" bei der vorgegebenen Interpretation J wahr ist oder nicht, ob die Gesamtheit, die „F" unter J bezeichnet, also mit B identisch ist oder nicht, sondern ist vielmehr der Grad, in dem ich als rationaler Mensch berechtigt bin, auf Grund dessen, was ich über den Bereich B und die in Betracht gezogenen Eigenschaften weiß, an die Hypothese „$a_{20} \in F$" zu glauben. Der c-Wert ist, mit anderen Worten, davon abhängig, ob „F" eine Abkürzung für „rot" oder aber für „grot" ist, während er unabhängig davon ist, ob das Ding a_{20} *rot* oder *nicht rot* (d. h. *grot*) ist.

Philosophen verwenden den Ausdruck „Begriff" oft jedoch in einer ganz anderen Weise. So sagt etwa Kant, daß unsere Anschauungen *durch Begriffe geordnet werden*, und nach Wittgenstein *verstehen wir einen Begriff genau dann, wenn wir wissen, wie wir ihn zu gebrauchen haben*. Durch einen Begriff ist nach diesen Ansichten also ein (vollständiges oder unvollständiges, mechanisch entscheidbares oder oft

nur unter großen Schwierigkeiten anwendbares) Verfahren gemeint, durch das entschieden werden kann, ob ein Gegenstand Element der Klasse ist oder nicht, die der Begriff bei der vorgegebenen Interpretation der Sprache bezeichnet. Im Sinne Hilberts ist dieses Verfahren mit der axiomatischen Charakterisierung der Grundbegriffe identisch. Mir will scheinen, daß man mit dem bisher noch nirgendwo präzisierten Begriff der Intension nichts anderes als ein solches Verfahren meint.[13] Die Axiome der Objektsprache legen aber nur die *Form* der Interpretation J und die logische Struktur der Identität, die die deskriptiven Ausdrücke durch J bezeichnen, fest, während diese Entitäten die *Inhalte* dieser Form sind, oder kurz: die Begriffsumfänge (die Extensionen) sind die Inhalte der Begriffsinhalte (der Intensionen), und die Intensionen sind die Formen der Extension.

In diesem Sinn stellt ein Begriff (ein Ausdruck zuzüglich seiner Intension) eine *Ordnung* des Gegenstandsbereichs dar; mit dem Begriff „rot" etwa habe ich ein Verfahren zur Hand, einen amorphen Bereich in zwei Klassen, nämlich die der roten und die der nichtroten Dinge, aufzuteilen und damit zu ordnen.[14] Diese Ordnung wird durch den Begriff „grot" nun teilweise zerrüttet und in ihr Gegenteil verkehrt, zu einer *Anti-Ordnung* gemacht. Ist die Art der Zerrüttung bekannt, ist also bekannt, wie die Anti-Ordnung aus der Ordnung, wie die Anti-Uniformität aus der Uniformität entstanden ist, so kann unter Zuhilfenahme einer der Uniformität entsprechenden L-Methode eine Bestätigungsmethode charakterisiert werden, die in der vorgegebenen Situation optimal ist; auch hier gilt, daß es nicht *eine* optimale Methode für *alle* möglichen Arten der Zerrüttung gibt, sondern daß man in *jedem* einzelnen Fall nach *einer* (nämlich der der Anti-Uniformität entsprechenden) optimalen induktiven Methode suchen muß. Dieses Wissen um die Anti-Ordnung und die Wahl einer entsprechenden Bestätigungsmethode ist die Lösung des Falles 2.

---

[13] Ich habe diese Gedanken in Essler [3] weiter ausgeführt.

[14] Diese Ordnung kann im Sinne des Phänomenalismus vorgenommen werden, indem in jedem einzelnen Fall entschieden wird, ob ein Ding mit einem vorgegebenen Standardobjekt ähnlich ist (die Ähnlichkeitsrelation wird dabei axiomatisch festgelegt), sie kann aber auch physikalistisch vorgenommen werden, indem bezüglich jedes einzelnen Dinges entschieden wird, ob es elektromagnetische Wellen von einer bestimmten Länge aussendet oder nicht (hier werden große Teile der Axiome der Physik als gültig vorausgesetzt).

Nicht immer wird die Objektsprache so ausdrucksreich sein, daß in ihr dieses Wissen um die Ordnung oder um die Anti-Ordnung formuliert werden kann. Wenn sie etwa eine Sprache der Junktorenlogik oder der engeren Quantorenlogik ist und an Prädikaten nur den Eigenschaftsausdruck Π enthält, so wird, wenn Π unter der vorgegebenen Interpretation die Gesamtheit der roten Dinge des Gegenstandsbereichs bezeichnet, die Intension von Π nicht in ihr ausdrückbar sein. Ist die Sprache hingegen ausdrucksreich genug, um diese Zusammenhänge zu beschreiben, so sind diese ihre Aussagen in der induktiven Logik gleich den Postulaten der Logik zu behandeln: Es sind entweder überhaupt nur Zustandsbeschreibungen zu formulieren, die mit ihnen logisch verträglich sind, oder die Methode c ist so einzuführen, daß sie allen Zustandsbeschreibungen, die diesen Axiomen widersprechen, den Aprioriwert 0 zuordnet, oder aber dieses Wissen um die Homogenität bzw. Anti-Homogenität der Prädikate ist im Sinne der Forderung des Gesamtdatums jeder Prämisse eines induktiven Argumentes konjunktiv hinzuzufügen.

Ein solches Wissen um die Homogenität bzw. Anti-Homogenität der Begriffe würde Kant vermutlich als ein Beispiel für ein synthetisch-apriorisches Urteil angesehen haben, während Carnap es eher den Analytizitätspostulaten zurechnen wird. Es ist hier nicht sehr sinnvoll, über Begriffe zu streiten, insbesondere nicht darüber, ob es sich *in Wahrheit* um analytische oder um synthetische Behauptungen handelt — eine solche Wahrheit gibt es nicht, wohl aber gibt es zweckmäßige und unzweckmäßige Festsetzungen über die Verwendung der Begriffe „analytisch" und „synthetisch". Ohne Zweifel aber ist jenes Wissen ein apriorisches Wissen: auch hier nicht im Sinne einer unumstößlichen Wahrheit, da in den Einzelwissenschaften dieses Aprioriwissen je nach Zweckmäßigkeit beibehalten oder aber durch ein anderes ersetzt wird, und auch nicht in dem Sinn, daß man es formuliert, noch bevor man etwas über die Welt ausgemacht hat; vielmehr akzeptiert man es, weil es sich als zweckmäßig, oder aber, weil sich ein alternatives Wissen als unzweckmäßig erwiesen hat.[15] Apriorisch ist

---

[15] So hat sich das Aprioriwissen über die Euklidizität des physikalischen Raumes als unzweckmäßig erwiesen und ist deshalb von Physikern durch ein anderes ersetzt worden.

dieses Wissen jedoch, wenn man einen Wissenszweig axiomatisch charakterisiert und so die Intensionen der Grundbegriffe festlegt, und wenn man sodann beginnt, dieses System deduktiv und induktiv zu verwenden.

Mir scheint, daß das von Goodman aufgeworfene Problem damit auf eine sehr einfache Weise gelöst ist, auf eine Weise, die naheliegt, wenn man sich die Schwierigkeiten an geeigneten Beispielen vergegenwärtigt. Daß diese Lösung bisher noch von keinem Logiker gefunden worden ist, liegt vermutlich daran, daß außer Heinrich Scholz und seinen Schülern kein Logiker bisher darangegangen ist, ein adäquates System der Semantik zu entwerfen, in dem streng zwischen Zeichen und Bezeichnetem sowie zwischen Extension und Intension unterschieden wird und in dem der Interpretationsbegriff seinen ihm gebührenden Platz hat, und daß die Mitglieder der Schule von Münster sich um die von Goodman entdeckten Probleme nicht sonderlich gekümmert haben. Andererseits scheint es im Lichte dieser Lösung, daß sie von den meisten *intendiert* worden ist, ohne daß dabei die richtigen Worte gefunden worden sind. Carnap etwa hat in diesem Zusammenhang ausdrücklich auf die Forderung des Gesamtwissens aufmerksam gemacht,[16] ohne dann allerdings diesen Gedanken weiter zu verfolgen, und nach Goodman darf man nicht von einem Standpunkt des vollständigen Nichtwissens ausgehen, sondern hat zu berücksichtigen, welche Naturgesetze bereits formuliert worden sind. Die Vertreter der mathematischen Wahrscheinlichkeitslehre hingegen können für sich in Anspruch nehmen, dieses Problem noch vor seiner Formulierung gelöst zu haben, da sie stets und überall vorausgesetzt haben, daß eine eindeutige Beschreibung des betreffenden Experiments vorliegt und da eine derartige Beschreibung doch als ein System von Axiomen aufgefaßt werden kann, das die zur Beschreibung benötigten Begriffe implizit charakterisiert.

Der dritte Unterschied zwischen der deduktiven und der induktiven Logik besteht somit darin, daß in dieser bei der Anwendung die Intensionen der Grundbegriffe angegeben werden müssen, was in jener nicht nötig ist; selbstverständlich bleibt die induktive Logik damit extensional und wird nicht zu einer sogenannten intensionalen Logik, da

---

[16] Vgl. Carnap [7], S. 140.

diese Intensionen nichts anderes als gewisse Extensionen (Klassen von Sätzen) sind.[17] Dieser Unterschied ist, genaugenommen, ein Spezialfall des zweiten, da er ebenfalls auf der Forderung des Gesamtdatums basiert; wegen seiner Wichtigkeit habe ich ihn jedoch nochmals besonders hervorgehoben.

Man wird mir, nicht ganz zu Unrecht, entgegnen, daß ich mit meinem Lösungsvorschlag das Problem nicht im strengen Sinn *gelöst* habe, sondern daß ich damit bestenfalls auf dem Wege zu seiner Lösung bin. Mit der Forderung, daß die Art und Weise, wie wir die Dinge begrifflich ordnen, als Axiome formuliert werden muß, ist ja noch nicht viel gewonnen. Wenn man etwa für den Begriff „rot" das folgende Axiom formuliert: „Alle Dinge sind rot oder nicht rot", so wird niemand sagen, daß dadurch die Intension von „rot" irgendwie festgelegt ist; daran ändert sich auch nichts, wenn man diesem Postulat noch den Grundsatz: „Wenn alle Dinge rot sind, dann gibt es rote Dinge" oder ähnliche Behauptungen hinzufügt. Fast immer bringt ein gelöstes Problem ein neues und noch zu lösendes mit sich; hier ist es die Frage, welchen Bedingungen ein Axiomensystem zu genügen hat, von dem gesagt werden darf, daß es die Intension eines Begriffs oder einer Gruppe von Begriffen festlegt.

Zunächst ist es naheliegend, die Forderung, daß ein Experiment und seine möglichen Ergebnisse eindeutig beschrieben sein müssen, als gleichwertig anzusehen mit der Bedingung, daß das Axiomensystem, das diese Beschreibung darstellt, vollständig zu sein hat. Es kann damit syntaktische oder semantische Vollständigkeit gemeint sein. Ein Axiomensystem T wird syntaktisch vollständig genannt genau dann, wenn von jedem Satz Φ, an dem an außerlogischen Begriffen nur die Grundbegriffe von T frei vorkommen, gilt, daß entweder er oder seine Negation aus T logisch folgt. T ist unter einer vorgegebenen Interpretation J semantisch vollständig genau dann, wenn von jedem Satz Φ, in dem an außerlogischen Begriffen nur die Grundbegriffe von T frei vorkommen, gilt, daß er nur dann bei J wahr ist, wenn er aus T logisch folgt.

Nun gilt jedoch bereits von einigen relativ elementaren Theorien, wie vom Axiomensystem der natürlichen Zahlen von Peano, daß sie un-

---

[17] Ausführlicher habe ich dies in Essler [2] begründet.

vollständig sind und nicht zu vollständigen widerspruchsfreien Systemen erweitert werden können.[18] Es kann daher nicht allgemein gefordert werden, daß die Intension durch ein semantisch vollständiges System wiedergegeben werden muß, auch wenn es einzelne semantisch vollständige Systeme gibt. Während die Bedingung der semantischen Vollständigkeit also zu stark ist, ist die der syntaktischen zu schwach: Wenn ein Axiomensystem syntaktisch vollständig ist, so folgt daraus nicht allgemein, daß es auch kategorisch ist.[19]

Es liegt daher nahe, die Kategorizität einer Theorie zu fordern. Ein Axiomensystem T wird dabei kategorisch genannt genau dann, wenn von zwei beliebigen Interpretationen J_1 und J_2, die beide sämtliche Sätze aus T wahr machen, gilt, daß sie isomorph oder strukturgleich sind, daß die Gegenstandsbereiche B_1 und B_2 von J_1 und J_2 also gleichzahlig sind und daß von jedem beliebigen Satz Φ, in dem an außerlogischen Begriffen nur die Grundbegriffe von T frei vorkommen, gilt, daß er unter J_1 genau dann wahr ist, wenn er es auch unter J_2 ist. Die Forderung der Kategorizität ist zunächst naheliegend, denn wenn von zwei verschiedenen Modellen J_1 und J_2 der Theorie T (von zwei verschiedenen, die Axiome von T wahrmachenden Interpretationen J_1 und J_2) gilt, daß sie isomorph sind, dann können zwei Menschen x_1 und x_2, die T akzeptieren und nur das Vokabular von T benützen, wobei x_1 T durch J_1 und x_2 T durch J_2 interpretiert, nicht feststellen, daß sie T verschieden deuten. Wenn T also nur isomorphe Modelle hat, dann ist sein Begriffsgerüst durch die Axiome eindeutig bestimmt.

Trotzdem ist die Forderung der Kategorizität nicht akzeptabel, da sie gleichzeitig zu schwach und zu stark ist. Gegeben sei etwa eine Theorie, die aus dem folgenden Axiom besteht: „Es gibt genau 50 Dinge, die F sind". Diese Theorie ist kategorisch; trotzdem wird niemand behaupten, daß damit die Intension des Begriffs „F" irgendwie

---

[18] Vgl. z. B. Kurt Gödel, Über formal unentscheidbare Sätze der principia mathematica und verwandter Systeme, Monatshefte für Mathematik und Physik 38 (1931) S. 173—198.

[19] Vgl. Hao Wang, A Survey of Mathematical Logic, Peking und Amsterdam 1963, S. 20. Die maximalkonsistente Menge von Sätzen, die Leon Henkin zum Beweis der Vollständigkeit der engeren Quantorenlogik konstruiert (vgl. auch Essler [1], S. 280—283), ist ein Beispiel für eine zwar syntaktisch vollständige, aber nicht kategorische Menge von Axiomen.

aufgehellt worden ist. Andererseits ist durch die Forderung der Kategorizität eines beliebigen vorgegebenen Axiomensystems bereits eindeutig festgelegt, welche Anzahl der Gegenstandsbereich eines jeden Modells dieses Systems haben muß, was einerseits unseren intuitiven Vorstellungen widerspricht, daß allgemeine Aussagen über die Natur nichts über die Anzahl der Dinge aussagen, und was andererseits den Übergang von endlichen zu unendlichen Sprachen mittels der Limeskonvention und umgekehrt den Übergang von unendlichen zu endlichen Sprachen unmöglich macht, da die Anzahl von B ja eindeutig festliegt (und zwar entweder endlich oder abzählbar unendlich oder aber von höherer Mächtigkeit ist). Diese Nachteile vermeidet man, wenn man statt der Kategorizität von solchen Axiomensystemen, die die Intension der Grundbegriffe beschreiben, die Kategorizität in einer gewissen Anzahl (der Dinge des Gegenstandsbereichs) verlangt. Ein Axiomensystem T ist dabei kategorisch in der Anzahl n genau dann, wenn zwei Modelle J_1 und J_2 von T, deren Gegenstandsbereiche B^1 und B^2 beide von der Anzahl n sind (wobei n jede endliche oder unendliche Anzahl sein kann), isomorph oder strukturgleich sind. Wenn T kategorisch in einer Anzahl n ist und J_1 sowie J_2 zwei verschiedene Modelle von T mit den n-zahligen Gegenstandsbereichen B^1 und B^2 sind, die von den Personen x_1 und x_2 benützt werden, so können auch hier x_1 und x_2, wenn sie nur die Begriffe von T verwenden, nicht feststellen, daß sie T verschieden interpretieren; x_1 gelangt dann zur Überzeugung, daß x_2 die Begriffe korrekt benützt, und umgekehrt. Es ist dann nicht möglich, daß x_1 und x_2 einen der Grundbegriffe strukturverschieden deuten, daß x_1 etwa den Ausdruck „F" im Sinne von „rot" und x_2 ihn im Sinne von „grot" auffaßt.

Dieser Begriff der Kategorizität in einer gewissen Anzahl ist auch verträglich mit der Art und Weise, wie wir Begriffe erlernen. Gegeben seien etwa die Begriffe „schwerer als" und „gleich schwer wie". Durch Hinweise sei eine Person x mit einer Gesamtheit von n (etwa mit 20) Objekten bekannt gemacht worden, und sie habe dabei die Extension dieser beiden Begriffe bei der Interpretation der Sprache über dieser Gesamtheit erfahren. Sie wird dann versuchen, die Struktur dieser Extension zu ermitteln, d. h. Sätze zu formulieren, die bei dieser Interpretation wahr sind, die keine Gegenstandsausdrücke ent-

halten und die in dieser Gesamtheit kategorisch in dieser Anzahl n (d. h. in 20) sind, und sie wird aus diesem Axiomensystem jene Aussagen herausnehmen, die für den vorgelegten n-zahligen Bereich spezifisch sind, d. h. die falsch werden, wenn der Gegenstandsbereich (und entsprechend die Gesamtheit der Gegenstandsausdrücke) erweitert wird.

Die Umgangssprachen unterscheiden sich von derartigen einfachen Modellsprachen dadurch, daß sie nicht aus einer Summe von solchen einfachen Axiomensystemen bestehen, sondern daß die einzelnen Begriffe mehr oder weniger lose (und zum Teil auch in einer schwer zu beschreibenden Weise) miteinander verknüpft sind. So stellen in ihr Aprioribehauptungen (d. h. Sätze, an deren Wahrheit wir bis auf weiteres festzuhalten gedenken) Beziehungen zwischen *schwerer als* und *gleich schwer wie* einerseits und *Balkenwaage*, *liegt auf*, *überwiegt*, *wiegt auf* und *gleich lang wie* andererseits her.

Der bis jetzt entwickelte Lösungsvorschlag kann aus folgendem Grund noch nicht voll befriedigen: Wenn ein Axiomensystem für einen Begriff „F" (etwa für „rot") kategorisch in einer Anzahl n ist, so kann man einen Ausdruck „G" auf folgende Weise definieren:
„$\Lambda\, x\, [x \in G \leftrightarrow (x \in F \wedge x \neq a_1) \vee (\neg\, (x \in F) \wedge x = a_1)]$".

Zu klären bleibt die Frage, wieso wir dann berechtigt sind, zu sagen, daß „F" ein unzerrüttetes und „G" ein zerrüttetes Prädikat ist. Die naheliegende Antwort, daß „F" ja ein Grundprädikat *und daher* unzerrüttet ist, kann nicht befriedigen; das vorgegebene Axiomensystem ist nämlich mit einem anderen logisch äquivalent, das die gleichen Axiome wie dieses besitzt, das jedoch an Stelle von „F" die Grundbegriffe „G" und „a_1" und für „F" folgende Definition hat:
„$\Lambda\, x\, [x \in F \leftrightarrow (x \in G \wedge x \neq a_1) \vee (\neg\, (x \in G) \wedge x = a_1)]$".

Die Lücke im Lösungsvorschlag ist geschlossen, wenn ein plausibler Grund dafür gegeben werden kann, warum wir ein Axiomensystem der ersten Art dem der zweiten Art vorziehen. Die Frage, warum wir von zwei äquivalenten Axiomensystemen das eine, aber nicht das andere verwenden, ist ein metatheoretisches Problem und kann nicht in der Sprache des Axiomensystems (der Objektsprache) beantwortet werden. Ich werde im 4. Abschnitt dieses Kapitels darauf zurückkommen. Zuvor möchte ich noch zwei eng mit dem Problem der zerrütteten Prädikate zusammenhängende Fragen erörtern, nämlich die

des Zusammenhangs zwischen den Bedingungen (IV) (a) und (II) (a) und die der adäquaten Einführung der Dispositionsprädikate.

Zur Verdeutlichung des Zusammenhangs zwischen (IV) (a) und (II) (a) möchte ich nochmals auf das zu Beginn dieses Abschnitts gebrachte Beispiel zurückkommen und es folgendermaßen abändern: Der Gegenstandsbereich bestehe nicht nur aus a_1 und a_2, sondern auch aus a_3; zu ihrer Bezeichnung enthalte die Sprache die Ausdrücke α_1, α_2 und α_3. Die den Überlegungen zugrundegelegte Eigenschaft werde durch Π ausgedrückt. Dann sind folgende Zustandsbeschreibungen formulierbar:

$\alpha_1 \varepsilon \Pi \wedge \alpha_2 \varepsilon \Pi \wedge \alpha_3 \varepsilon \Pi$	[Abkürzung: $_3\Sigma_1$]
$\alpha_1 \varepsilon \Pi \wedge \alpha_2 \varepsilon \Pi \wedge \neg (\alpha_3 \varepsilon \Pi)$	[Abkürzung: $_3\Sigma_2$]
$\alpha_1 \varepsilon \Pi \wedge \neg (\alpha_2 \varepsilon \Pi) \wedge \alpha_3 \varepsilon \Pi$	[Abkürzung: $_3\Sigma_3$]
$\neg (\alpha_1 \varepsilon \Pi) \wedge \alpha_2 \varepsilon \Pi \wedge \alpha_3 \varepsilon \Pi$	[Abkürzung: $_3\Sigma_4$]
$\alpha_1 \varepsilon \Pi \wedge \neg (\alpha_2 \varepsilon \Pi) \wedge \neg (\alpha_3 \varepsilon \Pi)$	[Abkürzung: $_3\Sigma_5$]
$\neg (\alpha_1 \varepsilon \Pi) \wedge \alpha_2 \varepsilon \Pi \wedge \neg (\alpha_3 \varepsilon \Pi)$	[Abkürzung: $_3\Sigma_6$]
$\neg (\alpha_1 \varepsilon \Pi) \wedge \neg (\alpha_2 \varepsilon \Pi) \wedge \alpha_3 \varepsilon \Pi$	[Abkürzung: $_3\Sigma_7$]
$\neg (\alpha_1 \varepsilon \Pi) \wedge \neg (\alpha_2 \varepsilon \Pi) \wedge \neg (\alpha_3 \varepsilon \Pi)$	[Abkürzung: $_3\Sigma_8$]

Unter der Voraussetzung, daß Π nicht zerrüttet und $_3c^1$ fair ist, ergeben sich folgende Bewertungen für diese Sätze:

$_3c^1 (_3\Sigma_1, T) = {}_3c^1 (_3\Sigma_8, T) = {}^5/_{16}$.

$_3c^1 (_3\Sigma_2, T) = {}_3c^1 (_3\Sigma_3, T) = \ldots = {}_3c^1 (_3\Sigma_7, T) = {}^1/_{16}$.

Π bezeichnet unter der vorgelegten Interpretation J die Gesamtheit der roten Dinge dieses Bereichs und unter J' die Gesamtheit jener Dinge, die rot und von a_2 verschieden oder aber, die nicht rot und mit a_2 identisch sind (da die Intension von Π nicht durch Axiome festgelegt ist, sind beide Interpretationen zulässig). Wenn die Interpretation J' intendiert ist und Π somit eine zerrüttete Eigenschaft ist, wird man eine Funktion $_3c$ als optimal ansehen, die durch folgende Bewertungen eindeutig festgelegt ist:

$_3c (_3\Sigma_3, T) = {}_3c (_3\Sigma_6, T) = {}^5/_{16}$.

$_3c (_3\Sigma_1, T) = {}_3c (_3\Sigma_2, T) = {}_3c (_3\Sigma_4, T) =$
$= {}_3c (_3\Sigma_5, T) = {}_3c (_3\Sigma_7, T) = {}_3c (_3\Sigma_8, T) = {}^1/_{16}$.

Diese Bewertung erhält man, wenn man zunächst annimmt, daß Π nicht zerrüttet ist, sodann mittels Π und α_2 ein neues Prädikat Π^+ definiert und schließlich überall simultan Π durch Π^+ und Π^+ durch Π ersetzt. Es wird dann also angenommen, daß der Gegenstands-

bereich bezüglich Π^+ einen gewissen Grad an Uniformität besitzt, während er bezüglich Π in dem gleichen Grad anti-uniform ist.
Die Bedingung (II) (a) ist durch die so charakterisierte Funktion $_3c$ nicht mehr erfüllt; es gilt ja z. B.:
$_3c\,(_3\Sigma_4, T) = {}^1/_{16} \neq {}^5/_{16} = {}_3c\,(_3\Sigma_3, T)$,
wobei $_3\Sigma_4$ aus $_3\Sigma_3$ durch simultane Vertauschung der Ausdrücke α_1 und α_2 entsteht und umgekehrt. Es gilt ganz allgemein, daß Funktionen, die man auf diese Weise gewinnt (die den Glauben zum Ausdruck bringen, daß einige der Eigenschaften anti-homogen sind), für B_n mit der Anzahl $n > 2$ die Bedingung der Symmetrie der Gegenstandsausdrücke und damit die unabhängige Wiederholbarkeit eines Experiments verletzen. Nun gehören $_2c$ und $_3c$ zu einer passenden Folge von Bestätigungsmethoden, denn es gilt für jede Zustandsbeschreibung $_2\Sigma_i$ aus der Sprache mit nur zwei Gegenstandsausdrücken: $_2c\,(_2\Sigma_i, T) = {}_3c\,(_2\Sigma_i, T)$. Man erhält damit das etwas merkwürdige Resultat, daß α_1 und α_2 allein durch $_3c$ als symmetrisch, zusammen mit α_3 jedoch als nicht symmetrisch behandelt werden, in der Sprache der mathematischen Wahrscheinlichkeitslehre: daß nach der vorgegebenen Bewertung die Experimente a_1 und a_2 unabhängig wiederholbar sind, solange es sich um die einzigen Experimente handelt, und daß sie es nicht mehr sind, wenn noch ein weiteres Experiment a_3 durchgeführt wird.
Der Beweisteil 5 (c) des Theorems 18 zeigt, daß mit wachsendem n die Bedingungen (I), (II) und (III) immer größere Teile von (IV) (a) implizieren; für $n = \infty$ ist (IV) (a) ganz in diesen Bedingungen (unter Hinzunahme der Limeskonvention für Gegenstandsausdrücke) enthalten. Wie Humburg gezeigt hat,[20] kann (IV) (a) bereits aus den Bedingungen (I) und (II) (a) sowie der Limeskonvention für Gegenstandsausdrücke hinsichtlich einer passenden Folge von Bestätigungsmethoden deduziert werden. Da man hinsichtlich zerrütteter Prädikate eine Funktion c als optimal ansehen wird, die (I) und die Limeskonvention [und die wohl auch (II) (b) und (III)] erfüllt, die jedoch (IV) (a) verletzt, wird diese auch (II) (a) verletzen. *Für den speziellen Fall, daß der*

[20] Vgl. Humburg [1], S. 9f. Seine Bedingung der Unabhängigkeit von c von der Anzahl der Elemente des Gegenstandsbereichs ist gleichwertig mit der Limeskonvention für Gegenstandsausdrücke hinsichtlich einer passenden Folge von Bestätigungsmethoden.

Gegenstandsbereich B abzählbar unendlich ist, ist demnach die Voraussetzung, daß die Prädikate zerrüttet sind, gleichwertig mit der Annahme, daß die Experimente nicht unabhängig wiederholbar sind (es sei denn, daß die Prädikate irreparabel zerrüttet sind oder daß der Gegenstandsbereich hinsichtlich der Prädikate durchgehend disuniform ist: dann ist die Methode c^{∞} optimal, die der Bedingung (II) (a) genügt und die somit unabhängige Wiederholbarkeit des betreffenden Experiments voraussetzt).

Für diesen Fall ist die Lösung des Problems der zerrütteten Prädikate einfacher zu formulieren; es braucht dann im Grunde nur noch wiederholt zu werden, was bereits im 4. Abschnitt des V. Kapitels ausgeführt worden ist: Wenn die Art und Weise der Abhängigkeit, die zwischen den Wiederholungen der Experimente besteht, in einer Objektsprache ausdrückbar ist, für die Bestätigungsmethoden definiert sind, und wenn Γ das Gesetz ist, das diese Abhängigkeit formuliert, dann kann eine für diese Situation optimale reguläre Bestätigungsmethode c folgendermaßen definiert werden: $c(\Psi, \Phi) = c^{L}(\Psi, \Phi \wedge \Gamma)$; so wie man in der deduktiven Logik *empirische Schlüsse* zu *logischen Argumenten* macht, indem man zu den Prämissen noch das Wissen hinzunimmt, auf Grund dessen man empirisch geschlossen hat, wird hier ein *empirischer Wahrscheinlichkeitsschluß* auf ein *logisches Wahrscheinlichkeitsargument* reduziert, indem c in ein empirisches Wissen und in eine reine induktive Methode c^{L} aufgespalten wird. Wenn hingegen die Art und Weise der Abhängigkeit, die zwischen den Wiederholungen der Experimente besteht, nicht in einer Objektsprache formuliert werden kann, wie etwa beim Relais-Experiment im Fall 3, wenn also total zerrüttete Prädikate vorliegen und der Gegenstandsbereich damit ein Minimum an Uniformität besitzt, ist c^{∞} optimal.

3. Das Problem der Dispositionsprädikate

Gelegentlich ist ein Zusammenhang zwischen den Problemen der zerrütteten Prädikate und der Dispositionsprädikate gesehen worden von der Art, daß die Lösung des einen die des anderen nach sich zieht.

Solange ein derartiger Zusammenhang nicht streng bewiesen ist, werde ich nicht versuchen, die Lösung des einen Problems mit Gewalt auch auf einen ähnlich gelagerten Komplex zu übertragen und sie so unzulässig zu verallgemeinern und zu überdehnen. Während ich der Ansicht bin, daß das Problem der zerrütteten Prädikate zwar die induktive Logik betrifft, aber nicht auf sie beschränkt ist und auch nicht durch sie gelöst werden kann, scheint mir umgekehrt die Lösung des Problems der Dispositionsprädikate nur durch eine adäquate induktive Logik möglich zu sein.

Dispositionsprädikate werden üblicherweise durch sogenannte *operationale Definitionen* eingeführt: Es werden gewisse Testbedingungen formuliert und es wird gesagt, daß ein Gegenstand genau dann die betreffende Disposition besitzt, wenn er beim Test in einer bestimmten Weise reagiert. Daß ein Gegenstand die Disposition der (permanenten) Wasserlöslichkeit besitzt, kann dann so entschieden werden, daß man ihn dem Test unterwirft und in Wasser gibt und sodann beobachtet, ob er sich darin auflöst oder nicht. Schon die einfachsten Begriffe des Alltags können in eine phänomenalistische Sprache als Dispositionsprädikate eingeführt werden; für den Begriff „rot" lautet die operationale Definition dann, daß ein Gegenstand genau dann rot ist, wenn ich unter gewissen Bedingungen (den sogenannten *Normalbedingungen*: daß ich nicht rot-grün-blind bin, daß normale Sichtverhältnisse vorliegen, usw.) Wahrnehmungen einer bestimmten Art habe. Aber auch in physikalisch-realistischen Sprachen sind viele der grundlegenden Begriffe Dispositionsprädikate, wie etwa „magnetisch"; man kann ja festsetzen, daß ein Gegenstand genau dann magnetisch ist, wenn er unter gewissen Testbedingungen (z. B. Heranbringen von Eisenstücken) in einer bestimmten Weise reagiert (Anziehen dieser Eisenstücke).

Für den Logiker und Wissenschaftstheoretiker ergibt sich hierbei die Frage, wie das allgemeine Schema zur Einführung dieser Begriffe in die Wissenschaftssprachen genau lautet. Zunächst hat Carnap folgende Definition für „wasserlöslich" im Auge gehabt: „Ein Gegenstand ist wasserlöslich genau dann, wenn er sich zu allen Zeiten, zu denen er in Wasser gegeben wird, auflöst", symbolisch:
„$\Lambda\, x\, [x \in Wl \leftrightarrow \Lambda\, t\, (\langle x, t\rangle \in Wg \rightarrow \langle x, t\rangle \in Al)]$".
Die Definition für „magnetisch" lautet dann: „Ein Gegenstand ist

magnetisch genau dann, wenn er jeden Eisenkörper, der in seine Nähe gebracht wird, anzieht", symbolisch:
„$\Lambda\, x\, [x \in Mg \leftrightarrow \Lambda\, t\, \Lambda\, y\, (\langle x, y, t\rangle \in Nb \rightarrow \langle x, y, t\rangle \in Az)]$".
Allgemein lautet das Definitionsschema für Dispositionsprädikate:
„$\Lambda\, x\, [x \in F \leftrightarrow \Lambda\, y_1 \ldots \Lambda\, y_n\, (\langle x, y_1, \ldots, y_n\rangle \in U_1 \rightarrow \langle x, y_1, \ldots, y_n\rangle \in U_2)]$".
Später hat Carnap dann entdeckt, daß Dispositionsprädikate durch dieses Schema nicht adäquat in eine Wissenschaftssprache eingeführt werden. Wenn ein Gegenstand a zu keinem Zeitpunkt seiner Existenz in Wasser gegeben worden ist, so daß (unter der üblichen Interpretation der Sprache) der Satz „$\Lambda\, t \neg (\langle a, t\rangle \in Wg)$" wahr ist, dann ergibt sich daraus logisch die Wahrheit von „$\Lambda\, t\, (\langle a, t\rangle \in Wg \rightarrow \langle a, t\rangle \in Al)$" und somit nach der obigen Definition auch die von „$a \in Wl$"; diese Definition führt also dazu, daß *alle* Gegenstände, die während ihrer Existenz nie in Wasser gegeben worden sind, als wasserlöslich angesehen werden müssen, was absurd ist.[21] Ersetzt man im Definiens den Allquantor durch den Existenzquantor und das Implikationszeichen durch das Symbol für die Konjunktion, so erhält man umgekehrt, daß kein Ding, das zeit seiner Existenz nicht in Wasser gegeben worden ist, wasserlöslich ist, was ebenfalls absurd ist.[22]
Carnap hat daher[23] das intuitive Verfahren der sogenannten operationalen Einführung von Begriffen genauer analysiert und dabei festgestellt, daß es aus zwei Teilverfahren besteht, die folgendermaßen beschrieben werden können:[24] „Wenn man einen Gegenstand x, der

---

[21] Wenn man auf diese Weise auch „wasserunlöslich" definiert, so daß ein Gegenstand wasserunlöslich ist genau dann, wenn er sich zu keiner Zeit, in der er in Wasser gegeben wird, auflöst, so ergibt sich für einen solchen Gegenstand a der zwar nicht formale, aber intuitive Widerspruch, daß er sowohl wasserlöslich als auch wasserunlöslich ist (nach Hinzufügung der Prämisse, daß a genau dann wasserlöslich ist, wenn er nicht wasserunlöslich ist, erhält man daraus auch einen formalen Widerspruch).

[22] Wenn man auf diese Weise auch „wasserunlöslich" definiert, so daß ein Gegenstand wasserunlöslich ist genau dann, wenn es einen Zeitpunkt gibt, zu dem er in Wasser gegeben worden ist und sich aufgelöst hat, so erhält man für a den zwar wiederum nicht formalen, aber intuitiven Widerspruch, daß er weder wasserlöslich noch wasserunlöslich ist.

[23] In Carnap [1], S. 439—453.

[24] Im folgenden werde ich das Problem der Dispositionsprädikate stets an einem konkreten Beispiel, und zwar, um Zusatzkomplikationen zu vermeiden, an einem einfachen Beispiel, nämlich dem der Wasserlöslichkeit, behandeln. Dabei

wasserlöslich ist, zu irgendeiner Zeit in Wasser gibt, dann löst er sich darin auf; und wenn man einen Gegenstand x, der nicht wasserlöslich ist, zu irgendeiner Zeit in Wasser gibt, dann löst er sich darin nicht auf", symbolisch:

„Λ x Λ t [x ε Wl ∧ ⟨x, t⟩ ε Wg → ⟨x, t⟩ ε Al] ∧
Λ x Λ t [¬ (x ε Wl) ∧ ⟨x, t⟩ ε Wg → ¬ (⟨x, t⟩ ε Al)]".

Dieser Satz ist nun logisch äquivalent mit: „Wenn ein Gegenstand x zu irgendeinem Zeitpunkt t in Wasser gegeben wird, dann ist x wasserlöslich genau dann, wenn er sich zur Zeit t darin auflöst", symbolisch:

„Λ x Λ t [⟨x, t⟩ ε Wg → (x ε Wl ↔ ⟨x, t⟩ ε Al)]".

Dieser Satz (nach Carnap ein *bilateraler Reduktionssatz*) hat die Form einer bedingten (oder partiellen) Definition.[25] Er legt die Intension von „wasserlöslich" also nur insoweit fest, als die Bedingung erfüllt ist, d. h. als der betreffende Gegenstand zu einem Zeitpunkt in Wasser gegeben worden ist oder noch einmal während seiner Existenz in Wasser gegeben wird. Solche bedingten Definitionen sind auch in

setze ich voraus, daß aus dem Satz „⟨x, t⟩ ε Wg" folgt, daß x in einer vernünftigen Relation zur Wassermenge steht, daß also nicht etwa ein 20 kg schweres Zuckerstück in eine Menge von 1 g Wasser gegeben wird. Da sich x nicht *momentan* im Wasser auflösen kann, wird angenommen, daß das Zeitkontinuum in vernünftige Zeitintervalle aufgeteilt wird, so daß man, wenn x sich nicht innerhalb eines solchen Intervalls auflöst, mit Recht von der Falschheit des Satzes „⟨x, t⟩ ε Al" sprechen kann; daraus ergibt sich als weiterer Vorteil, daß man es mit höchstens abzählbar unendlich vielen Zeitintervallen zu tun hat. Die Relationen Wg und At seien so bestimmt, daß ihre Nachbereiche mit der Gesamtheit der Zeitintervalle identisch sind.

25 Bedingte Definitionen erfüllen nicht Kriterien der vollständigen Eliminierbarkeit und der durchgehenden Nichtkreativität, sondern nur Abschwächungen derselben, nämlich, daß der so eingeführte Ausdruck eliminierbar und in keinem Kontext kreativ ist, *in dem die Bedingung der partiellen Definition erfüllt ist.* Verschiedene Autoren haben erklärt, daß bilaterale Reduktionssätze nicht als Definitionen angesehen werden können, da man z. B. „magnetisch" nicht nur durch das Experiment der Anziehung von Eisenkörpern, sondern auch durch den Test mit Eisenfeilspänen operational einführen kann (vgl. z. B. Carl G. Hempel, A Theoreticians Dilemma, in: H. Feigl, M. Scriven und G. Maxwell, Minnesota Studies in the Philosophy of Science, Bd. II, Minneapolis 1958, S. 71). Ich halte dieses Argument nicht für überzeugend, da auch bei Explizitdefinitionen der Fall eintreten kann, daß man sich durch einen Willkürakt (durch eine Zweckmäßigkeitsbetrachtung) dafür entscheidet, eine Gleichung als Definition und eine andere als Axiom zu nehmen, so daß man eine dritte als Theorem erhält (man denke nur an die verschiedenen Gleichungen für den physikalischen Kraftbegriff, von denen im Prinzip jede als Definition in Frage kommt).

den Aprioriwissenschaften wie der Logik und der Mathematik gang und gäbe. Die Operation der Division etwa wird auf diese Weise eingeführt, d. h. es wird definiert:
„$\Lambda\, r\, \Lambda\, s\, \Lambda\, t\, [s \neq 0 \rightarrow (r/s = t \leftrightarrow r = s \cdot t)]$".
Zur Definition des Begriffs der logischen Wahrheit benötigt der Logiker den der Wahrheit. Diesen führt er auf folgendem Weg ein: Es wird zunächst (mit den Begriffen der Metasprache) aus dem Gegenstandsbereich eine Teilklasse ausgesondert, die als sprachliche Zeichen dienen soll. Der Begriff des Satzes wird dann nicht mehr für *alle* Gegenstände, sondern nur noch für solche, die sprachliche Zeichen sind, eingeführt, und der Wahrheitsbegriff wird nicht mehr für alle sprachlichen Zeichen (geschweige denn für alle Gegenstände), sondern nur noch für solche, die Sätze sind, definiert, d. h. es wird gesagt: „Wenn Φ ein Satz ist, dann ist Φ wahr genau dann, wenn ...". Durch eine bedingte Definition ist schließlich in D 9 auch der Begriff „gn" eingeführt worden.
Für Fälle, in denen die Bedingung nicht erfüllt ist, ist durch solche partielle Definitionen auch der betreffende Begriff nicht definiert; so ist das Divisionszeichen in dem Satz „$r/0 = t$" nicht definiert, und gleiches gilt für „Der Baum vor meinem Fenster ist ein Satz" und „Dieses Parkett ist wahr". Wenn jemand solche Wortfolgen formuliert, so wird er vom Mathematiker oder vom Logiker hören, daß sie keine sinnvollen Sätze ergeben, da einige der darin vorkommenden Begriffe nicht für die entsprechenden Gegenstände definiert sind. Ähnlich kann man im Fall der bedingten Definitionen in den Erfahrungswissenschaften, durch die Dispositionsprädikate eingeführt werden, nicht argumentieren; wenn ein Zuckerstück a oder ein Streichholz b, bevor sie verbrannt worden sind, nie in Wasser gegeben worden sind, so hat es für uns durchaus einen Sinn, zu sagen, a sei löslich und b sei nicht löslich, obwohl der Test nie durchgeführt werden konnte. Hierin sehe ich einen entscheidenden Unterschied zwischen den Begriffen, die in Aprioriwissenschaften, und solchen, die in Erfahrungswissenschaften durch bedingte Definitionen eingeführt werden.
Wenn die Dispositionsprädikate auch dann sinnvoll angewendet werden können, wenn die betreffenden Testbedingungen nicht erfüllt sind, so muß man versuchen, das Definitionsschema so abzuändern

oder zu ergänzen, daß es unsere intuitiven Vorstellungen vollständig und adäquat wiedergibt. Dazu sehe ich im Prinzip folgende drei Möglichkeiten: 1) Man kann versuchen, das Definitionsschema durch ein anderes zu ersetzen. 2) Man kann versuchen, die (extensionale, auf dem zweiwertigen Wahrheitsbegriff beruhende) deduktive Logik durch eine andere zu ersetzen oder durch eine andere zu ergänzen. 3) Man kann versuchen, den Sinn der Dispositionsprädikate von den Fällen, in denen die Testbedingungen erfüllt sind, auf die Fälle, in denen sie es nicht sind, *induktiv* zu erweitern.

Der erste Ausweg ist der naheliegendste; da er jedoch trotz eifrigstem Bemühen nicht zum Erfolg geführt hat, wird er heute nicht mehr begangen. Daraus folgt selbstverständlich nicht, daß nicht eines Tages doch noch ein adäquates Schema zur Einführung von Dispositionsprädikaten in der Form von Explizitdefinitionen gefunden wird, da bisher kein Beweis für die Unmöglichkeit einer derartigen Definierbarkeit vorliegt. Da jedoch scheinbar alle halbwegs vernünftigen Möglichkeiten der Definition schon ausprobiert sind, beurteile ich weitere Definitionsversuche skeptisch.

Die zweite Möglichkeit ist bisher noch nirgendwo im Detail durchgeführt worden, so daß es schwerfällt, darüber etwas Genaues zu sagen. Eine Möglichkeit wäre, der üblichen Logik, die auf Indikativsätzen beruht, eine Logik der Konjunktivsätze hinzuzufügen und den Begriff der Löslichkeit (und analog die anderen Dispositionsprädikate) folgendermaßen zu definieren: „Wenn ein Gegenstand x zu einer Zeit t in Wasser gegeben wird, dann ist x wasserlöslich genau dann, wenn x sich zur Zeit t auflöst, und wenn x zu keiner Zeit in Wasser gegeben wird, dann ist x wasserlöslich genau dann, wenn x sich zu einer Zeit t in Wasser auflösen *würde*, wenn man ihn da in Wasser *gäbe*". Zu einer solchen Logik der Konjunktivsätze existieren jedoch keine zum Ziel führenden Ansätze.

Eine andere Möglichkeit wäre die Einführung von Modaloperatoren (von Junktoren „notwendig" und „möglich", die wie das Negationssymbol als objektsprachliche Zeichen aufgefaßt werden). Es könnte dann der Begriff der Wasserlöslichkeit folgendermaßen eingeführt werden: „Ein Gegenstand x ist wasserlöslich genau dann, wenn es notwendig (vielleicht: wenn es ein Naturgesetz) ist, daß sich x immer dann auflöst, wenn x in Wasser gegeben wird" (vermutlich ist der

irreale Konditionalsatz „Wenn man x zu irgendeiner Zeit in Wasser geben *würde*, so *würde* es sich darin auflösen" sinngleich mit dem Modalurteil „Es ist notwendig, daß sich x immer dann auflöst, wenn man x in Wasser gibt"). Ich möchte nur eine der mit diesem Vorschlag verbundenen Schwierigkeiten diskutieren, nämlich das Problem, wie der Begriff „notwendig" zu charakterisieren ist; denn ich glaube, damit hinreichend klarmachen zu können, daß auch dieser Ausweg entweder zu keinem Ziel führt oder aber zu einem Ziel, das sich auf andere Weise leichter erreichen läßt.

Im folgenden wird der Modalkalkül S 5 entwickelt.[26] Neben dem Zeichen „$\rightarrow$" für „wenn — dann" wird noch „$\Box$" für „notwendig" und „$\Diamond$" für „möglich" verwendet; „$\Diamond$" kann dabei als Abkürzung für „$\neg \Box \neg$" aufgefaßt werden, d. h. „$\Diamond$" ist auf folgende Weise definierbar: $\Diamond \Phi \leftrightarrow \neg \Box \neg \Phi$. „$\vdash$" sei eine metasprachliche Abkürzung für „herleitbar". Die Axiome von S 5 lauten:

A 1 $\vdash \Box \Phi \rightarrow \Phi$

A 2 $\vdash \Box (\Phi \rightarrow \Psi) \rightarrow (\Box \Phi \rightarrow \Box \Psi)$

A 3 $\vdash \Box \Phi \rightarrow \Box \Box \Phi$

A 4 $\vdash \Phi \rightarrow \Box \Diamond \Phi$

A 5 $\vdash \Diamond \Phi \rightarrow \Box \Diamond \Phi$

S 5 hat die beiden Regeln:

R 1 Wenn $\vdash \Phi$ und $\vdash \Phi \rightarrow \Psi$, dann $\vdash \Psi$.

R 2 Wenn $\vdash \Phi$, dann $\vdash \Box \Phi$.

Wer nicht weiß, was „$\Box$" und damit „$\Diamond$" bedeutet, wird es auch durch diese Axiome und Ableitungsregeln nicht lernen, d. h. diese legen den korrekten Gebrauch der beiden Begriffe nicht eindeutig fest. Dennoch sind zwei Deutungen dieser beiden Operatoren naheliegend: (a) „$\Box$" gibt objektsprachlich in etwa das wieder, was der metasprachliche Begriff „junktoren- und quantorenlogisch wahr" ausdrückt, und entsprechend charakterisiert „$\Diamond$" etwa das gleiche wie „erfüllbar". (b) „$\Box$" und „$\Diamond$" werden im gleichen Sinn wie „faktisch bestehend" gedeutet, d. h. es wird „notwendig" mit „möglich" und „faktisch bestehend" gleichgesetzt und damit ein Zufall ausgeschlossen; diese Deutung kommt dadurch zum Ausdruck, daß man

---

26 Vgl. hierzu auch Essler [2], wo ich ausführlicher auf diese Fragen eingehe und weitere Literaturhinweise gebe.

„□" als Abkürzung für „¬ ¬" und entsprechend „◇" als Abkürzung für „¬ ¬ ¬ ¬" ansieht, daß man also definiert: $\Box \Phi \leftrightarrow \neg \neg \Phi$.[27] Offensichtlich wird der Modalkalkül S 5 bei dieser Deutung zu einem unvollständigen Teilkalkül der Junktorenlogik.

Ich gebe dieser Interpretation (b) den Vorzug vor (a) und allen anderen Deutungen, weil sie fruchtbarer ist als die anderen, d. h. mehr Gesetze als diese zu formulieren gestattet, nämlich insbesondere alle Gesetze der Logik der Identität, die in (a) und verwandten Interpretationen nicht gelten. Trotzdem werden überzeugte Modallogiker wohl gerade auf diese Deutung verzichten wollen; sie werden versuchen, S 5 durch zusätzliche Axiome zu ergänzen, die diese Deutung ausschließen. Gegen die Adäquatheit einer derartigen Ergänzung kann folgendes eingewendet werden:

Saul Kripke hat für S 5 eine Semantik entworfen, die man in groben Umrissen folgendermaßen skizzieren kann: Es wird eine Relation R in der Metasprache eingeführt, die über Zustandsbeschreibungen (über die Klasse Zb) definiert ist und die intuitiv folgendes zum Ausdruck bringen soll: $\langle \Sigma_1, \Sigma_2 \rangle$ e R genau dann, wenn die Welt H_2, die durch die Zustandsbeschreibung Σ_2 beschrieben wird, *möglich* ist relativ zur Welt H_1, die durch die Zustandsbeschreibung Σ_1 beschrieben wird. Es wird dann axiomatisch festgelegt, daß R eine Äquivalenzrelation über Zb ist, d. h. R für Zustandsbeschreibungen reflexiv, transitiv und symmetrisch ist. Der Begriff der Wahrheit eines Satzes bei einer Interpretation oder, was in diesem Zusammenhang damit gleichwertig ist,[28] der Begriff des Geltens eines Satzes bei einer Zustandsbeschreibung, wird dann folgendermaßen eingeführt: Wenn Φ keinen Modaloperator (keines der Zeichen „□" und „◇") enthält,

---

[27] Es wird also die doppelte Negation durch ein einziges Zeichen ausgedrückt, wie ja in der Aussage „weder A noch B", die gleichwertig mit „nicht A und nicht B" ist, das „noch" für „und nicht" steht.

[28] Vgl. das auf S. 95 f. dieses Buches Gesagte. — Wenn nicht, wie hier, davon ausgegangen wird, daß der Gegenstandsbereich von einer fest vorgegebenen Anzahl ist, dann sind beliebige Interpretationen der Sprache über beliebig umfangreichen Gegenstandsbereichen möglich (nach einem Theorem von Skolem braucht man allerdings keine überabzählbar unendlichen Bereiche in Erwägung zu ziehen). In diesem Fall benötigt man eine umfangreichere Klasse von Zustandsbeschreibungen, nämlich die Vereinigung der Klassen der Zustandsbeschreibungen für alle endlichen und abzählbar unendlichen Bereiche. Die Relation R wird hierbei ganz analog eingeführt. Vgl. auch Essler [2].

dann gilt Φ bei Σ genau dann, wenn Φ aus Σ logisch folgt; $\Box\,\Phi$ gilt bei Σ genau dann, wenn Φ bei jedem Σ' aus Zb gilt, so daß $\langle \Sigma, \Sigma' \rangle \in R$, und $\Diamond\,\Phi$ gilt bei Σ genau dann, wenn $\neg\,\Box\,\neg\,\Phi$ bei Σ gilt. Wie Kripke nun zeigen konnte, ist ein Satz Φ in S 5 genau dann herleitbar, wenn er bei allen Zustandsbeschreibungen gilt.

R teilt die Gesamtheit der Zustandsbeschreibungen in Äquivalenzklassen ein. Dabei können zwei Extremfälle eintreten: (α) R ist über Zb die universelle Relation, gilt also zwischen beliebigen Zustandsbeschreibungen, so daß überhaupt nur eine Äquivalenzklasse existiert. Dann gilt $\Box\,\Phi$ (bei einer Zustandsbeschreibung Σ) genau dann, wenn Φ bei allen Zustandsbeschreibungen gilt, was genau dann der Fall ist, wenn Φ logisch wahr ist; dann liegt also die Deutung (a) des Operators „$\Box$" vor. (β) R ist über Zb die Identitätsrelation, teilt also Zb in Äquivalenzklassen auf, die durchgehend Einerklassen sind; dann gilt $\Box\,\Phi$ bei einer Zustandsbeschreibung Σ genau dann, wenn Φ bei Σ gilt; es liegt hier also die Deutung (b) des Operators „$\Box$" vor.

Eine echte Verstärkung von S 5, d. h. eine Hinzufügung von Axiomen, die nicht bereits aus S 5 herleitbar sind [zum Zwecke der Ausschaltung der Deutung (b) und vielleicht auch der Deutung (a)], hätte nun zur Folge, daß R zusätzlich bestimmt werden müßte, und zwar so, daß für wenigstens zwei Zustandsbeschreibungen Σ_1 und Σ_2 gilt, daß $\langle \Sigma_1, \Sigma_2 \rangle \in R$, und eventuell auch, daß für wenigstens zwei derartige Sätze Σ_1 und Σ_2 gilt, daß nicht $\langle \Sigma_1, \Sigma_2 \rangle \in R$. Ich sehe jedoch nicht, mit welchem Recht man solche Behauptungen noch als logische Aussagen ansehen kann; ich halte es durchaus für *möglich*, daß jede Welt hinsichtlich einer beliebigen anderen möglich ist, und mir erscheint es andererseits auch als *möglich*, daß keine Welt hinsichtlich einer anderen möglich ist.

Der Modaloperator „$\Box$" ist eindeutig definiert, wenn für zwei beliebige Elemente Σ_1 und Σ_2 aus Zb feststeht, ob $\langle \Sigma_1, \Sigma_2 \rangle \in R$ oder nicht, und damit, wie Zb durch R in Äquivalenzklassen aufgeteilt wird. Eine derartige Aufteilung kann nun willkürlich oder nach einem allgemeinen Verfahren geschehen; im ersten Fall liegt eine Welt vor, die ganz oder teilweise indeterminiert ist, während im zweiten Fall alle Zusammenhänge durch Gesetze von deterministischer Form beschrieben werden können.[29] Was immer vorliegen mag: Ich halte es

[29] Man hat hier also ein Quasi-Kontinuum von möglichen Aufteilungen von

für weniger umständlich, diese Zusammenhänge in der Objektsprache zu beschreiben und den Umweg über diese Semantik und S 5 zu vermeiden; wenn diese objektsprachlichen Sätze keine Lösung des Problems der Dispositionsprädikate bringen, wird eine solche durch S 5 ebenfalls nicht gelingen.

Man wird vielleicht geneigt sein, die Verwendung einer Modallogik auf folgende Weise zu rechtfertigen: Notwendig sind offenbar die Naturgesetze und nur sie; daß ein Sachverhalt notwendig ist, besagt demnach, daß der Satz, der ihn beschreibt, aus den Naturgesetzen logisch folgt. Dem kann folgendes entgegengehalten werden:

(1) Bei dieser Interpretation des Notwendigkeitsbegriffs wissen wir in keinem einzigen Fall, ob ein Sachverhalt notwendig ist oder nicht, da wir von keinem einzigen Naturgesetz (im strengen Sinn) *wissen*, ob es gilt oder nicht. Für den Fall, daß jene Aussagen, die wir heute als Naturgesetze annehmen, sich einmal als falsch erweisen würden, hätten wir dann Sachverhalte als notwendig angesehen, die gar nicht bestehen.

(2) So wie die Quantoren*logik* uns nicht sagt, welche Allsätze wahr und welche falsch sind, sondern uns nur logische Folgebeziehungen liefert, die zwischen einigen dieser Behauptungen bestehen, so soll uns auch die Modal*logik* nicht mitteilen, welche Notwendigkeitsbehauptungen wahr und welche falsch sind, sondern uns lediglich logische Folgebeziehungen liefern, die zwischen einigen von ihnen bestehen; sie soll uns also nicht sagen, was *tatsächlich* notwendig ist, sondern sie soll eine *Logik* des Notwendigkeitsbegriffs sein. Jene Interpretation liefert jedoch viel mehr und ist also inadäquat.

(3) Hinsichtlich des Problems der Dispositionsprädikate ist diese Interpretation hingegen wiederum zu schwach; mit einer so gedeuteten Modallogik sind nur solche Probleme lösbar, die bereits durch die Logik jener modalfreien Sprache, in der die Naturgesetze formuliert sind, gelöst werden können, entgegen den Intentionen der Modallogiker. Da für ein System der Modallogik die Deutung (b) zulässig sein muß, die besagt, daß genau das faktisch Bestehende notwendig ist, kann es nur für solche Probleme Lösungen liefern, die auch schon

---

Zb in Äquivalenzklassen mit den Extremfällen (α) und (β) und dementsprechend ein Quasi-Kontinuum von Modallogiken (von modalen Methoden) mit den Extremfällen (a) und (b).

mit den Methoden der Junktoren- und Quantorenlogik angegangen werden können.

(4) Hinsichtlich des vorliegenden Problems ist die Anwendung einer so gedeuteten Modallogik schließlich zirkulär. Denn in den Naturgesetzen werden auch den unbeobachteten Dingen Dispositionen zugeschrieben, da sie nicht nur für die schon *beobachteten* Dinge, sondern generell gelten. Das Problem der Dispositionsprädikate soll also mittels einer Theorie gelöst werden, die das Problem, wie Dispositionsprädikate für nichtgetestete Dinge zu definieren sind, schon als gelöst voraussetzt, was absurd ist.

Ich bin dafür, nicht immer dann, wenn man längere Zeit vergeblich nach der Lösung eines Problems sucht, gleich an der Logik herumzubasteln und sie so zu verbiegen, daß das Problem gegenstandslos wird. Oft wird damit mehr verschleiert als geklärt. Als Beispiel mögen die Versuche einiger Philosophen und Physiker[30] betrachtet werden, gewisse Schwierigkeiten bei der Deutung quantenmechanischer Phänomene durch Schaffung einer sogenannten *Quantenlogik* zu überbrücken. Ausgehend von einer zweifelhaften Interpretation der Heisenbergschen Unschärferelation[31] gelangen sie zu dem Schluß, daß in der Quantenmechanik das junktorenlogische Gesetz $\Phi \rightarrow (\Psi \rightarrow \Phi)$ nicht gelten könne; denn wenn Φ etwa besage, daß ein Teilchen a einen bestimmten Ort habe, und Ψ, daß sein daraufhin gemessener Impuls von einer bestimmten Größe sei, so dürfe dann nicht mehr Φ behauptet werden, da durch die Messung des Impulses ja der Ort des Teilchens in einer nicht zu berechnenden Weise verändert worden ist. Hiergegen ist (wenigstens) zweierlei zu sagen: Einmal wird in der Logik der Ausdruck „wenn — dann" *nicht* im Sinne einer zeitlichen Aufeinanderfolge verstanden, wie es bei dieser Argumentation offensichtlich geschieht, und zum anderen wird bei dieser Argumentation, wenn man sie genau analysiert, gesagt, daß

---

30 Zu diesen zählt insbesondere Peter Mittelstaedt mit seinen Schriften „Logik und Quantenlogik", in: Physikalische Blätter 18 (1962) S. 23—29, und „Philosophische Probleme der modernen Physik", Mannheim ³1968, Kap. VI.

31 Vgl. hierzu Patrick Suppes, The Role of Probability in Quantum Mechanics, in: B. Baumrin (Hrsg.), Philosophy of Science. The Delaware Seminar, Vol. I (1961/62), New York 1963, sowie auch Stegmüller [2], S. 486—517. Vgl. auch Popper [18], S. 18—28.

folgendes *nicht allgemein* gilt: Wenn ein Teilchen a zu einer Zeit t_1 einen gewissen Ort hat und wenn zu einer Zeit t_2 sein Impuls von einer bestimmten Größe ist, dann hat es zu t_2 noch den gleichen Ort; es wird also bestritten, daß der Satz $\Phi(\alpha, \tau_1) \rightarrow [\Psi(\alpha, \tau_2) \rightarrow \Phi(\alpha, \tau_2)]$ allgemein gilt. Dieser Satz ist jedoch keine logische Wahrheit, da die Aussage $\Phi(\alpha, \tau_1)$ nicht mit $\Phi(\alpha, \tau_2)$ identisch ist, und somit ist dieser Grund zur Schaffung einer Quantenlogik hinfällig. Im übrigen fällt auf, daß keiner dieser Physiker bemerkt zu haben scheint, daß sie die Heisenbergsche Unschärferelation *unter Verwendung der klassischen Mathematik* gewonnen haben, die stets mit den Mitteln der *gesamten klassischen (zweiwertigen) Logik* operiert. Konsequenterweise müßten sie eigentlich eine *Quantenmathematik* entwickeln, doch ist eine solche strenge Konsequenz bei ihnen nicht zu entdecken.

Da ich weder dem ersten noch dem zweiten Ausweg großes Vertrauen entgegenbringe, bin ich somit gezwungen, die dritte Möglichkeit auszuprobieren, also die durch den bilateralen Reduktionssatz partiell festgelegte Intension von „wasserlöslich" induktiv auf die nicht beobachteten Fälle zu übertragen. Ich denke dabei an folgendes Verfahren: Zunächst muß festgelegt werden, für welche Stoffe der Löslichkeitsbegriff überhaupt festgelegt werden soll. Im Alltag wird er nur hinsichtlich der festen Stoffe charakterisiert; die Chemiker hingegen sprechen auch gelegentlich davon, daß sich Sauerstoff oder Erdöl in geringerem Umfang in Wasser auflöst, und haben demnach einen umfassenderen Begriff der Löslichkeit vor Augen. Sodann wird der Bereich der Dinge in Teilklassen aufgegliedert von der Art, daß man von jeder von ihnen hinsichtlich des Löslichkeitsbegriffs ein hohes Maß an Uniformität erwarten kann, also etwa in Zucker (d. h. in die Klasse der Zuckerstücke), Holz, Kochsalz, Eisen, Kunstleder, ... und eventuell auch in die einzelnen Gase und Flüssigkeiten.

In dieser Aufzählung können selbstverständlich nur solche Gegenstandsarten vorkommen, die bis jetzt bekannt sind, also nicht etwa ein bestimmter Kunststoff, der erst in 150 Jahren entdeckt wird. Folglich kann die Intension von „wasserlöslich" auch nicht induktiv *auf alle Dinge schlechthin*, sondern nur *auf alle Dinge der uns bis jetzt hinreichend bekannten Arten* erweitert werden. Das schließt selbstverständlich nicht aus, daß von einer Art (etwa von einer bestimmten Plutonium-Silizium-Verbindung) nur relativ wenige Gegenstände auf ihre

Wasserlöslichkeit hin untersucht worden sind; wesentlich ist nur, daß von ihr so viele Dinge auf diese Eigenschaft hin überprüft worden sind, daß sich mit einer geeigneten L-Methode bestimmte Werte für den singulären Voraussageschluß ergeben, nämlich solche, die entweder sehr nahe bei 1 oder sehr nahe bei 0 liegen. Bei Gegenstandsarten, von denen dies (noch) nicht gilt, ist eine solche induktive Erweiterung (noch) nicht möglich, so daß es dann keinen Sinn hat, von ihren unbeobachteten Elementen über Löslichkeit oder Unlöslichkeit zu reden; dies steht, wie mir scheint, durchaus im Einklang mit der Verwendung dieses Paares von Prädikaten im Alltag.

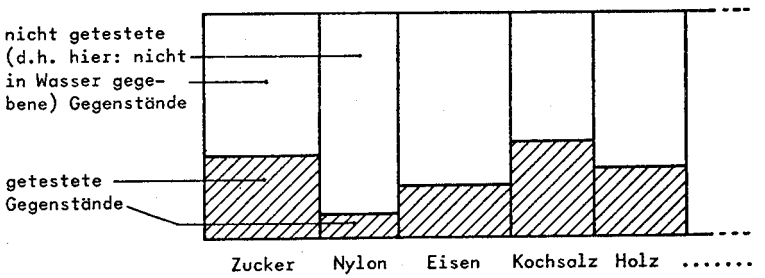


Fig. 12

Diese Aufteilung des Gegenstandsbereichs darf selbstverständlich nicht unter Zuhilfenahme des Dispositionsprädikates vorgenommen werden; bei einer Aufteilung der Dinge in wasserlösliche und wasserunlösliche entstehen zwei sich ausschließende und die Klasse der Dinge erschöpfende Teilbereiche, die selbstverständlich hinsichtlich der Eigenschaft *wasserlöslich* durchgehend uniform sind, da aus rein logischen Gründen dann jedes Ding der einen und keines des anderen Bereichs wasserlöslich ist. Ein solches zirkuläres Vorgehen führt jedoch zu keiner Lösung des Problems der Erweiterung der Intension des partiell definierten Prädikates auf den noch nicht getesteten Teilbereich der Dinge. Bei den folgenden Überlegungen wird eine solche Zirkularität dadurch vermieden, daß vorausgesetzt wird, daß jene Aufteilung ohne Bezugnahme auf das Dispositionsprädikat vorzunehmen ist. Am einfachsten sieht man dies, wenn man die folgende abstrakte Formulierung wählt: Man geht von einer Sprache aus, in der unter anderem die Eigenschaftsausdrücke „F", „G_1", „G_2",

„G_3", ... und die zweistelligen Relationsausdrücke „Q" und „R" vorkommen. An Axiomen und Definitionen enthält die Theorie, die mit den Mitteln dieser Sprache formuliert wird, eine partielle Definition von „F" in „Q" und „R" sowie Grundsätze, die besagen, daß die Klassen G_1, G_2, G_3, ... eine Division bilden, und eventuell noch andere Postulate. Keines der Axiome und Definitionen behauptet jedoch einen Zusammenhang zwischen F und G_1, G_2, G_3, ... Für diese Teilklassen wird nun (aus welchen Gründen auch immer) angenommen, daß ihre Uniformitätsgrade hinsichtlich der Eigenschaft F (für die das Dispositionsprädikat „F" steht) nahe genug bei 1 liegen (da nicht vorausgesetzt werden muß, daß diese Grade mit 1 identisch sind, kann man auch nicht so vorgehen, daß man definiert: „Ein x ist wasserlöslich genau dann, wenn x Zucker oder Kochsalz oder ... ist"). Im folgenden wird der Anschaulichkeit halber auf diese abstrakte Darstellung zugunsten des konkreten Beispiels verzichtet.

Es wird jetzt also versucht, dem Schlagwort „induktive Erweiterung der Intension eines Dispositionsprädikates auf die nichtgetesteten Fälle" einen klaren Sinn zu geben. Dies kann selbstverständlich nicht darin bestehen, daß man sich bemüht, so etwas wie eine *induktive* Definition solcher Begriffe anzugeben; Definitionen fallen selbstverständlich stets in den Bereich der deduktiven Logik. Versucht werden soll hingegen, 1. die Verwendung von Dispositionsprädikaten auch hinsichtlich der noch nicht beobachteten Dinge induktiv zu rechtfertigen und 2. dabei zu zeigen, daß bei dieser induktiven Erweiterung der durch (B) partiell festgelegten Intension das Schema (A) erreicht wird, daß also dann z. B. die Behauptung „$x \in Wl$" als induktiv gleichwertig mit „$\Lambda\, t\, (\langle x, t\rangle \in Wg \rightarrow \langle x, t\rangle \in Al)$" anzusehen ist, gleichgültig, wie man „x" interpretiert. Das Schema (A) ist ja nicht so schlecht, wie es zunächst erscheinen mag; es funktioniert für alle Gegenstände, die schon einmal der Testbedingung unterworfen worden sind oder ihr noch einmal unterworfen werden (die im Falle der Disposition *wasserlöslich* schon einmal in Wasser gegeben worden sind oder die noch einmal in Wasser gegeben werden); lediglich in den Fällen, in denen Gegenstände Zeit ihrer Existenz nie der Testbedingung unterworfen worden sind und ihr auch nie mehr unterworfen werden (und die man in diesem Sinne „pathologisch" nennen könnte), führen zu inadäquaten Resultaten. Eine induktive

Lösung des Problems soll also dieses Schema im wesentlichen rechtfertigen und gleichzeitig die inadäquaten Ergebnisse vermeiden. Wie dies möglich ist, wird im folgenden gezeigt.

Gegeben sei die Hypothese „Zucker ist wasserlöslich", die symbolisch durch den folgenden Satz wiedergegeben wird:

(a) „$\Lambda\, x\, [x \varepsilon Zr \rightarrow x \varepsilon Wl]$"

Der Satz „$X^a{}_k$" beschreibe eine Erfahrung über k Zuckerstücke hinsichtlich ihrer Wasserlöslichkeit (und enthalte eventuell noch ein Zusatzdatum, auf das später hingewiesen wird); auf Grund dieses Erfahrungsdatums sind wir bereit, (a) zu akzeptieren. Wenn wir nach dem Sinn von (a) gefragt werden, so werden wir entgegnen, daß wir diesen Satz als gleichwertig betrachten mit:

(a_1) „$\Lambda\, x\, [x \varepsilon Zr \rightarrow \Lambda\, t\, (\langle x, t\rangle \varepsilon Wg \rightarrow \langle x, t\rangle \varepsilon Al)]$"

Offensichtlich ist (a_1) mit (a) *deduktiv äquivalent*, wenn man die folgende (inzwischen bereits verworfene, aber *für den Normalfall plausible*) Definition benützt:

(A) „$\Lambda\, x\, [x \varepsilon Wl \leftrightarrow \Lambda\, t\, (\langle x, t\rangle \varepsilon Wg \rightarrow \langle x, t\rangle \varepsilon Al)]$"

Hingegen kann man genauso leicht sehen, daß (a_1) nicht mit (a) deduktiv äquivalent ist, wenn man die folgende bedingte Definition benützt:

(B) „$\Lambda\, x\, \Lambda\, t\, [\langle x, t\rangle \varepsilon Wg \rightarrow (x \varepsilon Wl \leftrightarrow \langle x, t\rangle \varepsilon Al)]$"

Es muß daher nach einem Weg gesucht werden, auf dem unter Verwendung der induktiven Logik gezeigt werden kann, daß das, was wir mit (a) meinen, und das durch (a_1) beschrieben wird, mit (a) gleichwertig ist, d. h. es muß gezeigt werden, daß unter Zugrundelegung von (B) der Satz (a) mit (a_1) *induktiv äquivalent* ist. Dazu wird (a_1) zunächst logisch umgeformt in:

(a_2) „$\Lambda\, x\, \Lambda\, t\, [x \varepsilon Zr \rightarrow (\langle x, t\rangle \varepsilon Wg \rightarrow \langle x, t\rangle \varepsilon Al)]$"

Wegen (B) ist (a_2) äquivalent mit:

(a_3) „$\Lambda\, x\, \Lambda\, t\, [x \varepsilon Zr \rightarrow (\langle x, t\rangle \varepsilon Wg \rightarrow x \varepsilon Wl)]$"

Dieser Satz ist nun logisch äquivalent mit:

(a_4) „$\Lambda\, x\, [V\, t\, (\langle x, t\rangle \varepsilon Wg) \rightarrow (x \varepsilon Zr \rightarrow x \varepsilon Wl)]$"

Zum Zwecke der kürzeren Schreibweise sei der Eigenschaftsausdruck „irgendwann in Wasser gegeben" („Wi") folgendermaßen definiert:

(C) „$\Lambda\, x\, [x \varepsilon Wi \leftrightarrow V\, t\, (\langle x, t\rangle \varepsilon Wg)]$"

Dann ist (a_4) äquivalent mit:

(a_5) „$\Lambda\, x\, [x \varepsilon Wi \rightarrow (x \varepsilon Zr \rightarrow x \varepsilon Wl)]$"

Aus (a) folgt logisch (a_5), doch nicht umgekehrt, da weder „$x \varepsilon Wi$" noch „$x \varepsilon Zr \rightarrow x \varepsilon Wl$" logisch wahr ist. Es muß also versucht werden, (a) auf induktivem Weg von (a_5) aus zu erreichen. Wegen Theorem 20 ist es zwecklos, als Hypothese eines induktiven Arguments den Satz (a_5) oder (a) zu verwenden, da diese durch adäquate induktive Methoden in *unendlichen* Gegenstandsbereichen den Wert 0 erhalten; man kann jedoch auch unendliche Bereiche durch hinreichend große *endliche* Gesamtheiten beliebig approximieren, weshalb im folgenden angenommen wird, daß er endlich ist.[32] Da uns bei generellen Behauptungen keine Mittel zur Verfügung stehen, um die Irrelevanz der Bedingung „$x \varepsilon Wi$" für „$x \varepsilon Zr \rightarrow x \varepsilon Wl$" zu zeigen, sind wir genötigt, von der auf Grund des Erfahrungsdatums „Xa_k" akzeptierten Allbehauptung (a_5) zum singulären Voraussageschluß überzugehen. Nun ist ein nahe bei 1 gelegener c-Wert beim singulären Voraussageschluß auf keinen Fall eine hinreichende Bedingung dafür, daß eine Allbehauptung auf Grund eines Erfahrungsdatums akzeptiert werden kann, und im allgemeinen ist er auch nicht einmal eine notwendige Voraussetzung dafür, wie schon gezeigt worden ist. Es gibt jedoch Fälle, in denen er als notwendige Voraussetzung für die Akzeptierung eines Gesetzes angesehen werden kann, und es werden im nächsten Abschnitt die Bedingungen erörtert, unter denen dies zutrifft. Ich nehme hier an, daß diese Bedingungen erfüllt sind; dann gilt für eine reguläre Bestätigungsmethode c [die eine L-Methode oder eine durch L-Methoden definierbare induktive Methode sein wird, wobei wegen der vorausgesetzten (durchgehenden oder fast durchgehenden) Uniformität der Teilbereiche hinsichtlich Wl von nahe bei 0 liegenden reellen Zahlen L ausgegangen werden darf]:

(a_6) $c(„a_{k+1} \varepsilon Wi \rightarrow (a_{k+1} \varepsilon Zr \rightarrow a_{k+1} \varepsilon Wl)", „Xa_k") = 1 - r$

Dabei ist r eine sehr kleine reelle Zahl, d. h. es gilt etwa: $r < 0{,}005$. Nach dem *Prinzip von Cournot* ist man hier berechtigt, einen kleinen Fehler zu machen und $1 - r = 1$ zu setzen; dann gilt:

(a_7) $c(„a_{k+1} \varepsilon Wi \rightarrow (a_{k+1} \varepsilon Zr \rightarrow a_{k+1} \varepsilon Wl)", „Xa_k") = 1$

---

32 Für mathematische Zwecke, d. h. zur Anwendung der Hilfsmittel der Analysis, ist es gelegentlich *notwendig*, von der Idealisierung auszugehen, daß der Gegenstandsbereich *unendlich* ist; für induktive Argumente, die nicht mit Sätzen operien, in denen All- oder Existenzquantoren vorkommen, ist diese Idealisierung ohne Belang und liefert keine Verfälschung der c^L-Werte.

Unter Verwendung des Lehrsatzes 4 erhält man dann:

(a_8) $c(\text{„}a_{k+1} \varepsilon \text{Wi} \rightarrow (a_{k+1} \varepsilon \text{Zr} \rightarrow a_{k+1} \varepsilon \text{Wl})\text{“}, \text{„}X^{a}{}_{k}\text{“}) =$
$c(\text{„}a_{k+1} \varepsilon \text{Zr} \rightarrow a_{k+1} \varepsilon \text{Wl}\text{“}, \text{„}a_{k+1} \varepsilon \text{Wi} \wedge X^{a}{}_{k}\text{“})$

Aus intuitiven Überlegungen heraus ergibt sich nun, daß „$a_{k+1} \varepsilon$ Wi“ für „$a_{k+1} \varepsilon$ Zr $\rightarrow a_{k+1} \varepsilon$ Wl“ hinsichtlich „$X^{a}{}_{k}$“ irrelevant ist. Im formalen System wird dies entweder so wiedergegeben, daß sich diese Irrelevanz auf Grund des Datums „$X^{a}{}_{k}$“ ergibt, daß dieses also noch zusätzlich aussagt, daß bei den bisher in Wasser gegebenen Dingen etwa gleichviele sich aufgelöst bzw. nicht aufgelöst haben, oder aber c wird so charakterisiert, daß es keinen Analogieeinfluß von „$a_{k+1} \varepsilon$ Wi“ auf „$a_{k+1} \varepsilon$ Zr $\rightarrow a_{k+1} \varepsilon$ Wl“ wiedergibt.[33] Dann gilt:

(a_9) $c(\text{„}a_{k+1} \varepsilon \text{Zr} \rightarrow a_{k+1} \varepsilon \text{Wl}\text{“}, \text{„}a_{k+1} \varepsilon \text{Wi} \wedge X^{a}{}_{k}\text{“}) =$
$c(\text{„}a_{k+1} \varepsilon \text{Zr} \rightarrow a_{k+1} \varepsilon \text{Wl}\text{“}, \text{„}X^{a}{}_{k}\text{“})$

Aus diesen Gleichungen folgt nun:

(a_{10}) $c(\text{„}a_{k+1} \varepsilon \text{Zr} \rightarrow a_{k+1} \varepsilon \text{Wl}\text{“}, \text{„}X^{a}{}_{k}\text{“}) = 1$

Korrigiert man den durch die Anwendung des Prinzips von Cournot gemachten Fehler, so erhält man schließlich:

(a_{11}) $c(\text{„}a_{k+1} \varepsilon \text{Zr} \rightarrow a_{k+1} \varepsilon \text{Wl}\text{“}, \text{„}X^{a}{}_{k}\text{“}) = 1 - r$

Also wird (a) auf Grund des Datums „$X^{a}{}_{k}$“ akzeptiert, da der Wert des singulären Voraussageschlusses nahe genug bei 1 liegt und die anderen Bedingungen zur Akzeptierung der Gesetzesaussage auf Grund dieses Datums als erfüllt vorausgesetzt worden sind. In diesem Sinne kann gesagt werden, daß (a) aus (a_1) *induktiv folgt*. Umgekehrt impliziert (a) deduktiv (a_5) und somit auch (a_1). Nach Theorem 1 (k) gilt somit:

$c(\text{„}\Lambda x\,[x \varepsilon \text{Zr} \rightarrow x \varepsilon \text{Wl}]\text{“}, \text{„}X^{a}{}_{k}\text{“}) \leqslant$
$\leqslant c(\text{„}\Lambda x\,[x \varepsilon \text{Zr} \rightarrow \Lambda t\,(\langle x, t\rangle \varepsilon \text{Wg} \rightarrow \langle x, t\rangle \varepsilon \text{Al})]\text{“}, \text{„}X^{a}{}_{k}\text{“}),$

so daß auch (a_1) aus (a) *induktiv folgt* (der oben gegebene Beweis, daß (a) aus (a_1) induktiv folgt, ist allerdings auch umkehrbar, wie man

[33] c ist dann nach den am Schluß des 4. Abschnittes des VI. Kapitels gemachten Bemerkungen eine aus verschiedenen L-Methoden zusammengesetzte Funktion. Im übrigen wird gar nicht die Gleichung (9) und somit *vollständige* Irrelevanz benötigt, sondern lediglich eine Approximation, also *fast vollständige* Irrelevanz, die bewirkt, daß der Wert $c(\text{„}a_{k+1} \varepsilon \text{Zr} \rightarrow a_{k+1} \varepsilon \text{Wl}\text{“}, \text{„}X^{a}{}_{k}\text{“})$ in der Nähe von $1 - r$ und so nahe bei 1 liegt, daß er nach dem Prinzip (d. h. nach dem Fehlverhalten) von Cournot immer noch mit der 1 identifiziert werden kann. Diese Voraussetzung, daß „$a_{k+1} \varepsilon$ Wi“ für „$a_{k+1} \varepsilon$ Zr $\rightarrow a_{k+1} \varepsilon$ Wl“ bezüglich „$X^{a}{}_{k}$“ *fast irrelevant* ist, dürfte empirisch, d. h. auf Grund dessen, was die Prämisse „$X^{a}{}_{k}$“ aussagt, immer erfüllt sein.

unmittelbar sehen kann, so daß der Rückgriff auf Theorem 1 (k) hier vermieden werden kann); also ist (a) mit (a_1) auf der Grundlage von (B) *induktiv äquivalent.*

Ähnlich wird nun die Verwendbarkeit des Begriffs „wasserlöslich" auf nicht in Wasser gegebene Holzstücke erweitert, d. h. es wird gezeigt, daß die folgenden Sätze (b) und (b_1) induktiv äquivalent sind:

(b) „$\Lambda\, x\, [x \varepsilon Hz \rightarrow \neg\, (x \varepsilon Wl)]$"

(b_1) „$\Lambda\, x\, [x \varepsilon Hz \rightarrow \Lambda\, t\, (\langle x, t\rangle \varepsilon Wg \rightarrow \neg\, (\langle x, t\rangle \varepsilon Al))]$"

Genauso wird für die anderen Arten (Kochsalz, Eisen, Kunstleder usw.) vorgegangen.

Da vorausgesetzt worden ist, daß die Aufteilung des Bereichs der Dinge *vollständig* war (zumindest, soweit es sich um bekannte Arten von Dingen handelt), folgt aus der induktiven Äquivalenz von (a) mit (a_1), von (b) mit (b_1) usw., daß der in (a), (b) usw. vorkommende Teilsatz „$x \varepsilon Wl$" mit dem in (a_1), (b_1) usw. vorkommenden Teilsatz „$\Lambda\, t\, (\langle x, t\rangle \varepsilon Wg \rightarrow \langle x, t\rangle \varepsilon Al)$" induktiv äquivalent ist, gleichgültig, wie man „x" interpretiert, da jeder Gegenstand Element einer der vorgegebenen uniformen Teilklassen ist, für die die induktive Gleichwertigkeit schon nachgewiesen worden ist. Da auf diese Weise das Schema (A), ausgehend von (B), induktiv erreicht worden ist, kann das vorgenommene Verfahren als adäquat angesehen werden.

Wesentlich dafür, daß man (a) als induktiv äquivalent mit (a_5) und damit auch mit (a_1) ansehen kann, ist die Irrelevanzbetrachtung bezüglich der Zusatzprämisse „$a_{k+1} \varepsilon Wi$". Da analoge Betrachtungen bei Begriffen, die in der deduktiven Logik oder in der Mathematik durch bedingte Definitionen eingeführt werden, nicht durchführbar sind, bleibt die Intension dieser Begriffe auf die Fälle beschränkt, in denen die Bedingung der Definition erfüllt ist.

Hinsichtlich der Frage, ob „Wl" als Grundprädikat oder als definiertes Prädikat anzusehen ist, tut man sich bei der obigen Argumentation am leichtesten, wenn man (B) als ein Axiom ansieht, das (mehr oder weniger zufällig) gewissen Bedingungen genügt wie etwa der der partiellen Eliminierbarkeit; in diesem Fall rechnet man „Wl" zu den Grundprädikaten. Im Prinzip kann so auch bei Explizitdefinitionen vorgegangen werden: Man rechnet die definierten Begriffe zu den Grundprädikaten und fügt die Definitionen den Axiomen der Theorie hinzu.

Es darf nicht verschwiegen werden, daß die Anwendbarkeit des Verfahrens der induktiven Erweiterung der partiell charakterisierten Intension eines Ausdrucks an zwei Grenzen stößt. Es kann der Fall eintreten, daß man eine Art von Dingen entdeckt, von denen noch keines der Testbedingung unterworfen worden ist. Gelegentlich kann man durch Analogiebetrachtungen (also durch spezielle induktive Argumente) vom einen auf den anderen Stoff schließen, etwa wenn der eine aus dem anderen durch den Austausch eines Kaliumatoms gegen ein Natriumatom entsteht; hierzu zählt wohl auch das sogenannte *mathematische Berechnen* der Wasserlöslichkeit der Gegenstände dieser Art: Es werden mathematisch Ähnlichkeiten der äußeren Elektronenbahnen zweier verschiedener Stoffe errechnet und sodann durch ein Analogieargument aus der Wasserlöslichkeit der Elemente des einen auf die der Elemente des anderen geschlossen. — Ist der neuentdeckte Stoff jedoch so andersartig im Vergleich mit allen bekannten, daß keine derartigen Betrachtungen möglich sind, so kann das Verfahren der induktiven Erweiterung der Intension eines derartigen Prädikates nicht angewandt werden. Das gleiche gilt, wenn es nicht möglich ist, den Gegenstandsbereich von vornherein so aufzuteilen, daß die Teilklassen hinsichtlich der fraglichen Disposition (fast) durchgehend uniform sind; der Bereich der Menschen etwa dürfte hinsichtlich der Disposition *intelligent* nicht in dieser Weise aufteilbar sein. — Ich habe jedoch den Eindruck, daß man in diesen beiden Fällen weder im Alltag noch in den Wissenschaften derartige Prädikate verwendet, daß man also den Elementen jener neuen Art etwa die Wasserlöslichkeit zunächst (d. h. bis zur Überprüfung der ersten Fälle) weder zu- noch abspricht und daß man es nicht für sinnvoll hält, etwa einem kurz nach der Geburt gestorbenen Kind oder sonst einem hinsichtlich seiner Intelligenz nie auf die Probe gestellten Menschen Intelligenz zu- oder abzusprechen. Wenn diese Vermutung begründet ist, kann jenes Verfahren als Explikation dessen betrachtet werden, was wir in solchen Fällen intuitiv tun. Im anderen Fall ist es lediglich eine *partielle* Rekonstruktion unseres faktischen Vorgehens.

Bei anderen Dispositionsprädikaten werden sich gelegentlich Zusatzkomplikationen ergeben. Ein Beispiel dafür ist der Begriff der objektiven Wahrscheinlichkeit, der in einer genügend ausdrucksreichen Objektsprache eingeführt werden kann. Zunächst wird man dabei an

folgende Definition denken,[34] bei der „Ntz" für „natürliche Zahl", „Rez" für „reelle Zahl" und „rh" für „relative Häufigkeit" steht:
„Λ F Λ G Λ r [r ε Rez → [w (F, G) = r ↔
Λ s V n Λ m Λ H (s ε Rez ∧ n ε Ntz ∧ m ε Ntz ∧ m > n ∧ H ⊆ G ∧
H ε m → r — s ⩽ rh (F, H) ⩽ r + s)]]"
Die relative Häufigkeit von F in H ist identisch mit dem Verhältnis der Anzahl der Dinge, die sowohl in F als auch in H sind, zur Anzahl der Dinge in H; rh (F, H) ist also stets eine rationale Zahl.
Unter Zugrundelegung dieser Definition ergibt sich dann folgendes Paradoxon: Gegeben sei ein Würfel, der bald nach seiner Herstellung verbrannt ist und mit dem zuvor wegen widriger Umstände nie gewürfelt worden ist. G sei die Gesamtheit der mit ihm gemachten Würfe, d. h. G sei mit der leeren Klasse identisch. Dann ist für jede von der 0 verschiedenen natürlichen Zahl m der Satz „H ε m" falsch und somit das ganze Definiens wahr. Folglich gilt dann der Satz „w (F, G) = r", gleichgültig, für welche reelle Zahl der Ausdruck „r" steht, was absurd ist.
Es liegt daher nahe, diese Definition so umzuformen, daß sie dem Schema (B) entspricht. Hierbei stört zunächst der Existenzquantor „V n". Er wird, nach dem Vorgehen Hilberts, folgendermaßen eliminiert: Intuitiv ist damit die kleinste natürliche Zahl n gemeint, für die das auf den Quantor Folgende gilt; welche Zahl dies ist, hängt jedoch davon ab, welche reelle Zahl s man zuvor gewählt hat (ist s groß, so genügt ein kleines n, und ist s klein, so muß n groß gewählt werden). Offensichtlich muß die Abhängigkeit der Zahl n von s genau bestimmt werden, bevor man mit der Überprüfung der Elemente von G auf die Eigenschaft F hin beginnt; denn es wäre unfair von mir, zu behaupten, daß die relativen Häufigkeiten von *Kopf* beim Münzwurf schon irgendwann im Intervall [0,995; 0,990] bleiben werden, wenn man nur lange genug die Münze wirft. Die Aussage ist unentscheidbar, wenn ich nicht zuvor sage, was „lange genug" heißt. Man kann bei der Abhängigkeit der Zahl n von s sehr großzügig sein und sie etwa durch eine Funktion f wiedergeben, die für 0,1 den Wert $f(0{,}1) = 10^{10}$, für 0,01 den Wert $f(0{,}01) = 10^{10^{10}}$ annimmt usw. Setzt man eine derartige Funktion f (deren Werte natür-

---

[34] Vgl. z. B. Mises [2], S. 14f.

liche Zahlen sind) als vorgegeben voraus, so geht die obige Definition in den folgenden Satz über:

„Λ F Λ G Λ r [r ε Rez → [w (F, G) = r ↔
Λ s Λ m Λ H (s ε Rez ∧ m ε Ntz ∧ m > f (s) ∧ H ⊆ G ∧ H ε m →
r — s ⩽ rh (F, H) ⩽ r + s)]]"

Den oben geschilderten Schwierigkeiten entgeht man, wenn man statt dieser Explizitdefinition eine bedingte Definition wählt:

„Λ F Λ G Λ r Λ s Λ m Λ H [r ε Rez ∧ s ε Rez ∧
m ε Ntz ∧ m > f (s) ∧ H ⊆ G ∧ H ε m →
[w (F, G) = r ↔ r — s ⩽ rh (F, H) ⩽ r + s]]"

Damit ist zunächst die Intension des Funktionsausdrucks „w" nur für die beobachteten Teilklassen H der verschiedenen Bezugsklassen G festgelegt. Durch einen Induktionsschluß, der darauf basiert, daß diese Bedingung irrelevant ist, wird sie auch auf den nicht beobachteten Teil erweitert. *Der Begriff der objektiven Wahrscheinlichkeit ist somit ein Dispositionsprädikat wie der Begriff der Wasserlöslichkeit.*

Hiermit werden freilich mehr Schwierigkeiten aufgezeigt als gelöst. Eines davon ist das Problem, daß sich Dispositionen im Laufe der Zeit gelegentlich ändern. So nimmt die menschliche Intelligenz im allgemeinen bis zum 30. Lebensjahr stark zu und vermindert sich meist mit zunehmendem Alter; der Intelligenzgrad, getestet durch bestimmte Verfahren und gemessen durch die sogenannten Intelligenzquotienten, ändert sich also im Laufe des Lebens. Da eine Münze beliebig deformiert werden kann, wenn man sie nur genügend lange wirft, ändert sich im Laufe der Zeit auch ihre Disposition, daß die objektive Wahrscheinlichkeit von *Kopf* in der Gesamtheit der Würfe genau $^1/_2$ ist. Da die mit f (s) identische natürliche Zahl n demnach *nicht* beliebig groß bewählt werden kann, ist der Wahrscheinlichkeitsgrad prinzipiell nicht genau (d. h. nur in gewissen Grenzen genau) bestimmbar, ähnlich wie ja auch der Intelligenzgrad nur in gewissen Grenzen genau bestimmbar ist. Mit welchen Grenzen wir uns dabei in jedem einzelnen Fall begnügen, dürfte eine theoretisch kaum zu beantwortende Frage sein.

Wenn bisher von Hypothesen die Rede gewesen ist, so sind damit selten Allbehauptungen, sondern fast immer singuläre Aussagen gemeint gewesen, die nicht zu dem Erfahrungsdatum gehören, das die Prämisse eines induktiven Arguments ausmacht und deren Wahrscheinlichkeitsgrad im Hinblick auf eben dieses Datum ermittelt werden soll. Im folgenden möchte ich noch kurz — ohne Zweifel *zu kurz* — einige Möglichkeiten zur Begründung von Allbehauptungen (von *Theorien*) aufzeigen bzw. darauf hindeuten.

Ich möchte (grob und ohne Anspruch auf Präzision) auf zwei Gruppen von Theorien hinweisen: Auf die Gruppe jener allgemeinen Behauptungen, die wir (im Moment) als kontingent ansehen, und die Gruppe jener Gesetzesaussagen, an deren Wahrheit wir (im Moment) auf jeden Fall festhalten werden, die wir also (in einem pragmatischen, nicht im Kantischen Sinn) als a priori wahr ansehen. In der Praxis ist oft nur schwer zu entscheiden, ob eine Theorie zur einen oder zur anderen Gruppe gehört, und zwar nicht nur im Alltag, sondern auch etwa in den heutigen Lehrbüchern der theoretischen Physik. Da aber kaum ein Zweifel daran aufkommen kann, daß die Sätze der einen Gruppe anders begründet werden müssen als die der anderen, muß hier notwendigerweise davon ausgegangen werden, daß eine solche Entscheidung getroffen werden kann. Im übrigen ist es zum Teil auch eine Frage der Zeit, wann in der Physik ein Präzisionsgrad erreicht wird, der den augenblicklichen Normen der Logik entspricht. Während diese Erfahrungswissenschaft noch um die Jahrhundertwende lediglich eine mit mathematischen Begriffen angereicherte Sprache des Alltags benützte, entwickelt sie sich seit den Untersuchungen Reichenbachs[35] immer mehr zu einer axiomatischen Wissenschaft.

Ohne Zweifel können die Allbehauptungen der ersten Gruppe, die wir als kontingent ansehen, nicht durch Aprioriüberlegungen begründet werden. Wenn etwa die These „Alle Lebewesen mit Herz sind Lebewesen mit Nieren" aufgestellt wird, so ist damit nicht intendiert, die darin vorkommenden Begriffe „Lebewesen mit Herz" und „Lebe-

---

[35] Vgl. Hans Reichenbach, Axiomatik der relativistischen Raum-Zeit-Lehre, Braunschweig [1]1924, [2]1965.

wesen mit Nieren" implizit zu charakterisieren, also ihre Intention (teilweise) festzulegen, sondern es wird behauptet, daß diesem Satz auf der Grundlage unseres Beobachtungswissens eine gewisse *Wahrscheinlichkeit* zukommt, d. h. ein Wahrscheinlichkeitsgrad, der nur um wenig kleiner als 1 ist. Da in solchen Behauptungen offensichtlich der logische Wahrscheinlichkeitsbegriff verwendet wird, muß erklärt werden, wie diese vom Standpunkt der *formalen* induktiven Logik aus zu beurteilen sind; da sie darüber hinaus als sinnvoll angesehen werden, muß gezeigt werden, wie das Theorem 20 und seine Konsequenzen umgangen werden können.

Am einfachsten ist dieses Problem zu lösen, wenn der Gegenstandsbereich B eine endliche Anzahl n hat und die Anzahl k der beobachteten Dinge sehr groß ist. Die Mindestgröße, die k haben muß, hängt von unserer Wahl des Faktors L ab: Ist L eine sehr kleine Zahl, so braucht k nicht allzu groß zu sein, damit eine Allbehauptung $\Lambda\,\xi\,\Psi\,(\xi)$ bezüglich Φ_k einen c^L-Wert erhält, der nur um eine kleine Zahl s von 1 verschieden ist, und wenn L groß ist, muß auch die Beobachtung entsprechend umfangreich sein.

Wenn eine Methode c^L mit relativ großem L verwendet wird, aber nur ein vergleichsweise kleiner Teil der Dinge des Bereichs untersucht worden ist, so hat man zwei Möglichkeiten: Man kann entweder versuchen, die Anzahl der Beobachtungen zu erhöhen, oder aber, wenn dies nicht möglich ist, durch Analogiebetrachtungen einen kleineren Parameter L′ zu erhalten und damit eine Funktion $c^{L'}$ verwenden zu können, deren Werte für den singulären Voraussageschluß sich rascher an die beobachteten relativen Häufigkeiten angleichen als die von c^L.

Gegeben seien etwa 10 Urnen mit je 100 einfarbigen Kugeln. Wenn ich nun aus einer dieser Urnen eine violette Kugel ziehe, so wird dieses Datum $\Psi\,(\alpha_1)$ für mich zunächst keine zureichende Grundlage dafür sein, zu behaupten, daß auch alle anderen 99 Kugeln dieser Urne violett sind; d. h. meine Bestätigungsmethode c^L wird von der Art sein, daß der Wahrscheinlichkeitswert der Theorie $\Lambda\,\xi\,\Psi\,(\xi)$, die besagt, daß alle Kugeln dieser Urne violett sind, hinsichtlich $\Psi\,(\alpha_1)$ ganz beträchlich kleiner ist als 1. Wenn ich nun aber (durch eigene Beobachtung oder durch Indiskretion eines anderen) erfahre, daß die übrigen 9 Urnen durchgehend uniform sind, also nur jeweils Kugeln

von einer Farbe enthalten (wobei etwa 2 Urnen nur gelbe Kugeln enthalten, drei nur rote, eine nur blaue, und drei nur orangefarbene), dann werde ich den Uniformitätsgrad der Kugeln der vorgegebenen Urne anders einschätzen: Ich werde annehmen, daß er mit 1 identisch oder nur um wenig kleiner als 1 ist[36] und entsprechend c^0 oder eine Funktion $c^{L'}$ mit kleinem L′ wählen. In diesem Fall erhalte ich: $c^L(\Lambda\,\xi\,\Psi(\xi), \Psi(\alpha_1)) = 1 - r$, wobei r eine verschwindend kleine Zahl ist, die man ohne großes Risiko vernachlässigen kann.
Wenn uns hingegen derartige Vergleichsmöglichkeiten und damit derartige induktive Argumente zur Ersetzung von c^L durch eine andere Methode $c^{L'}$ nicht zur Verfügung stehen, können wir der Allbehauptung auf Grund eines kleinen Datums Φ_k keine große Sicherheit beimessen und sie also nicht auf Grund dieses Datums akzeptieren.

Wenn der Gegenstandsbereich unendlich ist, sind derartige Überlegungen zunächst nicht anwendbar. Man kann sich jedoch auf folgende Weise behelfen: Selten wird man sich dafür interessieren, was *alle* Gegenstände ohne Einschränkung für Eigenschaften haben. Für unser praktisches Leben ist stets nur wichtig, wie es mit endlich vielen von ihnen steht. So etwa spielt die Frage, ob die Gravitationskonstante auch nach zwei Millionen Jahren noch den gleichen Wert wie jetzt hat oder nicht, ob sie also tatsächlich eine Konstante oder eine Variable (d. h. hier: *eine Funktion!*) ist, für die Probleme irgendeines Menschen oder auch der gesamten Menschheit kaum eine Rolle. Der Naturwissenschaftler wird gerne bereit sein, für die Hypothese, daß sich der

---

[36] Es handelt sich bei diesem Analogieschluß also um einen Induktionsschluß höherer Ordnung: Auf Grund des Wissens um die 9 anderen Uniformitäten wird induktiv auf die Uniformität des letzten Gegenstandsbereichs geschlossen. Wird bei diesem Schluß eine Funktion c^0 verwendet, so erhält man als Schätzwert den Uniformitätsgrad 1, und wenn man eine Funktion $c^{L'}$ mit relativ kleinem L′ benützt, so wird der Wert nur um weniges unter 1 liegen. Nimmt man hingegen an, daß die *Gesamtheit der zehn Urnen* ein (ganz oder fast ganz) disuniformer Gegenstandsbereich ist, dann wird man die Uniformität der letzten Urne mittels c^∞ oder mittels einer Funktion $c^{L'}$ mit großem L′ schätzen; man erhält dann das Ergebnis, daß das Wissen um die Uniformitäten der 9 Urnen ohne (große) Relevanz für die Frage der Uniformität der letzten Urne ist, und man wird daher für diese einen mittleren Uniformitätsgrad annehmen und somit eine Funktion $c^{L'}$ wählen, bei der L′ zwischen 1 und 2 liegt. Mit dieser Funktion $c^{L'}$ erhält dann $\Lambda\,\xi\,\Psi(\xi)$ hinsichtlich $\Psi(\alpha_1)$ einen relativ kleinen Wert.

Wert dieser Konstanten in den nächsten 100000 Jahren nicht merklich ändert, auf Grund seiner Erfahrungen (und seiner Einschätzung der Uniformität der Natur) eine beliebig hohen Grad an Sicherheit und Glaubwürdigkeit anzugeben bzw. diesen Grad sogar als (praktisch) mit der 1 identisch anzusehen. Er wird jedoch die Hypothese, daß dies auch für die nächsten 300 Trillionen Jahre gilt, im Hinblick auf die gleiche Erfahrung (und die gleiche Einschätzung der Uniformität) vielleicht schon weniger sicher beurteilen, und er wird sich weigern, der Hypothese, daß jene Konstante *stets und immer* konstant sein wird, auf Grund seiner Erfahrungen (und seiner Einschätzung der Uniformität) irgendeinen positiven Grad an Sicherheit zu geben. Selbstverständlich wird er auch der Hypothese, daß sich diese Konstante in ferner Zukunft einmal in einer ganz bestimmten Weise ändert, keinen größeren Sicherheitsgrad zumessen. Er wird allerdings, und, wie mir scheint, nicht ganz zu Unrecht, bemerken, daß solche Hypothesen niemanden interessieren, und auch von niemandem hinsichtlich ihres Sicherheitsgrades beurteilt werden können.

Er verwendet also eine Funktion c^L mit einem verschwindend kleinen L, das aber durchaus verschieden von der 0 ist. Wenn er den Bereich der Experimente auf die Gesamtheit der in den nächsten 100000 Jahren gemachten beschränkt, so rechnet er mit keinem Gegenbeispiel und ist bereit, jede Wette darauf einzugehen. Für die unendlich vielen möglichen Experimente, die theoretisch gemacht werden können, wenn die Menschheit nicht ausstirbt, ist er jedoch nicht bereit, ein Risiko einzugehen.[37]

Die Akzeptierung eines generellen Gesetzes setzt voraus, daß unsere Beobachtungen keine Gegenbeispiele enthalten. Dieser Fall wird jedoch kaum jemals eingetreten sein. In solchen Fällen sind wir aber geneigt, gegenteilige Beobachtungssätze aus dem Korpus der von uns als gesichert angesehenen Aussagen zu entfernen. Dieses Verfahren der Eliminierung gegenteiliger Beobachtungen wird von uns mit großem intuitiven Geschick, aber zunächst ganz irrational angewandt, nämlich im Hinblick auf die durch Beobachtungen erst zu rechtfertigende Hypothese. So fegen wir die im Mittelalter gemachten Beobachtungen, daß einigen Menschen nach bestimmten rituellen Hand-

[37] Dies ist, wie mir scheint, die adäquate Deutung der Aussagen Carnaps in [9], S. 571 ff.

lungen abgeschlagene Glieder wieder nachgewachsen sind, mit dem Hinweis vom Tisch, daß so etwas unseren als wahr angenommenen Naturgesetzen widerspricht.[38] Der einzige Weg, auf dem man hier einem Zirkel entgehen kann, ist vermutlich der, daß man auf die sich im Laufe der Zeit ändernden relativen Häufigkeiten in der Gesamtheit der Beobachtungen hinweist, d. h. darauf, daß mit zunehmenden Beobachtungen die relative Häufigkeit der Ausnahmen (zum Gesetz) in der Gesamtheit der Beobachtungen immer mehr gegen 0 konvergiert.
Es wird also für eine fest vorgegebene kleine Toleranzschranke s (etwa 0,05 oder 0,005)[39] angenommen, daß es eine natürliche Zahl i gibt, die wesentlich kleiner als k ist, so daß für die generelle Hypothese Ψ die Werte $c(\Psi, \Phi_i)$ $c(\Psi, \Phi_{i+1})$, ..., $c(\Psi, \Phi_{k-1})$, $c(\Psi, \Phi_k)$ alle größer als $1 - s$ sind. Durch ein metatheoretisches induktives Argument wird dann geschlossen, daß dies auch bei allen übrigen (endlich vielen) Werten $c(\Psi, \Phi_{k+j})$ so sein wird; des weiteren wird angenommen, daß sich die Toleranzschranke s bei entsprechend größerem k beliebig verkleinern läßt. In diesem Fall sind wir bereit, das (ausnahmslose Gültigkeit beanspruchende) Naturgesetz zu *akzeptieren* und alle gegenteiligen Beobachtungen als *unkontrollierbare Störungen des Versuchsablaufs* oder als *singuläre* bzw. *okkulte* Effekte abzutun und die sie beschreibenden Sätze aus der Gesamtheit der von uns akzeptierten Aussagen zu streichen. Mit dieser Operation haben wir dann erreicht, daß uns nur noch solche Fälle zur Verfügung stehen, die für das Gesetz positiv sind.
Ähnlich begründen wir auch statistische Gesetze (die in unserer einfachen Sprache als Adjunktionen von Konjunktionen formuliert werden müssen), indem wir zeigen, daß die Schätzung der relativen Häufigkeit einer Eigenschaft in einer Bezugsklasse schon längere Zeit in einem kleinen Intervall s um die reelle Zahl r (mit $0 < r < 1$) liegt und daß die Zahl s beliebig klein gewählt werden kann, falls man hinreichend große Anzahlen k von Beobachtungen zur Verfügung hat. Da $r + s$ stets kleiner als 1 ist, besteht demnach keine Veranlassung, mißliebige Erfahrungsdaten mittels undurchsichtiger Methoden als falsch zu deklarieren.

---

[38] Vgl. Hume [2], S. 123 f.
[39] Dies sind sogenannte *Erfahrungswerte*, die sich *bisher* in der Praxis gut bewährt haben; eine bessere Rechtfertigung für sie existiert nicht.

Die Lücke im Lösungsvorschlag zum Problem der Dispositionsprädikate kann nun folgendermaßen ausgefüllt werden: Es wird vorausgesetzt, daß der Gegenstandsbereich hinsichtlich der fraglichen Eigenschaften einen sehr großen (nur wenig von 1 verschiedenen, aber nicht mit 1 identischen) Uniformitätsgrad besitzt, so daß eine Funktion c^L mit kleinem L fair ist. Dabei möge sich ergeben, daß

$c^L(\text{„}\Lambda x\,[x \varepsilon \mathrm{Wi} \to (x \varepsilon \mathrm{Zr} \to x \varepsilon \mathrm{Wl})]\text{“}, \text{„}X^a{}_k\text{“}) = 1 - s,$

wobei $1 - s$ so nahe bei 1 liegt, daß wir bereit sind, einen kleinen Fehler zu machen und $1 - s = 1$ zu setzen. Nun gilt:

$$
\begin{aligned}
&c^L(\Lambda\,\xi\,\Delta(\xi), \Delta(\alpha_1) \wedge \ldots \wedge \Delta(\alpha_k)) = \\
&\quad = c^L(\Delta(\alpha_1) \wedge \ldots \wedge \Delta(\alpha_k) \wedge \Delta(\alpha_{k+1}) \wedge \ldots \wedge \Delta(\alpha_n), \Delta(\alpha_1) \wedge \ldots \wedge \Delta(\alpha_k)) \\
&\quad = c^L(\Delta(\alpha_1), \Delta(\alpha_1) \wedge \ldots \wedge \Delta(\alpha_k)) \cdot \ldots \cdot \\
&\qquad \cdot c^L(\Delta(\alpha_k), \Delta(\alpha_1) \wedge \ldots \wedge \Delta(\alpha_k) \wedge \Delta(\alpha_1) \wedge \ldots \wedge \Delta(\alpha_{k-1}) \cdot \\
&\qquad \cdot c^L(\Delta(\alpha_{k+1}), \Delta(\alpha_1) \wedge \ldots \wedge \Delta(\alpha_k) \wedge \Delta(\alpha_1) \wedge \ldots \wedge \Delta(\alpha_k)) \cdot \ldots \cdot \\
&\qquad \cdot c^L(\Delta(\alpha_n), \Delta(\alpha_1) \wedge \ldots \wedge \Delta(\alpha_k) \wedge \Delta(\alpha_1) \wedge \ldots \wedge \Delta(\alpha_{n-1})) \\
&\quad = 1 \cdot \ldots 1 \cdot c^L(\Delta(\alpha_{k+1}), \Delta(\alpha_1) \wedge \ldots \wedge \Delta(\alpha_k)) \cdot \ldots \cdot \\
&\qquad \cdot c^L(\Delta(\alpha_n), \Delta(\alpha_1) \wedge \ldots \wedge \Delta(\alpha_{n-1}))
\end{aligned}
$$

Der Wert für die erste singuläre Voraussage $\Delta(\alpha_{k+1})$ ist nun größer als der für den Allsatz, d. h. es gilt:

$$
\begin{aligned}
&c^L(\Lambda\,\xi\,\Delta(\xi), \Delta(\alpha_1) \wedge \ldots \wedge \Delta(\alpha_k)) = 1 - s < 1 - r = \\
&\quad = c^L(\Delta(\alpha_{k+1}), \Delta(\alpha_1) \wedge \ldots \wedge \Delta(\alpha_k)),
\end{aligned}
$$

und die Werte für die anderen singulären Voraussagen sind noch größer. Daher ist im letzten Abschnitt zu Recht der kleine Fehler gemacht und $1 - r = 1$ gesetzt worden (für die entsprechenden Gleichungen bezüglich der anderen singulären Voraussageschlüsse wäre der Fehler noch kleiner). Dann und nur dann, wenn die Zahlen r und s diese Größenordnungen haben, scheint es mir erlaubt zu sein, den Wert des singulären Voraussageschlusses als Gradmesser für den des Allsatzes zu nehmen.

Hintikka hat versucht, einen Begriff der induktiven Bestätigung zu charakterisieren, der mit dem Carnapschen Ansatz weitgehend übereinstimmt, der es aber gestattet, den Allbehauptungen auch in unendlichen Bereichen positive Wahrscheinlichkeitswerte zuzusprechen.[40] Gegeben sei eine Sprache, die die endlich vielen einstelligen Prädikate $\Pi_1, \ldots, \Pi_p$ und die (zunächst) endlich vielen Gegenstandsausdrücke

---

[40] Vgl. Hintikka [1], [2], [3], [4].

$\alpha_1, \ldots, \alpha_n$ besitzt. Es sind dann in ihr 2^p V-Prädikate $\Pi^*_1, \ldots, \Pi^*_q$ definierbar. Hintikka führt nun den Begriff des Konstituenten ein, den man folgendermaßen charakterisieren kann:
Ein Konstituent ist ein Satz der Form
$(\neg)\, V\, \xi\, (\xi \varepsilon \Pi^*_1) \wedge \ldots \wedge (\neg)\, V\, \xi\, (\xi \varepsilon \Pi^*_q)$,
eine Konjunktion von Existenzsätzen also, die teils negiert und teils unnegiert sind; offensichtlich gibt es 2^q derartige Konstituenten. Nicht alle von ihnen sind widerspruchsfrei. So ist etwa
$\neg\, V\, \xi\, (\xi \varepsilon \Pi^*_1) \wedge \ldots \wedge \neg\, V\, \xi\, (\xi \varepsilon \Pi^*_q)$
logisch falsch, da durch diese Aussage behauptet wird, daß es keine Dinge gibt [es ist ja der Satz $V\, \xi\, [\Phi\, (\xi) \vee \neg\, \Phi\, (\xi)]$, der besagt, daß es Dinge gibt, logisch wahr]. Wenn ferner $n < q$, so sind alle Konstituenten widerspruchsvoll, die mehr als n unnegierte Existenzsätze als Konjunktionsglieder enthalten. Ein Konstituent beschreibt, wieviele V-Prädikate *exemplifiziert* sind und wieviele nicht, wieviele von ihnen also nichtleere und wieviele von ihnen leere Eigenschaften bezeichnen. Er ist also eine allgemeinere Behauptung als eine statistische Verteilung (eine Strukturbeschreibung) oder gar eine individuelle Verteilung (eine Zustandsbeschreibung): Die erstere beschreibt die relativen Häufigkeiten der einzelnen Eigenschaften im Gegenstandsbereich und die letztere sagt sogar aus, welche Dinge die einzelnen Eigenschaften besitzen. Wie Hintikka gezeigt hat, ist jeder Satz der engeren Quantorenlogik, der keine Gegenstandsausdrücke frei enthält (in dem demnach quantifizierte Gegenstandsvariablen vorkommen), logisch äquivalent mit einer Adjunktion von Konstituenten und in diesem Sinn also in einer *adjunktiven distributiven Normalform* darstellbar.
Mit Carnaps Methode c^* in Gedanken und dem Ziel einer positiven Bestätigung von Allbehauptungen vor Augen hat Hintikka nun vorgeschlagen, zunächst eine Bewertung der einzelnen Konstituenten vorzunehmen, wobei die Summe dieser Werte 1 ist, sodann den Zahlenwert jedes Konstituenten zu gleichen Teilen auf die Strukturbeschreibungen aufzuteilen, die bei diesen Konstituenten gelten, und schließlich den Wert einer jeden Strukturbeschreibung wiederum zu gleichen Teilen auf die Zustandsbeschreibungen zu verteilen, die bei dieser Strukturbeschreibung gelten. Es liegt nahe, den Wert 1 auch auf die *widerspruchsfreien* Konstituenten gleichermaßen aufzuteilen

(und den *widerspruchsvollen* jeweils den Wert 0 zu geben); wie Hintikka jedoch hervorhebt, stolpert eine Person x, die eine so charakterisierte reguläre Bestätigungsmethode benützt, schon beim geringsten Anlaß in problematische Verallgemeinerungen. Es liegt daher nahe, diese Aufteilung anders vorzunehmen, und zwar so, daß einem Konstituenten ein um so geringerer Wert zugeordnet wird, je mehr negierte Existenzsätze er enthält. Hintikka hat vorgeschlagen, jedem der Konstituenten den Wert zu geben, den er durch eine Funktion c^{∞} bezüglich irgendeines (nicht des vorgegebenen) Universums von der Anzahl A erhält. Diese Anzahl A geht dann in die Definition als Parameter ein: Ist $A = 1$, so wird durchgehende Uniformität vorausgesetzt, da in diesem fiktiven Universum nur Allbehauptungen gelten; mit wachsendem A wird dieses Universum hingegen immer disuniformer, und mittels der dadurch festgelegten Folge von Bestätigungsmethoden ergeben sich immer zurückhaltendere Bestätigungsgrade.

Wenn man sich für $p = 1$ und $n \to \infty$ die Verteilungsfunktion vergegenwärtigt, so sieht man, daß sie an den Stellen 0 und 1 Sprünge hat: Der Konstituent $\neg V \xi (\xi \varepsilon \Pi_1)$ [genauer: $\neg V \xi (\xi \varepsilon \Pi^*_1) \wedge V \xi (\xi \varepsilon \Pi^*_2)$] hat für alle endlichen Bereiche B_n und für B_∞ den gleichen Wert, da (bei fest vorgegebenem A) die Verteilung der Aprioriwerte auf die Konstituenten ohne Berücksichtigung der Anzahl von B geschieht. Hat B die endliche und von der 0 verschiedene Anzahl n, so ist $\neg V \xi (\xi \varepsilon \Pi_1)$ gleichwertig mit $\neg (\alpha_1 \varepsilon \Pi_1) \wedge \ldots \wedge \neg (\alpha_n \varepsilon \Pi_1)$; ist B hingegen abzählbar unendlich, so wird durch ihn das gleiche ausgesagt wie durch $\neg (\alpha_1 \varepsilon \Pi_1) \wedge \neg (\alpha_2 \varepsilon \Pi_1) \wedge \ldots$. Während bei den L-Methoden der Aprioriwert für $\neg V \xi (\xi \varepsilon \Pi_1)$ mit wachsendem n gegen 0 konvergiert, bleibt er bei diesen regulären Bestätigungsmethoden konstant. Gleiches gilt für $V \xi (\xi \varepsilon \Pi_1)$, wie man unmittelbar sieht. Dies hat zur Folge, daß durch solche Methoden Allbehauptungen auch bei unendlichen Bereichen hinsichtlich eines endlichen Erfahrungsdatums positive Werte erhalten, die allerdings um so kleiner sind, je größer die natürliche Zahl A ist.

Obwohl der Ansatz Hintikkas attraktiv ist, scheinen doch folgende acht Punkte dagegen zu sprechen:

(1) Wie ich vorhin gezeigt habe, läßt sich das Problem der Bestätigung der kontingenten Allsätze (und nur um solche geht es mir zu-

nächst) auch dann in natürlicher Weise lösen, wenn man den in diesem Buch entwickelten Ansatz nicht verläßt. Man könnte gerechterweise entgegnen, daß sich eventuell bei anderen Ansätzen, wie etwa dem Hintikkas, das Problem *einfacher* lösen läßt.

(2) Es kann gezeigt werden, daß Hintikkas Bestätigungsmethoden die Adäquatheitsbedingung (III) bzw. (III') verletzen. Hierzu könnte man entgegnen, daß solche Funktionen durchaus für gewisse Situationen adäquat sein mögen. Dazu ist jedoch zu bemerken, daß Hintikka derartige Einschränkungen nicht macht, sondern die al gemeine Anwendbarkeit seiner Methoden voraussetzt. Er müßte also *Gründe* für die Ersetzung von (III) bzw. (III') durch andere Bedingungen angeben, und dann könnten (bei einer genügend ausdrucksreichen Objektsprache) Hintikkas Methoden durch L-Methoden, deren Prämissen zusätzlich noch diese Gründe enthalten, definiert werden.

(3) Die eineindeutige Entsprechung zwischen Uniformitätsgrad des Gegenstandsbereichs und der fairen induktiven Methode gilt nicht in entsprechender Weise für Hintikkas System von Bestätigungsmethoden. Dem könnte Hintikka entgegenhalten, daß stattdessen eine eineindeutige Beziehung zwischen einer anderen Art von Homogenität und seinen Bestätigungsmethoden besteht; diese andere Art wird durch die Konstituenten ausgedrückt: Je mehr Existenzsätze eines Konstituenten negiert sind, desto weniger Eigenschaften sind exemplifiziert und desto mehr allgemeine Gesetze können formuliert werden. Darüber hinaus kann gezeigt werden, daß die Werte der induktiven Methoden Hintikkas für singuläre Prämissen mit wachsendem Erfahrungsdatum Φ_k immer mehr zu den Werten der ihnen entsprechenden L-Methoden Carnaps konvergieren; ist die Anzahl k des Beobachtungsdatums Φ_k also hinreichend groß, so gelten diese Beziehungen approximativ auch zwischen seinen Methoden und den verschiedenen Uniformitätsgraden.

(4) Die Werte der singulären Voraussageschlüsse und die nach Theorem 23 darauf zurückführbaren Schätzungen von relativen Häufigkeiten werden durch Hintikkas Bestätigungsmethoden unnötig verzerrt. Da Hintikkas Methoden bei hinreichend großem Erfahrungsdatum Φ_k allerdings L-Methoden approximieren, fallen, wenn k genügend groß ist, derartige Verzerrungen bei Berechnungen nicht mehr ins Gewicht.

(5) Hintikkas Bestätigungsmethoden sind gleichzeitig Repräsentanten einer enumerativen und einer eliminativen Induktion, da sie nicht nur Allbehauptungen auf Grund von endlichen Erfahrungsdaten stets (auch hinsichtlich unendlicher Gegenstandsbereiche) positive Wahrscheinlichkeitswerte zuordnen, sondern hier auch differenzieren: Die einfachsten, mit dem Datum verträglichen Hypothesen erhalten den höchsten Wert. Wenn Φ_k das Erfahrungsdatum $\alpha_1 \varepsilon \Pi^*_{i_1} \wedge \ldots \wedge \alpha_k \varepsilon \Pi^*_{i_k}$ ist, so erhält hinsichtlich Φ_k jener Konstituent den höchsten Wert, bei dem genau die V-Prädikate $\Pi^*_{i_1}, \ldots, \Pi^*_{i_k}$ exemplifiziert sind, d. h. bei dem genau die Existenzsätze $V\,\xi\,(\xi \varepsilon \Pi^*_{i_1}), \ldots, V\,\xi\,(\xi \varepsilon \Pi^*_{i_k})$ unnegiert vorkommen.[41] Dies hat zur Folge, daß auch sonst generell die *einfachsten* Hypothesen den *größten* Bestätigungswert erhalten, daß also etwa hinsichtlich unseres Erfahrungsdatums der Satz (a) „Zucker löst sich in Wasser auf" einen höheren Bestätigungsgrad erhält als (a') „Zucker löst sich in Wasser auf oder färbt es grün". Hintikka sieht darin einen entscheidenden Vorteil seines Systems gegenüber der Gesamtheit der L-Methoden. — *Ich* sehe darin — falls nur *kontingente* Gesetzesaussagen in Betracht gezogen werden — eine entscheidende Schwäche seines Ansatzes. Nach meinen Intuitionen ist (a') im Hinblick auf unsere Erfahrungen wahrscheinlicher [er scheint mir eher wahr zu sein] als (a), da er schwerer zu widerlegen ist und der Tautologie nähersteht. Da jedoch die Intuitionen bei verschiedenen Menschen verschieden sein können, wird dieses Argument vielleicht nicht jeden überzeugen.

(6) Die einfachste mit einem Erfahrungsdatum verträgliche generelle Hypothese erhält durch Hintikkas induktive Methoden auf Grund dieses Datums deshalb einen größeren Bestätigungsgrad als die anderen generellen Hypothesen, die mit dieser Beobachtungsaussage verträglich sind, weil bereits jener Konstituent, der genau jene Existenzsätze unnegiert enthält, deren Prädikate durch das Datum exemplifiziert sind, auf Grund dieses Datums den höchsten Bestätigungsgrad erhält. Daß dies unseren intuitiven Vorstellungen vom Bestätigungsbegriff entspricht, versucht Hintikka folgendermaßen zu rechtfertigen:[42] Das Erfahrungsdatum sei hinreichend groß gewählt; es bestehe etwa aus 10000 einzelnen Beobachtungen. In ihm seien drei V-Prädi-

[41] Vgl. Hintikka [1], S. 285ff.
[42] Vgl. Hintikka [1], S. 276f.

kate exemplifiziert. Dann hatten wir genügend Gelegenheit, hinsichtlich der bislang nichtexemplifizierten Prädikate nach Beispielen zu suchen; daß wir keine gefunden haben, liegt vermutlich daran, daß es keine gibt. Also hat eine induktive Methode jenen Konstituenten, der genau die drei entsprechenden Existenzsätze unnegiert enthält, den höchsten Bestätigungsgrad zuzuordnen. — Dieses Argument läßt sich folgendermaßen paralysieren: Unter den 10000 Beobachtungen seien 9800 Exemplifizierungen von Π^*_1, 198 von Π^*_2 und 2 von Π^*_3; dabei haben a_{3227} sowie a_{7438} die durch Π^*_3 bezeichnete Eigenschaft F^*_3. Alle anderen V-Prädikate sind dann unexemplifiziert. Der Bestätigungsgrad, den der entsprechende Konstituent auf Grund dieses Datums hat, sei sehr groß. Nun erfüllen Hintikkas Methoden, wie die Carnaps, die Bedingung (II) (a). Also haben diese Konstituenten auf Grund eines anderen Erfahrungsdatums, das die gleichen relativen Häufigkeiten aufweist wie jenes, das aber besagt, daß Π^*_3 erst durch die Gegenstände a_{9999} und a_{10000} exemplifiziert worden sind, den gleichen Bestätigungsgrad. Wenn Π^*_3 aber erst *durch die letzten Beobachtungen* und sozusagen *im letzten Augenblick* exemplifiziert worden ist, so liegt die Vermutung nahe, daß auch Π^*_4 noch exemplifiziert wird, wenn man nur lange genug weitersucht. — Dem könnte man entgegenhalten, daß dann eben Φ_k genügend groß gewählt werden muß, etwa als Φ_{100000}, um diesem Gegenargument zu entgehen. Ein solcher Einwand würde jedoch verkennen, was ich zeigen wollte: Man kann das obige Gegenargument ja auch auf dieses respektable Datum Φ_{100000} anwenden, ohne daß sich daran substantiell etwas ändert. Durch eine geringfügige Ergänzung von Hintikkas Argument, daß seine Methoden bei der Anwendung adäquater sind als die Carnaps, erhält man also einen Nachweis dafür, daß sie weniger adäquat sind als diese.

(7) Induktive Methoden sollen nicht nur von großer formaler Eleganz sein, sondern uns insbesondere auch bei der Anwendung als Hilfsmittel dienen. Hierbei stellt sich heraus, daß Hintikkas Methoden, insoweit sie von den L-Methoden abweichen, uns induktiv inkonsistente Ergebnisse liefern. Gegeben sei ein Erfahrungsdatum Φ_{10000}, das besagt, daß alle beobachteten Einzelfälle die V-Eigenschaft F^*_1 haben. Der Gegenstandsbereich umfasse 1010000 Dinge. Der Wert $c\,(\alpha_{10001} \in \Pi^*_1, \Phi_{10000})$ sei dann bei Hintikkas wie bei Carnaps

Methoden annähernd gleich und sehr groß, etwa 0,999. Der Wert $c(\Lambda\xi(\xi \varepsilon \Pi^{\star}_1), \Phi_{10000})$ hingegen liegt bei Verwendung einer L-Methode zwar nahe bei 0, während er bei Benützung einer Funktion Hintikkas nahe bei 1 liegen wird und, für eine sehr kleine Zahl s, etwa 1 — s sein möge.[43] Angenommen, ich möchte auf den Satz $\Lambda\xi(\xi \varepsilon \Pi^{\star}_1)$ oder, was wegen der Prämisse Φ_{10000} und dem fest vorgegebenen Gegenstandsbereich damit deduktiv gleichwertig ist, auf die Behauptung $\alpha_{10001} \varepsilon \Pi^{\star}_1 \wedge \ldots \wedge \alpha_{1010000} \varepsilon \Pi^{\star}_1$ eine Wette eingehen; dann sagt mir Hintikkas Methode im einen Fall, daß der nahe bei 1 liegende Wert 1 — s ein fairer Wettquotient ist und daß ich also mit praktischer Sicherheit damit rechnen kann, daß es kein Gegenbeispiel gibt, während sie mir andererseits versichert, daß die relative Häufigkeit von $F^{\star}_1$ in $\{a_{10001}, \ldots, a_{1010000}\}$ 0,999 ist, daß also in diesem Bereich voraussichtlich 999000 Dinge die Eigenschaft $F^{\star}_1$ haben werden und daß ich demnach mit 1000 Gegenbeispielen rechnen muß.[44] Wenn wir mit 1000 Gegenbeispielen rechnen müssen, können wir aber nicht gleichzeitig den Allsatz als praktisch sicher ansehen. Die Methoden Carnaps zeigen diese induktive Inkonsistenz nicht, da sie dem Allsatz in diesem Fall einen äußerst geringen c-Wert zuordnen. Dieses Resultat ist in meinen Augen ein entscheidender Einwand gegen die *Benützung*[45] von Hintikkas Funktionen in solchen Anwendungsfällen.

(8) Hintikkas wie Carnaps Funktionen erfüllen die Bedingung (II) (a), der in der mathematischen Wahrscheinlichkeitslehre die Bedingung der unabhängigen Wiederholbarkeit eines Experiments entspricht. Hiergegen kann folgender Einwand erhoben werden: Im strengen Sinn gibt es in unserer Welt vielleicht überhaupt keine unabhängige Wiederholung eines Experiments. Es kann sein, daß sich unsere sogenannten Naturkonstanten (unmerklich langsam, aber doch) kon-

[43] s wird im allgemeinen nicht identisch mit 0,001 sein, doch spielt dies für die folgenden Überlegungen keine Rolle.

[44] Hintikkas Funktionen sind reguläre Bestätigungsmethoden und symmetrisch bezüglich der Gegenstandsausdrücke; für sie gilt also Theorem 23 (a) (β) bzw. (b) (α).

[45] Selbstverständlich kann dies kein Einwand dagegen sein, induktive Verfahren in Hintikkas Weise zu *charakterisieren*; diese Freiheit des Entwerfens neuer Systeme darf nicht beschnitten werden, da es sich ja stets im Nachhinein herausstellen kann, daß es für derartige zunächst scheinbar unplausible Methoden andere Anwendungsgebiete gibt, in denen sie äußerst fruchtbar sind.

tinuierlich ändern, und zwar in einer Weise, die nicht vorhersagbar ist; analog deformiert sich im Laufe der Zeit ja jeder Würfel, wenn man ihn lange genug benützt, so daß er schließlich nicht mehr ebenmäßig ist, und es kann der Fall eintreten, daß sich bestimmte Werte, wie etwa die durchschnittliche Monatstemperatur im Januar, periodisch ändern, ohne daß diese Perioden unter ein numerisches Gesetz subsumiert werden können. Das Experiment ist dann auf kurze Sicht (fast) unabhängig wiederholbar, und die Bedingung (II) (a) ist dann für eine Sprache, die nur für einen derartig begrenzten Objektbereich Gegenstandsausdrücke hat, (fast) adäquat. Wenn wir auf die mittlere Januartemperatur des nächsten Jahres induktiv schließen wollen, so erhalten wir zwar annähernd den gleichen Wert, wenn wir die Januartemperaturen der letzten 50 Jahre und wenn wir die der letzten 500 Jahre als Erfahrungsdatum benützen; wir erhalten jedoch einen anderen, wenn wir die Werte der letzten 50000 Jahre zugrundelegen, und einen dritten, wenn wir noch weiter zurückgehen. Da wir für die nächsten 12 Monate nicht mit dem Beginn einer neuen Eiszeit rechnen, werden wir Werte der letzteren Art für inadäquat ansehen. Wenn wir die Gesetzmäßigkeit wüßten, nach der sich die Temperaturschwankungen abspielen, könnten wir diese zu den Prämissen hinzunehmen und hätten damit keinen Grund, die Bedingung (II) (a) fallenzulassen. Da wir sie jedoch gerade nicht kennen, und wohl auch kaum jemals kennen werden, bleibt uns also nur die Möglichkeit offen, entweder auf (II) (a) zu verzichten oder aber jeweils den Bereich anzugeben, für den (II) (a) bzw. die unabhängige Wiederholbarkeit des Experiments (approximativ) gilt. Da es zur Bestimmung des Umfangs solcher Teilbereiche keine Regel geben kann (sonst hätten wir ja eine derartige Gesetzmäßigkeit), sind wir hier auf unser Geschick angewiesen, mit dem wir zwar häufig Erfolge erzielen, das uns jedoch bei der Frage nach den *Grundlagen* des induktiven Schließens nicht weiterhilft: Es soll bei derartigen Untersuchungen ja gerade dieses Geschick (wenn es korrekt angewandt wird) *gerechtfertigt* werden; die Rechtfertigung kann dann nicht in dem Hinweis auf eben dieses Geschick bestehen. — Diese Bemerkungen stellen natürlich keinen Einwand gegen Hintikkas System dar, da sie nicht für dieses spezifisch sind und, genauso wie etwa das von Goodman entdeckte Problem der zerrütteten Prädikate, alle bekannten Systeme

des induktiven Schließens betreffen. Es ist jedoch ein Einwurf, von dem man sich zunächst bei Hintikka eher eine Lösung erwarten wird als beim Ansatz Carnaps, ohne daß Hintikkas Methoden hierzu tatsächlich irgendetwas beitragen können.

Wenn man eine vage und schwer durchschaubare Theorie von einem ihr fremden Standpunkt aus — nämlich von dem der Logik aus — analysiert und kritisiert, so ist gegen dieses Verfahren nichts einzuwenden. Wenn man hingegen ein präzises und unzweideutig formuliertes System, das erst entwickelt wird, wie Hintikkas induktive Logik, vom Standpunkt eines anderen Systems der induktiven Logik aus mit sehr allgemeinen Argumenten kritisiert, wie ich es hier getan habe, so ist dies unfair. Außerdem gerät man in die Gefahr, daß die Zeit (die weitere Entwicklung der kritisierten Theorie) diese Einwände von selbst erledigt. So wie den Mathematikern (von Frege abgesehen) die Kardinalzahlen und Ordinalzahlen Cantors zunächst absurd erschienen, obwohl diese natürliche Erweiterungen des Systems der natürlichen Zahlen darstellen, so kann es sich auch erweisen, daß Hintikkas Methoden naheliegende Verallgemeinerungen der L-Methoden sind, die unter bestimmten, nicht allgemein gegebenen Voraussetzungen die Werte dieser Funktionen (mit genügender Annäherung) liefern.

Zu klären bleibt noch die Frage, unter welchen Gesichtspunkten wir Theorien bewerten, die wir nicht als kontingent ansehen, die vielmehr die Intension der von uns benützten Grundbegriffe beschreiben. Diese Bewertung muß (nach dem Ergebnis des zweiten Abschnittes dieses Kapitels) vorgenommen werden, *bevor* eine induktive Methode, die von c^∞ verschieden ist, adäquat angewandt werden kann. Insbesondere muß gezeigt werden, warum wir ein Axiomensystem mit Grundbegriffen, unter denen auch „F" (etwa „rot") vorkommt, einem damit logisch gleichwertigen System vorziehen, das bis auf „F" die gleichen Grundbegriffe, aber zusätzlich noch „G" und „a_1" enthält. Die Antwort, die jedermann darauf geben wird, ist, daß jenes System einfacher ist als dieses.[46] Ich sehe sie durchaus als befriedigend

---

[46] Es ist ja auch Keplers Theorie der Planetenbewegung mit einem hinreichend modifizierten ptolemäischen System logisch äquivalent und wird daher nicht wegen ihrer *größeren Wahrscheinlichkeit*, sondern wegen ihrer *größeren Einfachheit* akzeptiert. Gleiches gilt im Hinblick auf Einsteins Relativitätstheorie einerseits und die klassische Mechanik und Elektrizitätslehre andererseits.

an; unbefriedigend finde ich lediglich die bisherigen Untersuchungen zum Einfachheitsbegriff und die bisherigen Analysen desselben. Zum Teil sind sie zu vage und unverbindlich wie die an sich interessanten Ergebnisse Kemenys[47] oder die anregenden Aphorismen Sir Karls[48], und zum Teil ist noch nicht zu übersehen, welche Relevanz bestimmte Untersuchungen allgemein oder ganz konkret zu diesem Problem haben.[49] Es ist nicht ausgeschlossen, daß sich Hintikkas Bestimmung der Einfachheit einer Aussage[50] als fruchtbarster Ansatz erweist.

Mit dem Aufweis dessen, wieviele Fragen der induktiven Logik als gelöst gelten können, zeigt sich also, wieviele bisher noch nicht oder noch unzureichend gesehene Probleme einer eingehenderen Analyse bedürfen. Die Lösung solcher Probleme besteht gerade nicht in allgemeinen und unverbindlichen Behauptungen von der Art „Ein Kriterium der Wahrheit von Naturgesetzen ist ihre Einfachheit", die jeder kennt und die niemandem nützen, die sich später aber fast immer als nur bedingt richtig, d. h. als falsch erweisen. Klären kann man solche Schwierigkeiten nur durch detaillierte und konkrete Kleinarbeit, auch wenn die sogenannten großen Geister darüber anders denken. In diesem Sinne enthält die vorliegende Schrift mehr *Problemstellungen* als *Problemlösungen*, was aber vielleicht einer ernsthaften philosophischen Diskussion nicht schadet.

---

47 Vgl. Kemeny [4].
48 Man findet sie zum größten Teil in Popper [2].
49 Vgl. z. B. Hans Hermes, Zum Einfachheitsprinzip in der Wahrscheinlichkeitsrechnung, in: Dialectica 12 (1958) S. 113—127.
50 Vgl. Hintikka [1], S. 285.

Anhang

Der Erfolg mit adäquaten induktiven Methoden

Zur Abrundung dieser Arbeit seien hier Carnaps Untersuchungen geschildert über das Verhältnis der Wahl einer adäquaten induktiven Methode zum Erfolg, den man unter ihrer Verwendung bei einem bestimmten Zustand der Welt hat.[1] Es sollen dabei die verschiedenen Eigenschaften der einzelnen adäquaten induktiven Methoden hinsichtlich eines vorgegebenen Universums, das durch eine Zustandsbeschreibung charakterisiert wird, miteinander verglichen werden; die Ergebnisse dieses Vergleichs sind für jeden, der mit einer induktiven Methode eine rationale Grundlage seines Handelns erhalten will, von großer Bedeutung.

Zunächst werden einige Begriffe aus der Statistik charakterisiert. Es sei f eine Variable (eine Zufallsvariable, d. h. eine Funktion wie etwa *die Länge von* oder *die Anzahl von*). Vorgegeben seien m mögliche Fälle, für die f die Werte $f(a_1), \ldots, f(a_m)$ habe, die nicht notwendigerweise voneinander verschieden sein müssen. Dann möge gelten:

$$(1)\quad M(f) = \sum_{i=1}^{m} \frac{f(a_i)}{m} = \text{das Mittel von f}$$

$$(2)\quad Q(f,a) = \sum_{i=1}^{m} \frac{(f(a_i) - f(a))^2}{m}$$

$$= \text{die mittlere quadratische Abweichung von f bezüglich eines fest vorgegebenen Wertes } f(a)$$

---

[1] Vgl. Carnap [11], S. 56—80. Da dieses Buch nicht mehr neu aufgelegt worden ist und da es in manchen Bibliotheken fehlt, ist eine solche Skizze ohne Zweifel notwendig. Im folgenden wird praktisch immer ein endlicher Gegenstandsbereich und stets eine endliche Gesamtheit von Grundprädikaten vorausgesetzt; die Beweise der einzelnen Lehrsätze können deshalb, da dann keine Integrationstheorie vorausgesetzt werden muß, im Prinzip auch vom philosophischen Leser selbst gefunden werden und werden daher hier unterdrückt.

(3) $V(f) = \sum_{i=1}^{m} \frac{(f(a_i) - M(f))^2}{m}$

= die Varianz von f = die mittlere quadratische Abweichung von f bezüglich M (f).[2]

Im folgenden wird stets vorausgesetzt, daß die Bedingungen dieser Definitionen erfüllt sind. Die Varianz ist dann der kleinste Wert der mittleren quadratischen Abweichung, da für jeden von M (f) verschiedenen Wert f (a) gilt:

(4) $Q(f, a) = V(f) + (M(f) - f(a))^2$

Für $f(a) = 0$ gilt folglich:

(5) $M(f^2) = V(f) + (M(f))^2$

u und w seien irgendwelche konstante Werte; dann gilt:

(6) $M(u \cdot f) = u \cdot M(f)$

(7) $M(f + w) = M(f) + w$

(8) $M(u \cdot f + w) = u \cdot M(f) + w$

(9) $V(f + d) = V(f)$

(10) $V(u \cdot f) = V(u \cdot f + d) = u^2 \cdot V(f)$

Gegeben sei eine endliche Sprache, die n Gegenstandsausdrücke, p einstellige Grundprädikate und somit $q = 2^p$ V-Prädikate hat. Σ sei eine Zustandsbeschreibung und Φ_k (mit $k > 0$) eine individuelle Verteilung in den Ausdrücken $\alpha_1, \ldots, \alpha_k$, wobei Σ dem Prädikat Π (das die Eigenschaft F beschreibt) k′ Elemente aus dieser beobachteten Gesamtheit zuweist; Π muß nicht notwendigerweise ein Grundprädikat sein. Es sei A der nicht in Φ_k erwähnte Teil des Gegenstandsbereichs B, und es sei $n \gg k$, so daß A approximativ mit B und also n approximativ mit $n - k$ gleichgesetzt werden kann. Die relative Häufigkeit von F im Universum B sei $r = n'/n$, wobei n′ die Anzahl der Elemente von B ist, die nach Σ die Eigenschaft F besitzen.

Die Gesamtheit der Stichproben vom Umfang k des vorgegebenen Universums (das n Elemente enthält) ist die Menge aller Teilklassen des Universums, die genau k Objekte enthalten. Es gilt dann offenbar:

(11) Die Anzahl aller Stichproben der Größe k im Universum der Größe n ist $\binom{n}{k}$.

---

[2] Statt „V" wird in der Literatur auch oft „σ^2" geschrieben.

Eine jede Schätzungsmethode kann auf jede dieser Stichproben angewandt werden; wir betrachten daher die Verteilung der Werte unter diesen Stichproben, d. h. die *Stichprobenverteilung* dieser Werte. Es ist $0 \leqslant k' \leqslant k$. Dann gilt:

(12) Die Anzahl der Stichproben der Größe k im Universum der Größe n mit den vorgegebenen Werten k' und n' ist $\binom{n'}{k'} \cdot \binom{n-n'}{k-k'}$.

Folglich ist die relative Häufigkeit dieser Stichproben mit dem vorgegebenen Wert k' unter allen Stichproben der Größe k der Quotient aus den Werten von (12) und (11). Für $n \gg k$ gilt nun:[3]

(13) Die relative Häufigkeit der Stichproben der Größe k mit einem vorgegebenen Wert k' unter allen Stichproben der Größe k ist $\binom{k}{k'} \cdot r^{k'} \cdot (1-r)^{k-k'}$

Für das Mittel und die Varianz von k' bei fest vorgegebenem k gilt dann:[4]

(14) $M(k') = k \cdot r$

(15) $V(k') = k \cdot r \cdot (1-r)$

Mit (6) und (10) erhält man daraus:

(16) $M(k'/k) = r$

(17) $V(k'/k) = r \cdot (1-r)/k$

Der Irrtum Ir (s, r, Φ) eines Schätzwertes s auf der Grundlage von Φ hinsichtlich des tatsächlichen Wertes r ist die Differenz zwischen der (auf Grund von Φ vorgenommenen) Schätzung s und der Größe r selbst:

(18) $\mathrm{Ir}(s, r, \Phi) = s - r$
= der Irrtum einer Schätzung s auf der Grundlage Φ bezüglich eines Wertes r.

Mit (7) gilt dann für den mittleren Irrtum (oft auch „Verzerrung der Schätzungsmethode" genannt):

(19) $M(\mathrm{Ir}(s, r, \Phi)) = M(s) - r$

Die Varianz des Irrtums ist nach (9) gleich der Varianz der Schätzung:

(20) $V(\mathrm{Ir}(s, r, \Phi)) = V(s)$

---

[3] Vgl. Korollar 12 (b); für relative Häufigkeiten dieser Art ergeben sich demnach die gleichen Werte wie für direkte Ableitungen.

[4] Vgl. Cramér [1], S. 193.

Der mittlere quadratische Irrtum Mq (s, r) ist das Mittel des quadratischen Irrtums der Schätzung in der Stichprobenverteilung:

(21) $Mq(s, r) = M(Ir(s, r, \Phi)^2)$
$= $ der mittlere quadratische Irrtum bei der Schätzung s hinsichtlich des Wertes r.

Mit (5) bzw. (20) erhält man daraus:

(22) $Mq(s, r) = V(Ir(s, r, \Phi)) + M(Ir(s, r, \Phi)^2)$

(23) $Mq(s, r) = V(s) + M(Ir(s, r, \Phi)^2)$

Dieses mathematische Rüstzeug wird nun benützt, um festzustellen, welche induktive Methode hinsichtlich der Welt, die durch Σ beschrieben wird, optimal ist, also z. B. bei Wetten den geringsten Verlust bzw. den größten Gewinn verspricht. Die folgenden Überlegungen haben allerdings ein bloß theoretisches Interesse, da uns bei einer faktischen Anwendung der Zustand der Welt ja gerade nicht bekannt ist, sondern erst auf induktivem Wege erschlossen werden soll. Doch zeigen uns diese Überlegungen, wie gut und wie schnell wir diesen faktischen Zustand mit den einzelnen Methoden erschließen können, was für unsere Wahl hinsichtlich der jeweils optimalen L-Methode selbstverständlich nicht ohne Einfluß sein kann.

Das Mittel M (f) der Zufallsvariablen (der Funktion) f ist dabei das Zentrum der Streuwerte $f(a_1), \ldots, f(a_m)$, die mittlere quadratische Abweichung von f bezüglich des Wertes f (a) ist ein Maß für die Streuung der Werte um f (a), und die Varianz von f ist ein Maß der Streuung der Werte $f(a_i)$ um ihr Mittel M (f). Im folgenden wird als f die durch eine Funktion c^L bewerkstelligte Schätzung der relativen Häufigkeit genommen, also die Funktion srh_{c^L}. Eine derartige Schätzungsfunktion wird als schlecht betrachtet, wenn ihre Werte, die sich auf Stichprobenverteilungen der Größe k beziehen, weit gestreut sind; also soll die mittlere quadratische Abweichung der Werte der Stichprobenverteilung nicht zu groß sein. Als Bezugswert kann jedoch nicht das Mittel $M(srh_{c^L})$ genommen werden, da man sonst die Varianz $V(srh_{c^L})$ erhält, die lediglich die Weite der gestreuten Werte *untereinander* mißt, nicht jedoch die zum *tatsächlichen Wert* r. Dieser tatsächliche Wert r wird durch die Zustandsbeschreibung Σ bestimmt. Die Funktion srh_{c^L} ist demnach als um so brauchbarer anzusehen, je kleiner die mittlere quadratische Abweichung der Werte von srh_{c^L} bezüglich des tatsächlichen Wertes r ist, d. h. je klei-

ner der mittlere quadratische Irrtum in der Stichprobenverteilung bezüglich Σ ist.
Π habe die Weite m und damit die relative Weite m/q. Durch Theorem 23 (a) (β) erhält man dann aus Korollar 7 für adäquate induktive Methoden c^L:

$$(24)\quad c(\alpha_{k+1} \varepsilon \Pi, \Phi_k) = \mathrm{srh}_{c^L}(F, A, \Phi_k) = \frac{k' + (m/q) \cdot L}{k + L}.$$

In der gegebenen Zustandsbeschreibung Σ wird beschrieben, daß r die relative Häufigkeit von F in B ist. Das Mittel und die Varianz von srh_{c^L} (mit festem L) für alle Stichproben aus Σ mit fest vorgegebener Größe k erhält man aus (24), (14) und (15) mittels (8) und (10):

$$(25)\quad M(\mathrm{srh}_{c^L}) = \frac{k \cdot r + (m/q) \cdot L}{k + L}$$

$$(26)\quad V(\mathrm{srh}_{c^L}) = \frac{k \cdot r \cdot (1 - r)}{(k + L)^2}.$$

Eine Methode srh_{c^L} ist genau dann unverzerrt, wenn ihr Mittel stets mit r identisch ist, wenn also stets $M(\mathrm{srh}_{c^L}) = r$. Aus (25) folgt, daß dies nur für $L = 0$ gilt. Die Proportionalregel c^0 ist demnach die einzige unverzerrte Methode des L-Systems (wenn man die beiden Grenzmethoden noch hinzurechnet).
Nach Theorem 23 (b) (α) gilt für jede L-Methode c^L (übrigens einschließlich c^0 und c^∞), daß $c^L(\alpha_{k+1} \varepsilon \Pi, \Phi_k) = \mathrm{srh}_{c^L}(F, B, \Phi_k)$. Angenommen, die objektive Wahrscheinlichkeit von F in B existiert, d. h. angenommen, die relative Häufigkeit n'/n von F in B_n strebt mit wachsendem n einem Grenzwert zu, dann strebt demnach auch $\mathrm{srh}_{c^L}(F, B, \Phi_k)$ (falls $L \neq \infty$) mit wachsendem k ebenfalls diesem Limes zu. Also sind alle Funktionen c^L und die darauf basierenden Schätzungsfunktionen mit Ausnahme von c^∞ und srh_{c^∞} im Sinne Reichenbachs *selbstkorrigierend*.[5] Vom Standpunkt der objektiven Wahrscheinlichkeit aus sind also (abgesehen von c^∞) alle Schätzungsmethoden srh_{c^L} und damit alle Methoden c^L gleichwertig.
B sei wiederum endlich, aber hinreichend groß, so daß wir seine An-

[5] Vgl. Reichenbach [10], S. 445ff.

zahl n, ohne einen zu großen Fehler zu machen, mit n — k (und damit B mit A) identifizieren können. Es sei $s = \mathrm{srh}_{c^L}(F, A, \Phi_k)$. Hinsichtlich des Irrtums $\mathrm{Ir}(s, r, \Phi_k)$ der Schätzung s bezüglich Φ_k hinsichtlich der relativen Häufigkeit r, seines Mittels $M(\mathrm{Ir}(s, r, \Phi_k))$ und seiner Varianz $V(\mathrm{Ir}(s, r, \Phi_k))$ in der Stichprobenverteilung der Größe k hinsichtlich der Zustandsbeschreibung Σ gilt mit (24), (25) und (26) wegen (18), (19) und (20):

$$(27)\quad \mathrm{Ir}(s, r, \Phi_k) = \frac{k' + (m/q) \cdot L}{k + L} - r$$

$$(28)\quad M(\mathrm{Ir}(s, r, \Phi_k)) = \frac{(m/q - r) \cdot L}{k + L}$$

$$(29)\quad V(\mathrm{Ir}(s, r, \Phi_k)) = \frac{k \cdot r \cdot (1 - r)}{(k + L)^2}.$$

Für den mittleren quadratischen Irrtum $\mathrm{Mq}(s, r)$ in der Stichprobenverteilung nach Σ erhält man aus (29) und (28) unter Zuhilfenahme von (22):

$$(30)\quad \mathrm{Mq}(s, r) = \frac{k \cdot r \cdot (1 - r) + (m/q - r)^2 \cdot L^2}{(k + L)^2}$$

$$= \frac{(L^2 - k) \cdot r^2 + (k - 2 \cdot L^2 \cdot m/q) \cdot r + L^2 \cdot (m/q)^2}{(k + L)^2}.$$

$\mathrm{Mq}(s, r)$ ist demnach für $L^2 < k$ eine nach oben geöffnete und für $L^2 > k$ eine nach unten geöffnete Parabel und für $L^2 = k$ eine Gerade; für $L^2 = k$ ist der Wert von $\mathrm{Mq}(s, r)$, wie man leicht sehen kann, unabhängig von r. Dabei ist s nach Voraussetzung der durch c^L auf Grund von Φ_k geschätzte Wert der relativen Häufigkeit von F im noch nicht beobachteten Universum A. Nun ist F eine beliebige Eigenschaft, die in der Sprache ausdrückbar ist, und entsprechend ist Π ein beliebiger, in den Grundprädikaten definierbarer Ausdruck. Aus (30) folgt daher für ein beliebiges V-Prädikat Π^*_i mit der relativen Weite 1/q, das eine Eigenschaft F^*_i bezeichnet, womit nach Σ r_i die relative Häufigkeit von F_i in B ist:

$$(31)\quad \mathrm{Mq}(s, r_i) = \frac{k \cdot r_i \cdot (1 - r_i) + 1/q - r_i)^2 \cdot L^2}{(k + L)^2}$$

$$= \frac{(L^2 - k) \cdot r_i^2 + (k - 2 \cdot L^2/q) \cdot r_i + L^2/q^2}{(k + L)^2}.$$

u sei die Summe der Quadrate der relativen Häufigkeiten der V-Eigenschaften in B, so wie dies durch Σ beschrieben wird; es möge also per definitionem gelten:[6]

(32) $u = \sum_{i=1}^{q} r_i^2$

Als Hilfsbegriffe seien definiert:

(33) $u_1 = 1 - u$

(34) $u_2 = u - (1/q)$

Es wird nun Mq (s, u) nach (11) als der mittlere quadratische Irrtum der $q \cdot \binom{n}{k}$ Schätzungen s mit $s = srh_{c^k}(F_i, A, \Phi_k)$ eingeführt, wobei für jede der $\binom{n}{k}$ Stichprobenverteilungen der Größe k nach Σ und für jedes der q V-Prädikate die relative Häufigkeit geschätzt wird. Mq (s, u) ist deshalb das Mittel der q Werte Mq (s, r_i), wobei $1 \leqslant i \leqslant q$:

(35) $Mq(s, u) = \frac{1}{q} \sum_{i=1}^{q} Mq(s, r_i)$

Da $\sum_{i=1}^{q} r_i = 1$, folgt mit (31) und (32):

(36) $Mq(s, u) = \frac{k - L^2/q + (L^2 - k) \cdot u}{q \cdot (k + L)^2}$

Man hätte das Ergebnis (36) auch erhalten, wenn man nicht auf die Zustandsbeschreibung Σ, sondern auf die Strukturbeschreibung (auf die statistische Verteilung) Δ, die Σ als Adjunktionsglied enthält, Bezug genommen hätte; denn es sind nur die Anzahlen n und q, die die Sprache charakterisieren, und die Anzahlen $n_1, \ldots, n_q$ der q V-Eigenschaften nach Σ benützt worden, um (36) zu erhalten. Aus dieser Gleichung folgt nun, daß auch n und die n_i (abgesehen von der Voraussetzung, daß $n \gg k$) nicht relevant sind und daß von ihnen nur vorausgesetzt werden muß, daß nach (32): $(n_1/n)^2 + \ldots + (n_q/n)^2 = u$. Diese Zahl u ist demnach die einzige dieser die Zustandsbeschreibung betreffenden Größen, die für Mq (s, u) relevant ist.

u hat das Maximum 1, wenn eine der relativen Häufigkeiten 1 und die anderen 0 sind. Der Gegenstandsbereich ist dann hinsichtlich der V-Eigenschaften durchgehend uniform (und die Zustandsbeschrei-

[6] Vgl. hierzu auch die Definition 27 im V. Kapitel.

bungen, die ihn charakterisieren, entsprechend durchgehend homogen). u hat ferner das Minimum $1/q$; denn es gilt dann, daß die einzelnen relativen Häufigkeiten miteinander gleich sind, so daß $n_i = n/q$ und somit $r_i = n_i/n = 1/q$, sowie $r_i^2 = 1/q^2$ und schließlich $r_1^2 + \ldots + r_q^2 = q \cdot (1/q^2) = 1/q$. Dieser minimale Wert u kann in einem endlichen Gegenstandsbereich B nur erreicht werden, wenn n durch q teilbar ist (er kann aber ansonsten bei hinreichend großem n beliebig angenähert werden). Die Uniformität von B ist also am geringsten, wenn alle V-Eigenschaften von gleicher Anzahl sind. Da nach Voraussetzung $n \gg k$, sind somit auch im noch nicht beobachteten Teil A von B die V-Eigenschaften von (fast) gleicher Anzahl, und es wäre in diesem Fall unklug, dem durch Φ_k beschriebenen Erfahrungsdatum irgendein Vertrauen zu schenken. Während also bei $u = 1$ die Funktion c^0 fair erscheint, ist es für $u = 1/q$ offenbar die Methode c^∞. Gilt für den Uniformitätsgrad[7] u hingegen, daß $1/q < u < 1$, so ist entsprechend auch eine Methode c^L mit $0 < L < \infty$ fair.

Aus (36) folgt mit (33) und (34):

$$(37)\quad \mathrm{Mq}(s, u) = \frac{u_1 \cdot k + u_2 \cdot L^2}{q \cdot (k + L)^2}.$$

Für die beiden extremen Werte von u gilt dann:

(38) Ist $u = 1$, so gilt:
- (a) $u_1 = 0$
- (b) $u_2 = (q - 1)/q$
- (c) $\mathrm{Mq}(s, u) = \dfrac{(q - 1) \cdot L^2}{q^2 \cdot (k + L)^2}$.

(39) Ist $u = 1/q$, so gilt:
- (a) $u_1 = (q - 1)/q$
- (b) $u_2 = 0$
- (c) $\mathrm{Mq}(s, u) = \dfrac{(q - 1) \cdot k}{q^2 \cdot (k + L)^2}$.

Die L-Methoden, die durch den Faktor L in einem Kontinuum angeordnet werden, können nun daraufhin untersucht werden, wie erfolg-

[7] Ich verwende diesen Begriff „Uniformitätsgrad" anders, als Carnap den Ausdruck „degree of order" in [11], S. 66, offensichtlich intendiert; diesen halte ich allerdings für nicht sonderlich interessant.

reich sie in der Anwendung bei einem vorgegebenen Zustand der Welt sind (der durch eine Zustandsbeschreibung Σ charakterisiert ist). Es soll also untersucht werden, welche dieser Methoden hinsichtlich einer solchen Welt die erfolgreichste und damit die fairste ist, d. h. bei welcher der mittlere quadratische Irrtum am kleinsten ist. Hierbei sind zwei Fälle zu unterscheiden:

(A) Der Uniformitätsgrad u ist nicht extrem, es ist also $1/q < u < 1$ und damit $u_1 > 0$ und $u_2 > 0$. Die in (37) dargestellte Funktion $Mq\,(s, u)$ ist eine Funktion von L.[8] Wenn man sie partiell nach L differenziert, so sieht man, daß $Mq\,(s, u)$ mit wachsendem L zunächst abnimmt, dann einen minimalen Wert hat und schließlich monoton ansteigt. $L^\star$ sei dieser Wert L, für den $Mq\,(s, u)$ ein Minimum ist. Dann ist:

$$(40)\quad L^\star = \frac{u_1}{u_2} = \frac{1-u}{u-1/q}.$$

Der minimale Wert von $Mq\,(s, u)$, den man bei L erhält, ist dann wegen (37):

$$(41)\quad \min L\,(Mq\,(s, u)) = \min L\left(\frac{u_1 \cdot k + u_2 \cdot L^2}{q \cdot (k + L)^2}\right)$$

$$= \frac{u_1 \cdot k + u_2 \cdot L^{\star 2}}{q \cdot (k + L^\star)^2} = \frac{u_1 \cdot u_2}{q \cdot (u_1 + k \cdot u_2)}.$$

(B) der Uniformitätsgrad u sei einer der Werte 1 oder $1/q$; mit (38) und (39) erhält man dann:

(42) Es sei $u = 1$ und somit $u_1 = 0$; dann gilt:
- (a) $L^\star = 0$
- (b) $Mq\,(s, u) = 0$ bei $s = srh_c{}^0\,(F_i, A, \Phi_k)$

(43) Es sei $u = 1/q$ und somit $u_2 = 0$; dann gilt:
- (a) $L^\star = \infty$
- (b) $Mq\,(s, u) = 0$ bei $s = srh_{c\infty}\,(F_i, A, \Phi_k)$

Bei $u = 1$ hat $Mq\,(s, u)$ das Minimum 0 bei $L = 0$ und steigt von da ab mit wachsendem L an, bei $L = 1/q$ hingegen fällt es mit wach-

[8] Diese Abhängigkeit ist dort nicht explizit festgehalten worden, da an jener Stelle angenommen worden ist, daß L konstant bleibt; hier hingegen soll L ausdrücklich variiert werden, um feststellen zu können, welche Methode c^L bei Anwendungen optimal ist.

sendem L und hat das Minimum 0 bei $L = \infty$ (genaugenommen muß in diesem Fall selbstverständlich der Grenzwert der Funktion gebildet werden, die entsteht, wenn L gegen ∞ wächst). Die Gleichung (40) gilt übrigens, wie man leicht sehen kann, auch für $L^* = 0$ und für $L^* = \infty$ (bzw., wiederum genauer, mit dem Grenzwert von (40), den man erhält, wenn L über alle Grenzen anwächst). (40) gilt also für alle positiven reellen Zahlen einschließlich der beiden Grenzpunkte, und wegen (42) (b) und (43) (b) gilt dies dann auch für (41).

L^* ist der bei Σ optimale L-Wert und die durch L^* bestimmte adäquate induktive Methode demnach die in der durch Σ beschriebenen Welt *optimale* (oder *faire*) *induktive Methode*, da sie den kleinsten mittleren quadratischen Irrtum hat. L^* ist unabhängig von k, obwohl Mq (s, u) noch von k abhängt; L^* bestimmt also für Stichproben von *beliebiger* Größe k (falls nur $k \ll n$) die bei Σ optimale induktive Methode; das vorgegebene Erfahrungsdatum Φ_k ist davon nur ein Beispiel.

Nach (40) gibt es zu jedem Uniformitätsgrad u (und zu jeder Zustandsbeschreibung Σ, deren Uniformitätsgrad u ist) genau einen Wert L^*, so daß die dadurch bestimmte adäquate induktive Methode in einer solchen Welt fair ist. Das ist zwar von großer theoretischer Wichtigkeit, aber ohne unmittelbaren praktischen Nutzen. Denn wir Menschen kennen zu jeder Zeit nur einen Bruchteil der Welt, und eine induktive Methode soll uns ja gerade dazu verhelfen, mittels dieses bescheidenen Wissens auf die Beschaffenheit des übrigen Teils des Universums schließen zu können. Wir können deshalb die Gleichung (40) niemals dazu verwenden, um die für diese faktische Welt optimale und faire Methode zu bestimmen, da uns der Zustand und insbesondere der Uniformitätsgrad unbekannt ist. Um (40) doch anwenden zu können, müssen wir also *hypothetisch* einen solchen Grad u und damit den entsprechenden Wert L und die so festgelegte Methode c^L akzeptieren. Da wir niemals einen genügend großen Teil des Universums kennen werden, ist diese Hypothese durch kein (im Vergleich zum Universum hinreichend kleines) Erfahrungsdatum zu widerlegen. Im allgemeinen werden wir jedoch nicht mit einem derartigen dogmatischen Starrsinn an einer solchen mehr oder weniger willkürlichen Annahme beharren; vielmehr werden wir uns, wenn sich die Annahme, daß u der tatsächliche Uniformitätsgrad ist, als

allem Anschein nach unrichtig erweist, diese verwerfen und (bis auf weiteres) durch eine andere, bezüglich des vorliegenden Erfahrungsdatums zu besseren Resultaten führende ersetzen.
Die Gleichung (40) bestimmt eine umkehrbar eindeutige Funktion; ihre Umkehrung lautet nämlich:

$$(44) \quad u = \frac{L^*/q + 1}{L^* + 1}.$$

Betrachtet man also eine durch L^* bestimmte adäquate induktive Methode als fair, dann setzt man voraus, daß der Uniformitätsgrad die durch (44) bestimmte Zahl u ist. Da nun nach Voraussetzung der Gegenstandsbereich B von der endlichen Anzahl n ist, kann zunächst nicht ausgeschlossen werden, daß u hinsichtlich dieses n überhaupt kein möglicher Uniformitätsgrad ist, daß u dann also gar nicht der Gleichung (32) genügen kann. Ist u eine rationale Zahl (zwischen $1/q$ und 1), so können wir aber B und damit n entsprechend groß (aber immer noch endlich) wählen, so daß diese Gleichung erfüllt ist; ist u eine irrationale Zahl (zwischen $1/q$ und 1), so muß man zu diesem Zweck einen unendlichen Gegenstandsbereich B wählen, wobei man aus den bisherigen Gleichungen durch Grenzwertbildung für n gegen ∞ neue erhält. Das gleiche gilt für die Uniformitätsgrade zwischen 0 und $1/q$. Zwar kann man $1/q$ beliebig der 0 annähern, wenn man die Anzahl p der Grundprädikate (und damit die Anzahl $q = 2^p$ der V-Prädikate) nur genügend groß wählt; doch die Zahl 0 als möglichen Uniformitätsgrad erhält man erst, wenn man p über alle Grenzen wachsen läßt, also als Grenzwert, wenn p gegen ∞ geht.
Jeder dieser Werte u ist Uniformitätsgrad wenigstens einer Zustandsbeschreibung Σ und damit wenigstens einer möglichen Welt (die durch Σ beschrieben wird). u ist jedoch (bei nichtleeren Bereichen) stets auch der Uniformitätsgrad *mehrerer* möglichen Welten, und zwar in endlichen Sprachen endlich vieler, in Sprachen mit abzählbar unendlich vielen Gegenstandsausdrücken oder Prädikaten überabzählbar unendlich vieler, so daß die Umkehrung an dieser Stelle nicht eindeutig ist; man kann also zwar aus der Zustandsbeschreibung den Uniformitätsgrad errechnen, aber aus diesem Grad nicht wiederum die Zustandsbeschreibung erschließen (selbst bei fest vorgegebenen n und p nicht).

Da wir den Zustand unserer faktischen Welt nicht kennen und also nicht wissen, welche Zustandsbeschreibung tatsächlich wahr ist, können wir auf diese Weise nicht das optimale L^* bestimmen. Trotzdem können wir unter Zuhilfenahme des (relativ kleinen) Erfahrungsdatums Φ_k einiges über das unbekannte L^* ausfindig machen und insbesondere die untere Grenze für diesen Wert angeben. Voraussetzung dafür ist, daß das Universum nicht durchgehend uniform ist, was bereits dann als gesichert gelten kann, wenn der beobachtete Teilbereich $\{a_1, \ldots, a_k\}$ nicht uniform ist, wenn also in Φ_k wenigstens zwei verschiedene V-Prädikate vorkommen. Dies wird im folgenden angenommen.

Σ^* sei die wahre Zustandsbeschreibung, die also jedem Gegenstand aus B gerade die V-Eigenschaft zuschreibt, die er tatsächlich besitzt. Die Anzahlen, die diese q Eigenschaften in diesem Bereich haben, seien $n^*_1, \ldots, n^*_q$; somit ist $r^*_i = n^*_i/n$ für $1 \leqslant i \leqslant q$. Wir können, da wir nur die Stichprobe kennen, deduktiv nicht auf Σ^* selbst schließen. Trotzdem können wir jedoch deduktiv einige Eigenschaften dieses wirklichen Zustands ausfindig machen. So folgt aus Φ_k, daß für beliebige i gilt: $n^*_i \geqslant k_i$, und somit: $r^*_i \geqslant k_i/n$. Es sei $u^* = \sum_{i=1}^{q} r^*_i{}^2$; k_j sei die größte dieser Zahlen k_i (bzw., wenn mehrere die größten sind, eine daraus willkürlich ausgewählt). Mit diesem Erfahrungsdatum Φ_k ist nun eine Zustandsbeschreibung Σ' logisch verträglich, die *fast homogen* ist, die also den restlichen Gegenständen (die in Φ_k nicht erwähnt werden), die V-Eigenschaft F^*_j zuschreibt, die ja im beobachteten Teil von der Anzahl k_j ist; durch Σ' wird also angenommen, daß wir im Datum Φ_k bereits sämtliche Gegenbeispiele gefunden haben. Σ' ist demnach sozusagen die homogenste aller mit Φ_k verträglichen Zustandsbeschreibungen. Es sei für jedes i mit $1 \leqslant i \leqslant q$ die Anzahl n'_i die Kardinalzahl, die F^*_i nach Σ' besitzt, und es sei entsprechend $r'_i = n'_i/n$. Dann gilt für jedes i mit $i \neq j$: $n'_i = k_i$, und es gilt ferner: $n'_j = n - k + k_j$. Es sei $u' = \sum_{i=1}^{q} r'_i{}^2$. Da Φ_k nicht homogen ist, gilt dies auch von Σ', so daß:

(45) $u' < 1$

Die wahre Zustandsbeschreibung Σ^* kann nicht homogener sein als Σ'; es gilt daher:

(46) $u^* \leqslant u' < 1$

u' ist also die obere Grenze für die nach Φ_k möglichen Uniformitätsgrade der Welt, d. h. u' ist das Maximum der Uniformitätsgrade der mit Φ_k logisch verträglichen Zustandsbeschreibungen. Die wahren Werte u^*, L^* und Σ^* sind uns nun zwar unbekannt, doch können wir sie mittels (46) unter Zuhilfenahme von (40) abschätzen. Es gilt dann:

(47) $L' > 0$

(48) $L^* \geqslant L' > 0$

Also ist L' eine untere Grenze für das unbekannte L^* und das Minimum der optimalen L-Werte für alle mit Φ_k logisch verträglichen Zustandsbeschreibungen.

Es sei $s^* = \mathrm{srh}_{c^{L^*}}(F, A, \Phi_k)$ und $s' = \mathrm{srh}_{c^{L'}}(F, A, \Phi_k)$. Nun gilt von jeder nichthomogenen Zustandsbeschreibung Σ (so daß $u < 1$), daß die Werte $Mq(s, r)$ stets kleiner werden, wenn L von 0 bis L^* wächst [dies folgt für $u > 1/q$ aus (40) und für $u = 1/q$ aus (43) (a)]. Es sei $s_0 = \mathrm{srh}_{c^0}(F, A, \Phi_k)$; dann gilt:

(49) $Mq(s^*, r) \leqslant Mq(s', r) < Mq(s_0, r)$.

Obwohl also die wahre Zustandsbeschreibung Σ^* unbekannt ist, kann doch auf Grund der aus Φ_k erschlossenen Nichthomogenität von Σ^* gesagt werden, daß in einer durch Σ^* beschriebenen Welt die Methode $c^{L'}$ erfolgreicher sein wird als die Proportionalregel c^0, da sie einen kleineren mittleren quadratischen Irrtum für die Schätzung bezüglich der V-Eigenschaften im Hinblick auf alle Stichproben der Größe k hat als diese in der Statistik meist benützte Proportionalregel.

Dieses Ergebnis darf nicht so mißverstanden werden, daß man auf der Grundlage des Erfahrungsdatums Φ_k mit $c^{L'}$ stets einen größeren Erfolg haben wird als mit c^0; wenn Σ^* identisch ist mit Σ', wird man im Gegenteil mit c^0 erfolgreicher abschneiden als mit $c^{L'}$, da sich c^0 bei zunehmend neuen Erfahrungen (die dann durchgehend uniform sind) schneller auf die veränderte Situation einstellt als $c^{L'}$ (und schneller als jede andere L-Methode mit positivem L). Gemeint ist mit diesem Ergebnis nur, daß man, wenn man von *irgendeiner* derartigen Stichprobe der Größe k ausgeht, *im Mittel* mit $c^{L'}$ einen größeren Erfolg haben wird als mit c^0. Denn $Mq(s', r)$ ist ja der mittlere qua-

dratische Irrtum der Schätzung der relativen Häufigkeiten aller V-Eigenschaften in A, wenn man hierbei die Funktion $c^{L'}$ verwendet und *alle Stichproben* der vorgegebenen Größe k in der durch $\Sigma^\star$ beschriebenen Welt zugrundelegt. Wenn wir nach den durch Φ_k beschriebenen Erfahrungen zusätzliche Beobachtungen machen, ohne dabei die früheren zu vergessen, erhalten wir ja immer umfassendere Stichproben; das Ergebnis (49) hingegen nimmt nicht auf diese immer umfassenderen, sondern auf jene der fest vorgegebenen Größe k Bezug. Ob wir in einem solchen konkreten Fall mit $c^{L'}$ erfolgreicher abschneiden werden als mit c^0, hängt also davon ab, welche Eigenschaften diese neu beobachteten Gegenstände haben. (49) sagt uns also nicht etwas über *diese konkrete* Folge von Beobachtungen etwas aus (die wir in unserem Leben tatsächlich machen), sondern spricht von *allen derartigen* Folgen von Beobachtungen und gibt uns von ihnen einen Mittelwert.

Ein derartiges mögliches Mißverständnis wird ausgeschlossen, wenn die Nichthomogenität von $\Sigma^\star$ nicht aus der von Φ_k erschlossen wird, sondern wenn unmittelbar vorausgesetzt wird, daß $\Sigma^\star$ nicht homogen ist, wie es in dem folgenden Satz geschieht:

(50) Gegeben sei eine endliche Sprache, die durch die Anzahlen n und q bestimmt ist. Die Methoden $c^{L'}$ und damit $\mathrm{srh}_{c^{L'}}$ seien festgelegt durch
$L' = 2 \cdot q \cdot (n-1)/[(q-1) \cdot n^2 - 2 \cdot q \cdot (n-1)]$
(so daß $L' > 0$). Dann gilt für jede nichthomogene Zustandsbeschreibung Σ und für jedes Erfahrungsdatum der Anzahl k (mit $k \ll n$): $\mathrm{Mq}(s', u) < \mathrm{Mq}(s_0, u)$.[9]

Der Wert L' ergibt sich dabei folgendermaßen: Den größten Wert u' für nichthomogene Zustandsbeschreibungen erhält man, wenn $n-1$ die Anzahl einer V-Eigenschaft ist, 1 die Anzahl einer anderen und 0 die Anzahl der übrigen; dann ist $u' = (n^2 - 2 \cdot n + 2)/n^2$, und der Wert für L' ergibt sich sodann nach (40).

Ist also nach (40) die Welt nicht durchgehend (hinsichtlich der uns interessierenden Eigenschaften) uniform und damit die sie charakterisierende Zustandsbeschreibung nicht durchgehend homogen, dann

---

[9] Ist Σ hingegen durchgehend homogen, so gilt selbstverständlich für jedes derartige Erfahrungsdatum von der Anzahl k: $\mathrm{Mq}(s', u) > \mathrm{Mq}(s_0, u)$.

arbeitet man mit verzerrten L-Methoden im Durchschnitt erfolgreicher als mit der unverzerrten Funktion c^0.

Von einem induktiven Standpunkt aus (sei es ein formaler oder ein intuitiver) erscheint die obere Grenze u′ für u^* als viel zu hoch und entsprechend die untere Grenze L′ für L^* als viel zu tief gewählt. Man wird also den Wert u^* und entsprechend den Wert L^* mit einer induktiven Methode zu *schätzen* versuchen. Schätzt man ihn mittels der Proportionalregel, so erhält man das Resultat, daß die Uniformität in A den gleichen Grad hat wie in dem beobachteten Teil $\{a_1, \ldots, a_k\}$ des Universums, daß diese Stichprobe hinsichtlich der Uniformität also *repräsentativ* für A (und damit approximativ auch für B) ist;[10] der durch c^0 geschätzte Uniformitätsgrad wird im allgemeinen (wenn Φ_k nicht homogen ist) kleiner als 1 sein, und es wird demnach mittels c^0 eine induktive Methode c^L als optimal eingeschätzt, für die $L > 0$ ist. Schätzt man den Uniformitätsgrad durch c^∞, so erhält man als Resultat, daß keine Uniformität vorliegt und also c^∞ gerechtfertigt ist, und schätzt man ihn mit einer dazwischenliegenden L-Methode, so erhält man einen dazwischenliegenden Wert.

Wenn das Universum hinsichtlich der uns interessierenden Eigenschaften also nicht durchgehend uniform ist, dann haben wir mit verzerrten Methoden im Durchschnitt bessere Erfolge als mit unverzerrten. Am erfolgreichsten wären wir (im Durchschnitt) dann allerdings mit der Funktion c^{L^*}, die nach (40) dem (uns allerdings unbekannten) Uniformitätsgrad u^* des Universums entspricht. Nun haben unsere Erfahrungswissenschaften nicht den Stand, daß sie *alle* Beobachtungen und Meßergebnisse unter strenge und ausnahmslos geltende Gesetzesaussagen subsumieren können, sondern müssen vielmehr unbekannte Eigenschaften annehmen, um derartige Behauptungen aufrechterhalten zu können (indem etwa gesagt wird, daß singuläre Gegenbeispiele, die scheinbar die Theorie widerlegen, aber nicht reproduzierbar sind, durch irgendeinen unbekannten Effekt

---

10 Ist umgekehrt diese Repräsentativität bekannt, so ist die Verwendung einer induktiven Methode, die sich hinsichtlich Stichproben dieser Größe praktisch nicht mehr von der Proportionalregel unterscheidet, gerechtfertigt, so daß man aus Einfachheitsgründen für Berechnungen einfach diese Proportionalregel selbst wählen kann.

beim Meßverfahren zustandegekommen sind, den man bei Ausschaltung aller möglichen Störungen hätte vermeiden können); im Alltag sind wir darüber hinaus sogar ständig darauf gefaßt, Ausnahmen für unsere Lebensregeln anzutreffen. Es ist daher sowohl in den Wissenschaften als auch im Alltag unvernünftig, beim Schluß von Beobachtungen auf Unbeobachtetes die unverzerrte Proportionalregel an Stelle von verzerrten Methoden des Kontinuums zu verwenden.

Literaturverzeichnis

Abkürzungen

BrJ The British Journal for the Philosophy of Science
Erk Erkenntnis
JPh The Journal of Philosophy
JSL The Journal of Symbolic Logic
MZs Mathematische Zeitschrift
PPR Philosophy and Phenomenological Research
PRv The Philosophical Review
PSc Philosophy of Science
PSt Philosophical Studies (Minneapolis)

Aristoteles

[an. pr.] Analytica priora (= Organon III, sog. Lehre vom Schluß), übersetzt von Eugen Rolfes als Band 10 der Philosophischen Bibliothek, Leipzig 1921.
[an. post.] Analytica posteriora (= Organon IV, sog. Lehre vom Beweis), übersetzt von Eugen Rolfes als Band 11 der Philosophischen Bibliothek, Leipzig 1922.
[top.] Topik (= Organon V), übersetzt von Eugen Rolfes als Band 12 der Philosophischen Bibliothek, Leipzig ^{2}1922.
[rhet.] Rhetorik, übersetzt von Paul Gohlke, Paderborn 1959.

Bacon, Franz

[1] Neues Organ der Wissenschaften, Darmstadt 1962 (erstmals Leipzig 1830).
[2] Über die Würde und den Fortgang der Wissenschaften, Darmstadt 1966 (erstmals Pest 1783).

Basler, Herbert

[1] Grundbegriffe der Wahrscheinlichkeitsrechnung und statistischen Methodenlehre, Würzburg und Wien 1968.

Bar-Hillel, Yehoshua

[1] A Note on State-Descriptions, in: PSt 2 (1951) S. 72—75.
[2] A Note on Comparative Inductive Logic, in: BrJ 3 (1953) S. 308—310.
[3] An Examination of Information Theory, in: PSc 22 (1955) S. 86—105.
[4] Comments on „Degree of Confirmation" by Professor K. R. Popper, in: BrJ 6 (1955/56) S. 155—157.

[5] Further Comments on Probability and Confirmation, in: BrJ 7 (1956/57) S. 245—248.
[6] Information and Content: A Semantic Analysis, in: Synthese 9 (1953—55) S. 299—305.
[7] On an Alleged Contradiction in Carnap's Theory of InductiveLogic: in Mind 73 (1964) S. 265—267.
[8] On Alleged Rules of Detachment in Inductive Logic, in: Lakatos [1], S. 120—128.

Bar-Hillel, Yehoshua — Carnap, Rudolf

[1] Semantic Information, in: BrJ 4 (1953) S. 147—157.
[2] Comments on „Degree of Confirmation" by Professor K. R. Popper, in: BrJ 6 (1955/56) S. 155—157.
[3] Further Comments on Probability and Confirmation. A Rejoinder to Professor Popper, in: BrJ 7 (1956/57) S. 245—248.

Bernoulli, Jakob

]1] Ars conjectandi, Basileae 1713.
[2] Wahrscheinlichkeitsrechnung, Leipzig 1899. (Übersetzung von [1] durch R. Haussner.)

Boole, George

[1] An Investigation of the Laws of Thougt, on which are Founded the Mathematical Theories of Logic and Probabilities, London 1854 (= Collected Logical Works, Vol. 2, hrsg. von P. E. B. Jourdain, Chicago—London 1916).

Braithwaite, R. B.

[1] On Unknown Probabilities, in: Körner [1], S. 3—11.

Carnap, Rudolf

[1] Testability and Meaning, in: PSc 3 (1936) S. 419—471 und 4 (1937) S. 1 bis 40.
[2] On Inductive Logic, in: PSc 12 (1945) S. 72—97.
[3] The Two Concepts of Probability, in: PPR 5 (1945) S. 513—532.
[4] Remarks on Induction and Truth, in: PPR 5 (1946) S. 590—602.
[5] Rejoinder to Mr. Kaufmann's Reply, in: PPR 6 (1946) S. 609—611.
[6] Probability as a Guide in Life, in: JPh 44 (1947) S. 141—148.
[7] On the Application of Inductive Logic, in: PPR 8 (1947/48) S. 133—148.
[8] Reply to Nelson Goodman, in: PPR 8 (1947/48) S. 461—462.
[9] Logical Foundations of Probability, Chicago ^{1}1950, ^{2}1962.
[10] The Problem of Relations in Inductive Logic, in: PSt 2 (1951/52) S. 75—80.
[11] The Continuum of Inductive Methods, Chicago 1952.
[12] Meaning Postulates, in: PSt 3 (1952) S. 65—73.
[13] On the Comparative Concept of Confirmation, in: BrJ 3 (1953) S. 311 bis 318.

[14] Inductive Logic and Science, in: Proceedings of the American Academy of Arts and Sciences 80 (1953) S. 189—197.
[15] Remarks to Kemeny's Paper, in: PPR 13 (1953) S. 375—376.
[16] What is Probability?, in: Scientific American 189 (1953) S. 128—138.
[17] The Aim of Inductive Logic, in: Logic, Methodology and Philosophy of Science (Proceedings of the 1960 International Congress), hrsg. von Ernest Nagel, Patrick Suppes und Alfred Tarski, Standford, Calif. 1962.
[18] Intellectual Autobiography, in: Schilpp [1], S. 3—84.
[19] Replies and Systematic Expositions, in: Schilpp [1], S. 859—1013.
[20] Philosophical Foundations of Physics, New York 1966.

Carnap, Rudolf — Stegmüller, Wolfgang

[1] Induktive Logik und Wahrscheinlichkeit, Wien 1958.

Cramér, Harald

[1] Mathematical Methods of Statistics, Princeton 1946.

Essler, Wilhelm Karl

[1] Einführung in die Logik, Stuttgart ([1]1966) [2]1969.
[2] Über Intensionen und Modalitäten, in: Notre Dame Journal of Formal Logic 11 (1970).
[3] Über die Interpretation von Wissenschaftssprachen, in: Philosophisches Jahrbuch 77 (1970).

Finetti, Bruno de

[1] La prévision: ses lois logiques, ses sources subjectives, in: Annales de l'Institut Henri Poincaré 7 (1937) S. 1—68.
[2] Recent Suggestions for the Reconciliation of Theories of Probability, in: Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, hrsg. von Jerzy Neyman, Berkeley—Los Angeles 1951, S. 217—225.

Fisher, Ronald A.

[1] Statistical Methods and Scientific Inference, Edinburgh 1959.

Gaifman, Haim

[1] Concerning Measures in First Order Calculi, in: Israel Journal of Mathematics 2 (1964) S. 1—18.

Goodman, Nelson

[1] A query on Confirmation, in: JPh 43 (1946) S. 383—385.
[2] On Infirmities of Confirmation-Theory, in: PPR 8 (1947/48) S. 149—151.

[3] Fact, Fiction, and Forecast, Cambridge, Mass. 1955.
[4] Axiomatic Measurement of Simplicity, in: JPh 52 (1955) S. 709—722.
[5] Recent Developments in the Theory of Simplicity, in: PPR 19 (1959) S. 429—446.
[6] The Structure of Appearance, New York ²1966.

Hempel, Carl G.

[1] Über den Gehalt von Wahrscheinlichkeitsaussagen, in: Erk 5 (1935/36) S. 228—260.
[2] On the Logical Form of Probability-Statements, in: Erk 7 (1937—39) S. 154—160 und S. 360—363.

[3] Studies in the Logic of Confirmation, in: Mind 54 (1945) S. 1—26 und S. 97—121.

Hilpinen, Risto

[1] Rules of Acceptance and Inductive Legic (Acta Philosophica Fennica 21) Amsterdam 1968.

Hintikka, Jaakko

[1] Towards a Theory of Inductive Generalization, in: Proceedings of the 1964 International Congress for Logic, Methodology and Philosophy of Science, Jerusalem 1964, S. 274—288.
[2] On a Combined System of Inductive Logic, in: Acta Philosophica Fennica 18 (1965) S. 21—30.
[3] A Two-Dimensional Continuum of Inductive Methods, in: Hintikka-Suppes [1], S. 113—132.
[4] Induction by Enumeration and Induction by Elimination, in: Lakatos [1], S. 191—216.
[5] Statistics, Induction, and Lawlikeness: Comments on Dr. Vetter's Paper, in: Synthese 20 (1969) S. 72—83.

Hintikka, Jaakko — Suppes, Patrick

[1] Aspects of Inductive Logic, Amsterdam 1966.

Humburg, Jürgen

[1] Die Problematik des induktiven Schließens bei Carnap und Richter, unveröffentlichte Diplomarbeit München 1964.

Hume, David

[1] A Treatise on Human Nature, being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects, London 1739.
[2] An Enquiry Concerning Human Understanding, London 1748.

Jeffreys, Harold

[1] Theory of Probability, Oxford 1939.

Kahlius, Ludwig Martin

[1] Elementa logicae probabilium, methodo mathematica in usum scientarum et vitae adornata, Halle 1735.

Kemeny, John G.

[1] Carnap on Probability, in: Review of Metaphysics 5 (1951) S. 145—156.
[2] Extension of the Methods of Inductive Logic, in: PSt 3 (1952) S. 38—42.
[3] A Contribution to Inductive Logic, in: PPR 13 (1953) S. 371—374.
[4] The Use of Simplicity in Induction, in: PRv 62 (1953) S. 391—408.
[5] A Logical Measure Function, in: JSL 18 (1953) S. 289—308.
[6] Fair Bets and Inductive Probabilities, in: JSL 20 (1955) S. 263—273.
[7] Two Measures of Complexity, in: JPh 52 (1955) S. 722—733.
[8] A Philosopher looks at Science, Princeton—London 1959.
[9] Carnap's Theory of Probability and Induction, in: Schilpp [1], S. 711—737.

Keynes, John Maynard

[1] A Treatise on Probability, New York ^{7}1962.

Kneale, William

[1] Probability and Induction, Oxford ^{2}1952.

Körner, Stephan

[1] (Hrsg.) Observation and Interpretation, A Symposium of Philosophers and Physicists, London 1957.

Kolmogoroff, A. N.

[1] Grundbegriffe der Wahrscheinlichkeitsrechnung, in: Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 2, Berlin 1933, S. 195—262.

Kries, J. von

[1] Die Prinzipien der Wahrscheinlichkeitsrechnung, eine logische Untersuchung, Freiburg i. Br. 1886.

Kyburg (Jr.), Henry E.

[1] Probability and the Logic of Rational Belief, Middletown (Connecticut) 1961.
[2] Probability and Inductive Logic, London 1970.

Lakatos, Imre

[1] (Hrsg.) The Problem of Inductive Logic, Amsterdam 1968.
[2] Changes in the Problem of Inductive Logic, in: Lakatos [1], S. 315—417.

Laplace, Pierre Simon de

[1] Mémoire sur la probabilité des causes par les evenements, in: Mémoires presentées à l'Academie des Sciences, 1774.
[2] Mémoire sur les approximations des formules qui sont fonctions de très grands nombres, in: Mémoire de l'Institut, 1810.
[3] Théorie analytique des probabilités, Paris 1812.
[4] Essai philosophique sur les probabilités, Paris 1814.
[5] Philosophischer Versuch über Wahrscheinlichkeiten, Heidelberg 1819.
[6] Oeuvres complètes, Paris 1891—98.

Lehman, R. Sherman

[1] On Confirmation and Rational Betting, in: JSL 20 (1955) S. 251—262.

Mill, John Stuart

[1] System der deductiven und inductiven Logik. Eine Darlegung der Grundsätze der Beweislehre und der Methoden wissenschaftlicher Forschung, Bände I—III (= J. St. Mills Gesammelte Werke, Bände 2—4), Aalen 1968 (erstmals Leipzig 1872/73).

Mises, Richard von

[1] Grundlagen der Wahrscheinlichkeitsrechnung, in: MZs 5 (1919) S. 52—99.
[2] Wahrscheinlichkeit, Statistik und Wahrheit. Einführung in die neue Wahrscheinlichkeitslehre und ihre Anwendung, Wien 1928, ^{2}1936.
[3] Über kausale und statistische Gesetzmäßigkeit in der Physik, in: Erk 1 (1930/31) S. 189—210.
[4] Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik, Wien 1931.

Philodemus

[1] On Methods of Inference — A Study in Ancient Empiricism (hrsg., übersetzt und kommentiert von Ph. H. de Lacy und E. A. de Lacy), Philadelphia (Pennsylvania) 1941.

Popper, Sir Karl R.

[1] „Induktionslogik" und „Hypothesenwahrscheinlichkeit", in: Erk 5 (1935) S. 170—172.
[2] Logik der Forschung. Zur Erkenntnistheorie der modernen Naturwissenschaft, Wien ^{1}1935, Tübingen ^{2}1966.

[3] A Set of Independent Axioms for Probability, in: Mind 47 (1938) S. 275 bis 277.
[4] Degree of Confirmation, in: BrJ 5 (1954/55) S. 143—149.
[5] Two Autonomous Axiom Systems for the Calculus of Probabilities, in: BrJ 6 (1955/56) S. 51—57.
[6] „Content" and „Degree of Confirmation": A Reply to Dr. Bar-Hillel, in: BrJ 6 (1955/56) S. 157—163.
[7] Reply to Professor Carnap, in: BrJ 7 (1956/57) S. 244—245.
[8] Adequacy and Consistency: A Second Reply to Dr. Bar-Hillel, in: BrJ 7 (1956/57) S. 249—256.
[9] A Second Note on Degree of Confirmation, in: BrJ 7 (1956/57) S. 350 bis 353.
[10] The Propensity Interpretation of the Calculus of Probability, and the Quantum Theory, in: Körner [1], S. 65—70.
[11] A Third Note on Degree of Corroboration or Confirmation, in: BrJ 8 (1957/58) S. 294—302.
[12] The Propensity Interpretation of Probability, in: BrJ 10 (1959/60) S. 25—42.
[13] Probabilistic Independence and Corroboration by Empirical Tests, in: BrJ 10 (1959/60) S. 315—318.
[14] On Carnap's Version of Laplace's Rule of Succession, in: Mind 71 (1962) S. 69—73.
[15] Creative and Non-creative Definitions in the Calculus of Probability, Synthese 15 (1963) S. 167—186.
[16] Conjectures and Refutations, London 1963.
[17] The Mysteries of Udolpho: A Reply to Professors Jeffrey and Bar-Hillel in: Mind 76 (1967) S. 103—110.
[18] Quantum Mechanics without „The Observer", in: M. Bunge (Hrsg.), Quantum Theory and Reality, Berlin—Heidelberg—New York 1967, S. 7—44.

Post, H. R.

[1] Simplicity in Scientific Theories, in: BrJ 2 (1960/61) S. 32—41.

Ramsey, Frank P.

[1] The Foundations of Mathematics and other Logical Essays, London—New York 1931.

Reichenbach, Hans

[1] Der Begriff der Wahrscheinlichkeit für die mathematische Darstellung der Wirklichkeit, in: Zeitschrift für Philosophie und philosophische Kritik 161 (1917) und 162 (1918).
[2] Kausalität und Wahrscheinlichkeit, in: Erk 1 (1930/31) S. 158—188.
[3] Axiomatik der Wahrscheinlichkeitsrechnung, in: MZs 34 (1932) S. 568 bis 619.
[4] Die logischen Grundlagen des Wahrscheinlichkeitsbegriffs, in: Erk 3 (1932/33) S. 401—425.

[5] Wahrscheinlichkeitslehre. Eine Untersuchung über die logischen und mathematischen Grundlagen der Wahrscheinlichkeitsrechnung, Leiden 1935.
[6] Wahrscheinlichkeitslogik, in: Erk 5 (1935/36) S. 37—43.
[7] Experience and Prediction. An Analysis of the Foundations and the Structure of Knowledge, Chicago 1938.
[8] Über die semantische und die Objektauffassung von Wahrscheinlichkeitsaussagen, in: The Journal of Unified Science (= Erk) 8 (1939/40) S. 50—68.
[9] On the Justification of Induction, in: JPh 37 (1940) S. 97—103.
[10] The Theory of Probability, Berkeley—Los Angeles 1949 (= Übersetzung von [5]).

Richter, Hans

[1] Zur Grundlegung der Wahrscheinlichkeitstheorie, in: Mathematische Annalen 125 (1953) S. 129—139, 223—234, 335—343; 126 (1953) S. 362 bis 374; 128 (1954) S. 305—339.

[2] Zur Begründung der Wahrscheinlichkeitsrechnung, in: Dialectica 8 (1954) S. 48—77.
[3] Wahrscheinlichkeitstheorie, Berlin—Heidelberg—New York 1956, ²1966.

Savage, Leonard J.

[1] The Foundations of Statistics, New York 1954.
[2] Subjective Probability and Statistical Practice, Chicago 1959.

Schilpp, Paul Arthur

[1] The Philosophy of Rudolf Carnap, La Salle, Ill. 1963.

Scott, Dana — Kraus, Peter

[1] Assigning Probabilities to Logical Formulas, in : Hintikka-Suppes [13, S. 219-264

Shimony, Abner

[1] Coherence and the Axioms of Confirmation, in: JSL 20 (1955) S. 1—28.

Stegmüller, Wolfgang

[1] Bemerkungen zum Wahrscheinlichkeitsproblem, in: Studium generale 6 (1953) S. 563—593.

Todhunter, I.

[1] A History of the Mathematical Theory of Probability, from the Time of Pascal to that of Laplace, New York 1931 (erstmals London 1865).

Venn, John

[1] The Logic of Chance; an Essay on the Foundations and Province of the Theory of Probability, with Especial Reference to its Logical Bearings and its Application to Moral and Social Science and to Statistics, London—New York 1866.

Vetter, Hermann

[1] Wahrscheinlichkeit und logischer Spielraum. Eine Untersuchung zur induktiven Logik, Tübingen 1967.
[2] Logical Probability, Mathematical Statistics, and the Problem of Induction, in: Synthese 20 (1969) S. 56—71.

Waismann, Friedrich

[1] Logische Analyse des Wahrscheinlichkeitsbegriffs, in: Erk 1 (1930/31) S. 228—248.

Wald, Abraham

[1] Die Widerspruchsfreiheit des Kollektivbegriffs der Wahrscheinlichkeitsrechnung, in: Ergebnisse eines mathematischen Kolloquiums 8 (1937) S. 38—72.
[2] On the Principles of Statistical Inference, Notre Dame 1942.
[3] Statistical Decision Functions, New York 1950.
[4] Selected Papers in Statistics and Probability, New York 1955.

Wittgenstein, Ludwig

[1] Logisch-Philosophische Abhandlung (= Tractatus logico-philosophicus), erstmals in: Annalen der Naturphilosophie 14 (1921) S. 185—262, Neudruck Frankfurt 1960.

Wright, George Henrik de

[1] A Treatise on Induction and Probability, London 1951.
[2] The Logical Problem of Induction, Oxford ²1965.