Zur Problematik der Ableitung in sozialwissenschaftlichen Aussagensystemen

Ein Plädoyer für Formalisierung (Teil 2)

Hans J. Hummell

Forschungsinstitut für Soziologie der Universität zu Köln

On the Problem of Deducing Valid Conclusions in Systems of Social Science Propositions. Part 2

The structure of some kinds of arguments typically to be found in theoretical-empirical social science is examined with special attention to the rules of deduction which in the majority of cases remain implicit. Four types of propositions – describing deterministic relations between attributes (statements of the form ‘if x, then y’), describing monotonous deterministic relations between serials (‘the more x, the more y’) and their probabilistic counterparts – as well as two classes of rules of deduction (based on the properties of transitivity and conjunctivity of corresponding statement forms) are treated. Of the propositions only those of the deterministic if-then kind are unproblematic. The thesis is advanced that most of those ‘arguments’ presented in social science consisting of sentences of relatively complicated structures – using rules, however, which are exclusively valid for deterministic if-then-statements – are not correct.

Deterministic propositions being discussed in Part One, Part Two treats probabilistic ones with the following results: (I) By considering the ‘statistical syllogism’ the non-conjunctivity of probabilistic implications is shown; (II) their non-transitivity follows from the fact that in complex causal structures direct and indirect effects may vary independently of each other. For the simplest case of three attributes a qualitative analogon to the Simon-Blalock-Procedure is constructed and illustrated by an example from mobility research. (III) If monotonous probabilistic relations are characterized by coefficients of linear correlation, in the general case no deductions are possible which would use a rule of transitivity.

As a general solution to the problem of deducing valid conclusions it is suggested that the verbal language – which, though enriched by a technical vernacular, with respect to its formal apparatus is entirely based on elementary logics – should be replaced by languages of richer logical structures; this, however, means formalization. According to this function to warrant valid deductions the concept of formalization is specified and delimited from similar concepts as the construction of propositions within a linguistic framework of relatively rich logical structure which is given as a special calculus.

Inhalt: Untersucht werden einige in den theoretisch-empirischen Sozialwissenschaften typische Formen von Argumenten im Hinblick auf die (meist implizit gelassenen) Deduktionsregeln. Es werden vier Arten von Aussagen – (a) Deterministische Wenn-Dann-Aussagen; (b) Deterministische Je-Desto-Aussagen; (c) Probabilistische Wenn-Dann-Aussagen („W-Implikationen“); (d) Probabilistische Je-Desto-Aussagen – sowie zwei Klassen von Deduktionsregeln (basierend auf den Eigenschaften der Transitivität und Konjunktivität) behandelt. Problematisch sind nur Aussagen der Art (b), (c) und (d). Es wird die These vertreten, daß die meisten der in den Sozialwissenschaften vorgelegten „Argumente“, welche mit solchen Aussagen relativ komplizierter Struktur arbeiten, zur Ableitung jedoch nur für deterministische Wenn-Dann-Aussagen gültige Regeln verwenden, nicht korrekt sind.

Nachdem in Teil I bereits deterministische Aussagen diskutiert wurden, werden hier probabilistische behandelt, wobei sich folgende Resultate ergeben: 1. Anhand des „statistischen Syllogismus“ wird die Nicht-Konjunktivität von W-Implikationen gezeigt. 2. Ihre Nicht-Transitivität ergibt sich daher, daß in Wirkungsstrukturen direkte und indirekte Effekte unabhängig voneinander variieren können. Als Lösungsmöglichkeit wird für den einfachsten Fall von drei Attributen eine qualitative Analogie zum SIMON-BLALOCK-Verfahren konstruiert und an einem Beispiel aus der Mobilitätsforschung erläutert. 3. Stellt man monotone probabilistische Beziehungen durch Koeffizienten der linearen Korrelation dar, können ebenfalls i.a. keine Ableitungen formuliert werden, die von einer Transitivitätsregel Gebrauch machen.

Zur Lösung des Deduktionsproblems wird abschließend vorgeschlagen, an Stelle einer durch ein technisches Vokabular ergänzten Wortsprache, der ausschließlich die elementare Logik zugrundeliegt, generell Sprachen mit reicherer logischer Struktur zu verwenden, d.h. zu formalisieren. Von der Funktion her, Deduktionen zu ermöglichen, wird der Begriff der Formalisierung bestimmt und als Konstruktion von Aussagen in strukturreichen Sprachen, die als spezielle Kalküle gegeben sind, von verwandten Begriffen abgegrenzt.

II.

2. Probabilistische Wenn-Dann-Aussagen

Betrachten wir eine Menge von Attributen eines gegebenen Objektbereiches, zwischen denen probabilistische Beziehungen bestehen. Wie erwähnt, können wir diese durch „Wahrscheinlichkeitsimplikationen“ (kurz: *W-Implikationen*) beschreiben,

welche von der Form sind: „Wenn ein Objekt x die Eigenschaft F aufweist, dann besitzt es mit der Wahrscheinlichkeit r auch die Eigenschaft G“, symbolisch: $P(G/F) = r$. Dabei können wir zunächst ohne weiteres davon ausgehen, daß wir die W-Implikation $P(G/F) = r$ in dem Sinne interpretieren, daß damit die *relative Häufigkeit von Objekten mit der Eigenschaft G in der Re-*

ferenzklasse der Objekte mit der Eigenschaft F gemeint ist¹. Kompliziertere Aussagen können als aus Aussagen dieser Form aufgebaut gedacht werden.

Als Beispiel können wir die folgenden drei Dichotomien betrachten, welche der DURKHEIMschen Selbstmordtheorie entnommen sind: die Variable „Alter“ mit den beiden Ausprägungen „Alt“ (A_1) – „Jung“ (A_0), die Variable „Familienstand“ mit den Ausprägungen „Verheiratet“ (B_0) – „Unverheiratet“ (B_1) und die Variable „Begeht Selbstmord“ mit den Ausprägungen „Ja“ (C_1) – „Nein“ (C_0). Von DURKHEIM wird zwischen diesen drei Variablen eine ganz bestimmte Struktur angenommen: ältere Personen begehen eher Selbstmord als jüngere; unverheiratete Personen begehen eher Selbstmord als verheiratete; ob eine Person verheiratet ist oder nicht, variiert mit dem Alter. Die Kausalstruktur dieses Drei-Variablen-Falls läßt sich mithin in folgender Weise darstellen²:

Es interessieren uns also folgende W-Implikationen, die auch im einzelnen von DURKHEIM behandelt werden:

Wenn eine Person ein bestimmtes Alter erreicht

¹ Teilmengen der Bezugsklasse heißen auch „Ereignisse“. $P(G/F)$ wäre dann die („bedingte“) Wahrscheinlichkeit des Ereignisses G , falls das Ereignis F realisiert ist. Die erwähnte Interpretation ist nicht mit der statistischen Interpretation der Wahrscheinlichkeit im üblichen Sinne identisch: letztere bezieht sich auf die relative Häufigkeit, mit der bei wiederholten, aber voneinander unabhängigen Beobachtungen von Elementen aus F („Zufallsstichprobe“ aus F) das Merkmal G anzutreffen ist. Allerdings gilt nach dem Theorem von Bernoulli, daß mit steigendem Stichprobenumfang die beobachtete relative Häufigkeit (im stochastischen Sinne) gegen die relative Häufigkeit von G in der Population F , der die Stichproben entnommen werden, konvergiert.

² Der Einfachheit halber werden in der Zeichnung statt der Ausprägungen nur die Variablen aufgeführt, also z.B. A statt A_i ($i = 0,1$). Ähnlich wird im Text immer dann verfahren, wenn eindeutig ist, daß eine bestimmte Aussage für Variable als Menge von Aussagen für alle möglichen Ausprägungen der Variablen zu verstehen ist.

hat, hat sie eine bestimmte Wahrscheinlichkeit, unverheiratet zu sein³:

$$(I) P(B_1/A_1) = r_1$$

Wenn eine Person unverheiratet ist, hat sie eine bestimmte Wahrscheinlichkeit, Selbstmord zu begehen:

$$(II) P(C_1/B_1) = r_2$$

Wenn eine Person ein bestimmtes Alter erreicht hat, dann hat sie eine bestimmte Wahrscheinlichkeit, Selbstmord zu begehen:

$$(III) P(C_1/A_1) = r_3$$

Erweitern wir diese Menge von Aussagen um eine vierte:

Wenn eine Person ein bestimmtes Alter erreicht hat und wenn diese Person unverheiratet ist, dann hat sie eine bestimmte Wahrscheinlichkeit, Selbstmord zu begehen:

$$(IV) P(C_1/A_1 \wedge B_1) = r_4.$$

Zur Diskussion steht das Problem der logischen Beziehungen zwischen den Aussagen (I) bis (IV) im Hinblick auf die Möglichkeiten der Deduktion einer oder mehrerer Aussagen aus den restlichen. So kann man sich beispielsweise fragen, ob es möglich ist, etwas über den Zusammenhang von Alter und Selbstmord auszusagen (III), wenn die Zusammenhänge zwischen Alter und Familienstand (I) und zwischen Familienstand und Selbstmord (II) bekannt sind; anders formuliert: kann man aus (I) und (II) eine Konsequenz deduzieren, die sich auf (III) bezieht? Gesucht ist also eine Deduktionsfunktion, die r_3 in Abhängigkeit von r_1 und r_2 zeigt. Dies ist das *Problem der „Transitivität“*.

Ferner kann man fragen, ob es möglich ist, aus den bekannten Zusammenhängen zwischen Alter und Selbstmord bzw. Familienstand und Selbstmord etwas über die Selbstmordwahrscheinlichkeiten von alten und unverheirateten Personen oder von

³ Strenggenommen reicht die Berücksichtigung von $P(B_1/A_1)$ nicht aus, wenn man sich für die *Beziehung* zwischen den Dichotomien „Alter“ und „Familienstand“ interessiert, denn dann müßte man die bedingte Wahrscheinlichkeit, verheiratet zu sein, falls man ein bestimmtes Alter erreicht hat, entweder mit der anderen bedingten Wahrscheinlichkeit $P(B_1/A_0)$ oder mit der („unbedingten“) Wahrscheinlichkeit $P(B_1)$, verheiratet zu sein, gleichgültig wie alt man ist, vergleichen. Wegen der tautologischen Beziehung $P(B_1) = P(B_1/A_1) P(A_1) + P(B_1/A_0) P(A_0)$ gilt für diese Vergleiche: zwischen den beiden bedingten Wahrscheinlichkeiten besteht eine Differenz genau dann, wenn eine der bedingten Wahrscheinlichkeiten (und damit auch die andere) sich von der unbedingten unterscheidet.

jungen und verheirateten Personen etc. zu deduzieren. Mit anderen Worten: folgt aus den beiden Aussagen (II) und (III) etwas für die Aussage (IV); genauer: ist es möglich eine Deduktionsfunktion zu formulieren, die r_4 in Abhängigkeit von r_2 und r_3 angibt? Da sich die Referenzklasse, auf die in (IV) Bezug genommen wird, als Durchschnitt der Referenzklassen ergibt, auf die sich (II) und (III) beziehen, ist dies das *Problem der „Konjunktivität“* von W-Implikationen.

Beginnen wir mit dem zweiten Problem.

2.1 Die Nicht-Konjunktivität von W-Implikationen

Das hier angesprochene Problem ist in der wissenschaftstheoretischen Literatur ausführlich unter dem Stichwort „*Statistischer Syllogismus*“ behandelt worden⁴. Unter letzterem versteht man ein „Argument“, in dem als mindestens eine der Prämissen eine statistische Aussage verwandt wird, welches aber sonst dem klassischen Syllogismus sehr nahe kommt. Ein Beispiel wäre ein „Argument“, in welchem die Aussage „Andersen ist nicht römisch-katholisch“ aufgrund der Prämissen „Andersen ist Schwede“ und „Weniger als 2% der Schweden sind römisch-katholisch“ behauptet wird. In symbolischer Schreibweise würde der „Syllogismus“ lauten⁵:

- (1) $a \in S$ (Andersen gehört zur Klasse der Schweden)
- (2) $P(R/S) < .02$ bzw. (da R und $\sim R$ eine Zerlegung von S bilden) $P(\sim R/S) > .98$
- (3) Es gilt mit einer Wahrscheinlichkeit $r > .98$ (es ist beinahe sicher), daß $a \in \sim R$

4 Klassische Darstellungen dieser Problematik hat CARL G. HEMPEL (1960, 1962, 1965: Kap. 12 §3) gegeben, der auch die relevante Literatur im einzelnen anführt und diskutiert. Die HEMPELSchen Erörterungen bilden die Grundlage für WOLFGANG STEGMÜLLER (1969: Kap. IX). Wer die Arbeiten von HEMPEL und STEGMÜLLER kennt, weiß, wie stark sich unsere Ausführungen an diese anlehnen. Deshalb sei auf sie an dieser Stelle ein für allemal verwiesen.

5 Zu diesem „Syllogismus“ ist dreierlei zu bemerken: Erstens: Die Wahrscheinlichkeitsaussage (2) läßt sich als Aussage über Häufigkeiten konstruieren. Zweitens: Hingegen ist bei Aussage (3) nicht ohne weiteres eine Häufigkeitsinterpretation möglich, da einem einzigen Fall ein Wahrscheinlichkeitswert zugeschrieben wird. Eine alternative Auffassungsmöglichkeit bestände darin, nicht dem Fall, sondern der Aussage (3) einen Wahrscheinlichkeitswert zuzuschreiben. Drittens: Gleichgültig, wie man die Wahrscheinlichkeit interpretiert, welche in der „Konklusion“ auftritt, man benötigt eine Regel, welche angibt, zu welchem bestimmten Wert man für sie kommt. Im Beispiel wurde der gesuchte Wert einfach gleich dem gesetzt, welcher in Aussage (2) auftritt; aber das muß nicht notwendig so sein.

Nehmen wir an, wir erhalten über das Objekt „Andersen“ eine weitere Information: Andersen pilgerte nach Lourdes ($a \in L$); gleichzeitig sei uns folgende statistische Gesetzmäßigkeit bekannt: Weniger als 2% der Lourdes-Pilger sind nicht römisch-katholisch, d.h. mehr als 98% sind römisch-katholisch. ($P(R/L) > .98$). Unter Verwendung dieser Informationen ist es möglich, nach dem gleichen Schema einen zweiten „Syllogismus“ zu konstruieren, welcher den ersten paralyisiert:

- (1') $a \in L$
- (2') $P(R/L) > .98$
- (3') Es ist beinahe sicher, daß $a \in R$

Aufgrund von als gültig vorausgesetzten Prämissen ist es möglich, zwei Aussagen „Andersen ist mit großer Wahrscheinlichkeit nicht römisch-katholisch“ und „Andersen ist mit großer Wahrscheinlichkeit römisch-katholisch“ zu behaupten, welche sich widersprechen – wenn die dargestellten statistischen Syllogismen ihrer Form nach als gültig legitimiert werden können.

Die Ursache des Widerspruchs können wir in folgendem Sachverhalt lokalisieren: Für das Objekt a gibt es verschiedene Referenzklassen S und L , zu denen es gleichzeitig gehören kann. Und für beide Referenzklassen gibt es bezüglich der zur Diskussion stehenden Eigenschaft R unterschiedliche statistische Regelmäßigkeiten. Da diese Situation allgemein anzutreffen ist, kann man nach obigem Schema beliebig viele statistische Syllogismen konstruieren, die im Widerspruch zueinander stehen. Deshalb ein Beispiel aus dem Bereich der empirischen Sozialforschung: „Karl Müller ist römisch-katholisch“ und „Mehr als 61% der Katholiken wählen in Bundestagswahlen CDU“; hieraus würde mit einer Wahrscheinlichkeit von größer .61 folgen, daß Karl Müller CDU wählt. Weiterhin gelte: „Karl Müller ist Arbeiter“ und für diese Referenzklasse wiederum die Regelmäßigkeit, daß weniger als 36% CDU wählen. Hieraus würde dann mit einer Wahrscheinlichkeit von weniger als .36 folgen, daß Karl Müller CDU wählt. Wenn die Form des statistischen Syllogismus ein gültiges Schlußverfahren abgäbe, würden sich wiederum zwei unverträgliche Konklusionen ergeben – die Aussage „Karl Müller wählt CDU“ gilt zugleich mit einer Wahrscheinlichkeit von weniger als .36 und von mehr als .61 – , und dies obwohl sämtliche Prämissen der rivalisierenden Syllogismen als wahr vorausgesetzt werden können.

Im deterministischen Fall ist eine solche Situation nicht möglich. Zwar kann man sich vorstellen, daß einerseits die Aussage „a ist G“ ableitbar ist aus den Prämissen „alle F sind G“ und „a ist F“ und andererseits die Aussage „a ist kein G“ ableitbar ist aus den Prämissen „alle F sind nicht G“ sowie „a ist ein F“. Aber: Im deterministischen Fall können diese vier Prämissen nicht alle gleichzeitig wahr sein, denn da sich aus ihnen widersprüchliche Aussagen ableiten lassen, sind sie selbst inkonsistent. Für den deterministischen Fall gilt vielmehr: Wenn ein Objekt ein Merkmal F besitzt, das stets mit G verknüpft ist, kann es nicht gleichzeitig ein Merkmal F' besitzen, welches stets mit Non-G verknüpft ist.

Im statistischen Syllogismus können hingegen — wie gezeigt — sich widersprechende Aussagen „abgeleitet“ werden, obwohl die Prämissen wahr sind. Diese Widersprüche können jedoch zum Verschwinden gebracht werden, wenn man den Syllogismus in anderer Weise konstruiert, nämlich annimmt, daß solche Formeln wie „es ist beinahe sicher, daß . . .“, „es ist sehr wahrscheinlich, daß . . .“ oder „es gilt mit Wahrscheinlichkeit . . .“ nicht die Konklusion als solche qualifizieren, sondern eine *Relation zwischen Prämissen und Konklusion* beschreiben. Eine solche Annahme führt auf die Konzeption einer *induktiven Wahrscheinlichkeit*, wie sie insbesondere von RUDOLF CARNAP (1950, 1959) gearbeitet wurde. Dieser entsprechend wäre unter Wahrscheinlichkeit oder Gewißheit kein modaler Qualifikator einer Aussage zu verstehen, sondern eine Relation zwischen Aussagen, die den *Grad der Bestätigung einer Aussage (Hypothese H) durch eine Menge anderer Aussagen (Evidenzen E)* zum Ausdruck bringt.

Gemäß dieser Interpretation wären unsere Beispiele in folgender Weise zu rekonstruieren: Die Prämissen (1) und (2) verleihen praktische Gewißheit der Aussage (3) „Andersen ist nicht r.k.“. Dieses Resultat des Syllogismus steht jedoch nicht mehr im Widerspruch zu der Behauptung, daß die ganz anders gearteten Prämissen (1') und (2') praktische Gewißheit der Aussage (3') „Andersen ist r.k.“ verleihen. Das 2. Beispiel wäre entsprechend zu verstehen: Die Prämissen $K.M. \in R$ und $P(CDU/R) > .61$ verleihen der Aussage „Karl Müller wählt CDU“ einen Bestätigungsgrad (eine induktive Wahrscheinlichkeit) von $s > .61$; hingegen verleihen die Prämissen $K.M. \in A$ und $P(CDU/A) < .36$ der gleichen Aussage nur einen Bestätigungsgrad von $s < .36$.

Bei dieser Konstruktion des statistischen Syllogismus unter Verwendung des Konzeptes der induktiven Wahrscheinlichkeit sind die Widersprüche verschwunden, denn es besteht kein Widerspruch darin, daß sich aus ganz verschiedenen Prämissen (Evidenzen) für die gleiche Aussage (Hypothese)

unterschiedliche Grade der induktiven Bestätigung ergeben. Doch wenn auch das logische Problem möglicher Widersprüche beseitigt wurde, so tritt an seine Stelle nun das *methodologische Problem der Anwendung* verschiedener rivalisierender Syllogismen. Wenn sich der Bestätigungsgrad einer Hypothese bei Hinzuziehen von weiterem Wissen ändert bzw. die Wahrscheinlichkeit eines Ereignisses in Abhängigkeit davon variiert, in welche Bezugsklassen es eingeordnet wird, erhebt sich die Frage, welches Wissen im Einzelfall hinzugezogen werden bzw. welche der vielen möglichen Referenzklassen Berücksichtigung finden soll. Sollen wir im ersten Beispiel bei der Prognose der Religionszugehörigkeit von Andersen unser Wissen über seine nationale Identität oder über die Tatsache, daß er eine Pilgerfahrt nach Lourdes machte, verwenden? Oder sollen wir im zweiten Beispiel zur „Erklärung“ des Wahlverhaltens aufgrund demographischer Merkmale Karl Müller in die Bezugsklasse der Katholiken oder in die Klasse der Arbeiter einordnen? Da wir darüber hinaus noch sehr viel mehr über Andersen und Karl Müller wissen werden bzw. sie in viel mehr möglicherweise „relevante“ Referenzklassen einordnen können, besteht eine gewisse Willkürlichkeit in der Verwendung speziellen Wissens zur Stützung von Hypothesen bzw. zur Einordnung in spezielle Referenzklassen. Dies wird umso problematischer, je mehr sich die Konsequenzen voneinander unterscheiden, ja widersprechen. Wenn es mehrere rivalisierende Argumente in Form von statistischen Syllogismen gibt, auf welche der verschiedenen induktiven Relationen soll man sich bei der Formulierung theoretischer Erwartungen und bei dem Fällen praktischer Entscheidungen stützen?

Ein erster Vorschlag zur Lösung des Dilemmas der Anwendung statistischer Systematisierungen stammt von RUDOLF CARNAP (1950) und wurde formuliert als Forderung nach *Berücksichtigung des Gesamtdatums*: Bei der Anwendung statistischer Systematisierungen muß alles Wissen mit Ausnahme des irrelevanten berücksichtigt werden. Dabei ist ein Wissen für eine Hypothese dann irrelevant, wenn seine Hinzuziehung zu den als „Prämissen“ verwandten Evidenzen den Bestätigungsgrad der Hypothesen unverändert läßt. Zwar kann entsprechend dem CARNAPschen Vorschlag in den Fällen, in denen ein Anwendungsdilemma besteht, gezeigt werden, daß die Forderung des Gesamtdatums nicht erfüllt sein kann⁶,

6 Man betrachte dazu z.B. die induktive Bestätigung der Aussage „K.M. wählt CDU“ durch die Evidenzen

aber umgekehrt ist die Regel nicht praktikabel, da man nicht weiß, wie im Einzelfall die Forderung zu erfüllen ist. Im übrigen gilt, daß bei strikten Allaussagen alles weitere Wissen, das nicht in den Wenn-Komponenten formuliert wird, irrelevant ist. Das heißt: Im deterministischen Fall ist die CARNAPsche Forderung des Gesamtdatums automatisch immer erfüllt.

Von CARL G. HEMPEL (1962, 1965) wurde zur Lösung des Anwendungsdilemmas daher das *Prinzip der maximalen Bestimmtheit* vorgeschlagen. Ohne eine präzise Definition dieses Prinzips zu geben, können wir es folgendermaßen umschreiben: Wenn eine statistische Regelmäßigkeit $P(G/F) = r$ auf ein Objekt angewandt werden soll, das die Eigenschaft F aufweist (zur Referenzklasse F gehört) und wenn aus den Prämissen oder dem akzeptierten Wissen folgt, daß das Objekt auch zu F' gehört, dann muß dieses Wissen Berücksichtigung finden.

Falls die beiden Referenzklassen im Verhältnis der Inklusion zueinander stehen, dann muß eine statistische Regelmäßigkeit verwandt werden, die sich auf die engere Klasse bezieht: „Schärfere Information verdrängt stets die schwächere (bei der Anwendung statistischer Gesetzmäßigkeiten)“.

Falls das betreffende Objekt zu verschiedenen Referenzklassen F und F' gehört, die nicht im Verhältnis der Inklusion zueinander stehen, dann folgt daraus, daß es zum Durchschnitt von F und F' gehört (d.h. Eigenschaft F und Eigenschaft F' aufweist). Diese Folgerung ist aber wiederum eine schärfere Information, da der Durchschnitt von F und F' sowohl in F als auch in F' enthalten ist. Mit anderen Worten: es reicht nicht aus, statistische Regelmäßigkeiten für die Bezugsklassen F und F' zu verwenden, sondern es muß ein statistisches Gesetz für die aus dem Durchschnitt von F und F' bestehende Referenzklasse gesucht werden. *Und solange ein solches Gesetz nicht gefunden ist, müssen wir es unentschieden lassen, welcher der möglichen Syllogismen korrekt ist* bzw. ob man überhaupt ein korrektes Argument konstruieren kann. Für unsere Beispiele heißt dies: Da Andersen sowohl zur Klasse der Schweden als auch zur Klasse der Lourdes-Pilger gehört, muß für ein korrektes Argument eine statistische Regelmäßigkeit verwandt werden, die die Wahrscheinlichkeit, rö-

misch-katholisch zu sein, für die Referenzklasse derjenigen Personen spezifiziert, welche Schweden und Lourdespilger sind. Es mag zwar plausibel erscheinen, daß die Wahrscheinlichkeit katholisch zu sein sich für die schwedischen Lourdespilger kaum von der Wahrscheinlichkeit für alle Lourdespilger unterscheidet, aber logisch zwingend ist dies nicht. Möglich sind ganz andere Werte, beispielsweise auch, daß die Wahrscheinlichkeit für den Durchschnitt gleich derjenigen für die Klasse aller Schweden ist. Welchen möglichen Wert die Wahrscheinlichkeit in der Referenzklasse des Durchschnitts annimmt, ist ein empirisches Problem und nicht deduzierbar aus den Wahrscheinlichkeiten für die Klassen, deren Durchschnitt betrachtet wird.

In gleicher Weise gilt in unserem zweiten Beispiel, daß Karl Müller zum Durchschnitt der Klassen der Arbeiter und Katholiken gehört. Wenn also ein korrekter Syllogismus konstruiert werden soll, der eine Aussage über die CDU-Wahl von Karl Müller induktiv stützen soll, dann muß eine statistische Gesetzmäßigkeit der Form $P(CDU/A \wedge R) = r_3$ gesucht werden, welche wiederum logisch unabhängig ist von den beiden Gesetzmäßigkeiten $P(CDU/A) = r_1$ und $P(CDU/R) = r_2$. Solange wir nur die Wahrscheinlichkeiten des Wählens der CDU für Arbeiter bzw. Katholiken kennen, können wir nichts über den Grad der induktiven Bestätigung einer Aussage machen, in der behauptet wird, daß ein katholischer Arbeiter CDU wählen wird.

Fassen wir zusammen:

Gegeben sei eine Menge von W-Implikationen, in denen beschrieben wird, daß ein bestimmtes Attribut B in probabilistischer Weise von einer Reihe von Attributen A_i abhängig ist: $P(B/A_i) = r_i$. *Aus der Konjunktion* dieser elementaren statistischen Aussagen *folgt jedoch nichts* für eine entsprechende Aussage $P(B/A) = r$, in der die Referenzklasse A durch die Konjunktion der Eigenschaften A_i (als Durchschnitt der Referenzklassen A_i) definiert ist. *Eine solche Aussage muß vielmehr als zusätzliche empirische Aussage formuliert und überprüft werden*⁷.

Bis jetzt haben wir die Frage der Konjunktivität nur in Bezug auf die Wenn-Komponenten bzw. Referenzklassen behandelt. Die erwähnte Nicht-

7 Wir betonen nochmals, daß die Situation hier völlig anders ist als im deterministischen Fall: Aus der Konjunktion einer Menge strikter All-Aussagen mit gleicher Dann-Komponente folgt immer eine entsprechende Aussage, in der die komplexe Wenn-Komponente Konjunktion der Wenn-Komponenten der vorausgesetzten Aussagen ist.

„K.M. $\in R$ “ und „K.M. $\in A$ “. Beide Evidenzen können nicht gleich dem Gesamtdatum D sein, denn wäre es so, dann gälte
 $1 = P(\sim CDU/D) + P(CDU/D) > 0.64 + 0.61 = 1.25$,
 was einen arithmetischen Widerspruch darstellt.

Konjunktivität von W-Aussagen gilt aber auch für die Ereignisse, und dies wiederum im Gegensatz zur Konjunktivität von deterministischen Aussagen für die Dann-Komponente. Beispielsweise folgt, wie in Abschnitt I erwähnt wurde, aus der Konjunktion der Aussagen „Alle A sind B₁“ und „Alle A sind B₂“ immer die Aussage „Alle A sind B₁ und B₂“ und umgekehrt. Daß dies bei probabilistischen Aussagen nicht der Fall ist, kann man sich leicht an folgendem Würfelbeispiel klarmachen. Die Wahrscheinlichkeit, nicht eine 1 zu würfeln, betrage bei einem normalen Würfel 5/6; die Wahrscheinlichkeit, nicht eine 2 zu würfeln, ebenfalls 5/6 usw. . . Die Wahrscheinlichkeit des komplexen Ereignisses, nicht eine 1 und nicht eine 2 . . . und nicht eine 5 zu würfeln, beläuft sich hingegen, da dies die Wahrscheinlichkeit ist, gerade eine 6 zu würfeln, auf 1/6. *Die Konjunktion von 5 Ereignissen, die alle relativ wahrscheinlich sind, ist ein Ereignis, welches relativ unwahrscheinlich ist.*

2.2 Die Nicht-Transitivität von W-Implikationen

Kommen wir nun zu dem ersten, zu Beginn dieses Abschnitts skizzierten Problem. Wir hatten die Frage gestellt, ob es möglich sei, aus zwei bekannten W-Implikationen der Art $P(B/A) = r_1$ bzw. $P(C/B) = r_2$ Konsequenzen für eine W-Implikation zu formulieren, in der etwas über die Wahrscheinlichkeit des Attributs C unter der Bedingung A ausgesagt wird; spezifischer: ist es möglich, eine Deduktionsfunktion aufzustellen, die r_3 in Abhängigkeit von r_1 und r_2 beschreibt, wobei gilt: $r_3 = P(C/A)$?

Es wurde schon im Abschnitt I darauf hingewiesen, daß man D-Implikationen zwischen Attributen darstellen kann mit Hilfe der Inklusion von Klassen oder Mengen, wobei die betreffenden Klassen gerade als die Extensionen der fraglichen Attribute definierbar sind. Bezeichnen wir mit A das Attribut „ . . . ist Kölner“, mit B „ . . . ist Rheinländer“ und mit C „ . . . ist Deutscher“, so ist die D-Implikation „Alle Kölner sind Rheinländer“ übersetzbar in die Aussage „Die Menge der Personen, die A sind, ist enthalten in der Menge, die B sind“ usw. . . Hier gilt nun bekanntlich die Transitivität, d.h. wenn alle A in B und alle B in C sind, dann sind auch alle A in C; mit anderen Worten: aus den beiden D-Implikationen, daß alle Kölner Rheinländer und alle Rheinländer Deutsche sind, folgt die D-Implikation, daß alle Kölner Deutsche sind.

Will man analog hierzu Beziehungen zwischen

W-Implikationen ebenfalls durch Relationen zwischen Mengen darstellen, wobei die verwandten Mengen wiederum als Extensionen der zur Diskussion stehenden Attribute zu konstruieren sind, dann ist es erforderlich, für die auftretenden Mengen ein Maß zu definieren. Das folgende Diagramm ist die graphische

Darstellung der Beziehungen zwischen Mengen, in denen einige (nicht alle!) Elemente aus A auch B bzw. einige aus B auch C sind. Diese Beziehungen sollen entsprechende W-Implikationen zwischen den Attributen A, B und C zum Ausdruck bringen. Dabei bedeutet, daß eine von Null verschiedene Wahrscheinlichkeit dafür besteht, daß ein Objekt mit dem Attribut A auch das Attribut B besitzt, in dieser Übersetzung, daß der Durchschnitt der Mengen A und B (in der Zeichnung schraffiert) nicht leer ist. Formal kann man Wahrscheinlichkeit unter Verwendung von Maßfunktionen für Mengen in folgender Weise definieren⁸:

$$P(B/A) = \frac{m(A \cap B)}{m(A)}$$

Falls man mit endlichen Mengen arbeitet, ist es möglich, als spezielle Maßfunktion die Kardinalzahl (Anzahl der Elemente einer Menge) zu verwenden. Man erhält dann die schon angeführte Interpretation der Wahrscheinlichkeit als relative Häufigkeit.

Aus dem durch das Diagramm dargestellten speziellen Beispiel ist die Nichttransitivität der W-Implikation ersichtlich: Obwohl die Durchschnitte von A und B und von B und C jeweils nicht leer sind, ist dies für A und C der Fall: von Null verschiedene Wahrscheinlichkeiten, B zu besitzen, wenn man A hat, bzw. C zu besitzen, wenn man B hat, sind ohne weiteres damit vereinbar, daß die Wahrscheinlichkeit, auch C zu besitzen, für die Objekte mit A gleich Null ist.

Da im Falle probabilistischer Aussagen immer schon ein gewisses Ausmaß an Wahrscheinlichkeits-

8 Wie man sieht, kann bei einer Definition der Wahrscheinlichkeit durch das Maß von Mengen die D-Implikation als Spezialfall der W-Implikation aufgefaßt werden: Wenn nicht einige, sondern alle A auch B sind, ist ganz A in B enthalten, also der Durchschnitt von A und B gleich A. Damit ergibt sich für jede beliebige Maßfunktion $P(B/A) = 1$.

theoretischer Terminologie verwandt wird, sollte man sich fragen, ob es zur Lösung unseres Problems der Deduktion sinnvoll ist, ganz bewußt die Problematik in einen wahrscheinlichkeitstheoretischen Kontext einzubetten, um zu sehen, ob dieses Vorgehen zur Klärung und Lösung bestimmter Teilprobleme weiterhilft.

Wahrscheinlichkeitstheorie, so wie sie heute in der Mathematik aufgebaut wird, ist ein formaler, d.h. uninterpretierter Kalkül, bestehend aus einigen uninterpretierten Grundzeichen und drei abstrakten Aussagen, die den Status von Axiomen haben. Unter Verwendung dieser Axiome und geeignet definierter neuer Zeichen sind in der W-Theorie weitere abstrakte Aussagen beweisbar. Falls es möglich sein sollte, bestimmten Satzformeln des Kalküls bestimmte Aussagen zuzuordnen, können aus den im Rahmen des Kalküls abgeleiteten Satzformeln wiederum im Wege der Zuordnung bestimmte Aussagen gewonnen werden.

Die Grundzeichen der W-Theorie und die uns in unserem Zusammenhang interessierenden Zuordnungen sind:

- 1) Eine nichtleere Menge, der sog. *Stichprobenraum* Ω , interpretierbar als Menge aller betrachteten Einheiten (universe of discourse).
- 2) Eine Menge α von Teilmengen $A_i \subset \Omega$, der sog. *Ereignisraum*. Die Elemente A_i des Ereignisraums heißen Ereignisse. Diesen Ereignissen sind Attribute zuzuordnen, da Attribute extensional aufgefaßt werden können, was für unsere Zwecke ausreicht.
- 3) Im Ereignisraum sind die *mengentheoretischen Operationen* des Durchschnitts, des Komplements und der Vereinigung definiert, wobei gilt: der Ereignisraum ist abgeschlossen gegenüber Vereinigungsmengen- und Komplementbildung (er ist eine sog. Algebra⁹).
- 4) Auf α ist ein *W-Maß* definiert, d.h. eine positive, normierte und additive reellwertige Funktion, die jedem Ereignis eine reelle Zahl zwischen Null und 1 zuweist. Dies drücken wir auch so aus, daß wir sagen, die Funktion P erfüllt die folgenden drei W-Axiome:

- a) $A_i \in \alpha: 0 \leq P(A_i) \leq 1$
- b) $P(\Omega) = 1$
- c) $A_i \cap A_j = \emptyset \Rightarrow P(A_i \vee A_j) = P(A_i) + P(A_j)$

⁹ Wegen der Abgeschlossenheit gehören auch der ganze Stichprobenraum (das „sichere“ Ereignis) und die leere Menge (das „unmögliche“ Ereignis) zum Ereignisraum.

Wie oben interpretieren wir für unsere Zwecke das Maß P mit Hilfe der relativen Häufigkeiten, indem wir setzen:

$$A_i \in \alpha: P(A_i) = \frac{|A_i|}{|\Omega|}$$

Man sieht leicht, daß das unter Verwendung der Kardinalzahlen definierte Maß P tatsächlich die drei W-Axiome erfüllt.

Wir können also sagen, daß wir, wenn wir in der angegebenen Weise mit einem bestimmten (endlichen) Objektbereich, beliebigen Attributen und relativen Häufigkeiten arbeiten, eine Interpretation der W-theoretischen Begriffe Stichprobenraum, Ereignisraum und W-Maß angeben haben, bei der die Axiome der W-Theorie logisch wahr sind, d.h. die Axiome sind notwendigerweise wahr und nicht aufgrund irgendwelcher empirischen Gesetzmäßigkeiten, denen die Objekte gehorchen, auf die bei der Deutung Bezug genommen wird¹⁰.

Insbesondere ist also die Wahrheit der Axiome nicht davon abhängig, welche speziellen Objekte man zugrundelegt bzw. welche speziellen Attribute man betrachtet. Das bedeutet natürlich, daß aus der Wahrheit der Axiome keine empirisch gehaltenen Konsequenzen gezogen werden können.

Bis jetzt haben wir nur etwas genauer das reformuliert, was wir schon zu Beginn dieses Abschnittes ausführten. Zugleich wissen wir, daß wir keine speziellen empirischen Annahmen gemacht haben und daß wir deshalb auch keine neuen empirischen Aussagen erwarten können. Wohl aber wird es möglich sein, genauer die Struktur der Zusammenhänge von W-Implikationen zu analysieren, wie sie notwendigerweise von allen Objekten und Attributen erfüllt sein müssen, bzw. anzugeben, welche Aussagen wir behaupten müssen bzw. nicht behaupten können, wenn wir konsistent argumentieren wollen¹¹.

10 Das Erfülltsein der drei W-theoretischen Axiome ist nämlich einfach Folge der Definition von Häufigkeiten und damit Folge der Definition der natürlichen Zahlen bzw. der Operation des Zählens.

11 Es sei darauf hingewiesen, daß man statt mit dem Ereignisraum und einem auf ihm definierten W-Maß auch mit sog. *Indikatorvariablen* (häufig als „dummy variables“ bezeichnet) und entsprechenden Verteilungsfunktionen arbeiten kann. Dabei ist die Indikatorvariable eines Ereignisses A_i eine Abbildung aus dem Stichprobenraum in die Menge $\{0,1\}$, die für ein Element den Wert 1 genau dann annimmt, wenn es zu A_i gehört. Jedem Ereignis korrespondiert genau

Kommen wir nun zur *Darstellung der Struktur von Attributen mit Hilfe der W-Theorie*. Wir können sämtliche möglichen Verwendungen dieses speziellen Kalküls nicht diskutieren, sondern nur die Richtung angeben, in denen sich solche Anwendungsmöglichkeiten bewegen werden. Dabei soll uns nur der einfachste Fall interessieren, nämlich der von drei Attributen, wobei wir an das eingangs berichtete Beispiel von DURKHEIM über den Zusammenhang von Alter, Familienstand und Selbstmord denken können. Oben wurde die allgemeinste Struktur eines *Drei-Attribute-Falls* graphisch repräsentiert. Diese Struktur ist durch folgende von Null verschiedenen bedingten Wahrscheinlichkeiten gekennzeichnet¹²:

$$\begin{array}{ll} P(B_1/A_1) & P(B_1/A_0) \\ P(C_1/B_1) & P(C_1/B_0) \\ P(C_1/A_1 \wedge B_1) & P(C_1/A_1 \wedge B_0) \end{array}$$

Da die Axiome der W-Theorie erfüllt sind, ist es möglich, weitere Aussagen über Wahrscheinlichkeiten behaupten, und zwar alle, die aus den Axiomen der W-Theorie logisch deduzierbar sind. Z.B. gilt gemäß der W-Theorie in unserem Falle auch folgender Spezialfall des sog. *Theorems der totalen Wahrscheinlichkeit*

$$\begin{aligned} \text{(Th 1)} \quad P(C_1/A_1) &= P(C_1/A_1 \wedge B_1) \cdot P(B_1/A_1) + \\ &P(C_1/A_1 \wedge B_0) \cdot P(B_0/A_1) \end{aligned}$$

Unter Verwendung dieses Theorems – welches im Rahmen der W-Theorie, um es nochmals zu wiederholen, als analytische Identität notwendigerweise immer dann erfüllt ist, wenn auch die Axiome der W-Theorie erfüllt sind – kann man eine *qualitative Analogie zum Simon-Blalock-Verfahren* konstruieren¹³.

eine Indikatorvariable; und da wir andererseits bei unserer Interpretation der W-Theorie den Ereignissen Attribute zuordnen, entspricht jedem Attribut in eindeutiger Weise eine Indikatorvariable. Wir können also über W-Implikationen in der Form reden, daß wir Aussagen für Indikatorvariablen formulieren, was auf die von PAUL F. LAZARSFELD (1961, 1968) ausgearbeitete *Algebra dichotomer Systeme* führt, einen allgemeinen Kalkül, der logisch notwendige Zusammenhänge zwischen Relationen einer Menge dichotomer Merkmale beschreibt. (Zur Äquivalenz von Indikatorvariablen und Ereigniswahrscheinlichkeiten vgl. auch R. BOUDON 1968, insbes. § 6.6)

12 Gemäß Anm. 3 werden auch Wahrscheinlichkeiten unter den Bedingungen A_0 und B_0 aufgeführt; weiterhin ist die Liste noch um $P(C_1/A_0 \wedge B_1)$ und $P(C_1/A_0 \wedge B_0)$ zu erweitern.

13 Das SIMON-BLALOCK-Verfahren erlaubt es, unter bestimmten Annahmen nicht-zirkuläre Wirkungsstruk-

Nehmen wir den folgenden *ersten Spezialfall*, in dem B eine intervenierende Variable für den Zusammenhang zwischen A und C ist, und A *nur über B* auf C „wirkt“, d.h. *keine direkten Beziehungen* von A auf C vorliegen. Wie wird nun die Aussage, daß A allein deshalb auf C wirkt, weil die Variable B eine solche Wirkung vermittelt – was man im Graphen (s.u. S. 128, Fall (a)) durch das Wegfallen eines direkten Pfeils von A nach C andeuten kann – in unserem Kalkül formuliert, d.h. wie wird eine solche Aussage „formalisiert“?

Wenn die Beziehung zwischen A und C ausschließlich Resultat der Vermittlung durch B ist, bedeutet dies, daß bei „Konstanz“ von B die Beziehung „wegfällt“; d.h. betrachtet man nur die Objekte mit dem Attribut B_1 oder nur die Objekte mit dem Attribut B_0 , dann sind die Attribute A_i und C_j jeweils unabhängig voneinander: „In den Ereignissen B_1 und B_0 sind die Ereignisse A und C unabhängig“. Stochastische Unabhängigkeit von Ereignissen wird bekanntlich durch die Gültigkeit des Multiplikationssatzes ausgedrückt, den man für beliebige vorausgesetzte Bedingungen formulieren kann, also auch für die beiden Bedingungen B_1 und B_0 . Damit gelten in diesem Spezialfall die folgenden zusätzlichen Aussagen:

$$\begin{aligned} (1) \quad &P(C_1 \wedge A_1 | B_1) = P(C_1 | B_1) \cdot P(A_1 | B_1) \\ (2) \quad &P(C_1 \wedge A_1 | B_0) = P(C_1 | B_0) \cdot P(A_1 | B_0) \end{aligned}$$

Weiterhin ist die Wahrscheinlichkeit von C_1 unter der Bedingung A_1 und B_1 definiert als:

$$(3) \quad P(C_1 | A_1 \wedge B_1) = \frac{P(C_1 \wedge A_1 | B_1)}{P(A_1 | B_1)}$$

Unter Verwendung von (1) folgt hieraus

turen für metrische Variablen anhand von Koeffizienten der linearen Korrelation zu testen und u.U. bestimmte logisch mögliche Strukturen als mit den empirischen Daten nicht vereinbar zu eliminieren. Eine erste Einführung findet man bei HUBERT M. BLALOCK (1968); klassisch sind die Arbeiten von HERBERT SIMON (1954) sowie BLALOCK (1961 a, b; 1962). Während sich BLALOCK zunächst damit begnügte, qualitative Aussagen über die Vereinbarkeit von Annahmen über Wirkungsstrukturen mit Daten zu formulieren, werden in der unten in Abschnitt II.3 erwähnten Dependenz- oder Pfadanalyse aus den Korrelationskoeffizienten numerische Werte berechnet, die als Maß direkter Einflüsse der Variablen interpretierbar sind. (Siehe auch Anm. 19ff. und die dort zitierte Literatur).

$$(4) P(C_1 | A_1 \wedge B_1) = P(C_1 | B_1)$$

Analog gilt für die Wahrscheinlichkeit von C_1 unter der Bedingung A_1 und B_0 im Falle der stochastischen Unabhängigkeit von A_1 und C_1 in B_0

$$(5) P(C_1 | A_1 \wedge B_0) = P(C_1 | B_0)$$

Setzt man in Theorem 1 die Aussagen (4) und (5) ein, erhält man schließlich

$$P(C_1 | A_1) = P(C_1 | B_1) \cdot P(B_1 | A_1) + P(C_1 | B_0) \cdot P(B_0 | A_1)$$

bzw. bei geeigneter Umschreibung, die die Generalisierungsmöglichkeit für beliebige Zerlegungen des Stichprobenraums besser zum Ausdruck bringt:

$$(Th\ 2) \quad P(C_1 | A_1) = \sum_{k=0,1} P(C_1 | B_k) \cdot P(B_k | A_1)$$

Dieses Theorem ist die gesuchte D-Funktion, die r_3 in Abhängigkeit von r_1 und r_2 darstellt¹⁴. Mit (Th 2) ist uns eine *Lösung des Transitivitätsproblems im Drei-Attribute-Fall* mit einer intervenierenden Variablen und Abwesenheit von direkten Effekten gegeben. Wenn die bedingten Wahrscheinlichkeiten bekannt sind, die den stochastischen Zusammenhang zwischen A und B und zwischen B und C charakterisieren, dann kann man in dieser speziellen Wirkungsstruktur den numerischen Wert der bedingten Wahrscheinlichkeit deduzieren, durch die der Zusammenhang zwischen A und C charakterisiert wird.

Generalisieren wir (Th 2) für beliebige Polytomien. Polytomien sind dadurch definiert, daß sie eine Klasseneinteilung sämtlicher Objekte (eine Zerlegung des Stichprobenraums) liefern. Wenn beispielsweise B eine Polytomie mit K Ausprägungen ist, gilt:

$$\Omega = \bigcup_{k=1}^K B_k$$

Betrachten wir Zusammenhänge zwischen den Polytomien A und C, so können diese durch eine Menge von bedingten Wahrscheinlichkeiten charakterisiert werden, welche die Wahrscheinlichkeit des Ereignisses C_j unter der Bedingung A_i angeben. In unserer speziellen Wirkungsstruktur

tur gelten für alle diese bedingten Wahrscheinlichkeiten:

$$P(C_j | A_i) = \sum_{k=1}^K P(C_j | B_k) \cdot P(B_k | A_i)$$

oder in einer bekannteren Schreibweise durch „Übergangswahrscheinlichkeiten“

$$p_{ij} = \sum_{k=1}^K p_{ik} \cdot p_{kj}$$

Faßt man die Übergangswahrscheinlichkeiten, welche nichts anderes als unsere bedingten Wahrscheinlichkeiten sind, zu Matrizen zusammen, lautet (Th 2) in Form eines Matrizenproduktes¹⁵:

$$(Th\ 2') \quad P_{AC} = P_{AB} \cdot P_{BC}$$

Betrachten wir ein numerisches *Beispiel aus der Mobilitätsforschung* (G. CARLSSON 1969: 132ff.). Untersucht werden soll der Zusammenhang zwischen sozialer Herkunft, Ausbildung und erreichter Berufsposition. Dabei soll es sich bei allen drei Variablen um Trichotomien handeln. Ziel sei es, die Hypothese zu testen, daß die angeführte Wirkungsstruktur vorliegt, was hier bedeutet, daß die soziale Herkunft zwar in einem gewissen Ausmaß die erreichten Berufspositionen determiniert, daß dieser Zusammenhang aber allein Resultat der intervenierenden Variable Ausbildung ist. D.h. wenn bestimmte Personen besonders günstige Positionen erreichen, liegt dies daran, daß sie relativ gut ausgebildet sind, was selbst wiederum davon abhängt, daß sie aus relativ privilegierten Familien stammen. Aber: wenn man einmal eine bestimmte Ausbildung erlangt hat, ist die Wirkung der Herkunft erloschen; für Personen mit gleicher Ausbildung ist für die beruflichen Chancen die Herkunft irrelevant. Die Herkunft determiniert nur in indirekter Weise die Berufschancen, direkte Wirkungen – von CARLSSON „delayed effects“ genannt – liegen nicht vor.

Das Mittel zur Prüfung dieser Annahme besteht in der Berechnung einer fiktiven Übergangsmatrix von sozialer Herkunft zu den erreichten Positionen – einer Matrix, wie sie dann zu erwarten ist, wenn die Hypothese der Nicht-Existenz von „delayed effects“ gültig ist – und dem Vergleich der erwarteten Matrix mit der empirisch gegebenen.

15 (Th 2') bringt die Analogie des hier gewonnenen Resultats zum Fall metrischer Variablen am klarsten zum Ausdruck: Sind nämlich A, B und C von metrischer Struktur, dann erhält man nach SIMON-BLALOCK unter der Annahme der Nicht-Existenz direkter Effekte von A nach C für die Korrelationskoeffizienten die Beziehung: $r_{AC} = r_{AB} \cdot r_{BC}$.

14 Diese Aussage ist nicht ganz korrekt, da in die Berechnung von r_3 z. B. neben $r_2 = P(C_1 | B_1)$ auch die bedingte Wahrscheinlichkeit $P(C_1 | B_0)$ eingeht.

Wenn B die („Bildungs“-) Matrix ist, welche die Chancen des Zugangs zu bestimmten Ausbildungskategorien in Abhängigkeit von bestimmten Herkunftskategorien beschreibt, und wenn K die („Karriere“-) Matrix ist, welche die Chancen des Erreichens bestimmter Positionen in Abhängigkeit von der Zugehörigkeit zu einzelnen Bildungskategorien spezifiziert, ergibt sich gemäß (Th 2') die Matrix V_e der Transmission („Vererbung“) sozialer Positionen in der Generationenfolge unter der genannten Hypothese als

$$V_e = B \cdot K$$

Da K und B aufgrund der empirischen Daten bekannt sind, kann V_e mit der Matrix V der Übergangswahrscheinlichkeiten von bestimmten Herkunftskategorien in bestimmte Kategorien erreichter sozialer Positionen verglichen werden. Als Vergleichskriterium verwendet der Autor die Quote der Immobilen, d.h. den Prozentsatz der untersuchten Personen, die bezüglich der erreichten Positionen und der Herkunft in die gleichen Kategorien fallen. Weiterhin wird zum Vergleich die Immobilenquote herangezogen, die dann zu erwarten ist, wenn sämtliche Kategorien des erreichten Status und der Herkunft stochastisch unabhängig voneinander sind – also für das Erreichen eines bestimmten Status die jeweilige Herkunft irrelevant ist –, was in einem technischen Sinne als „perfekte Mobilität“ bezeichnet wird und dadurch definiert ist, daß die Matrix der Übergangswahrscheinlichkeiten von Herkunftskategorien in Kategorien erreichter Positionen zeilenweise gleich ist.

Die Ergebnisse sind in der folgenden Tabelle für zwei verschiedene Populationen Schwedens wiedergegeben.

Prozentsatz der Immobilen

Situationen	Männl. Bevölke- rung im Alter von 20-55 J. Schweden 1955	Willkürliche Auswahl Stockholm 1949
Perfekte Mobilität	45 %	37 %
Keine „delayed effects“	48 %	49 %
Tatsächl. beobachtete Situation	63 %	53 %

Zur Interpretation soll folgendes vermerkt werden: Wenn Schwedens Schichtsystem im Jahre 1955 durch „perfekte Mobilität“ gekennzeichnet wäre, wäre bei der gegebenen Struktur der Verteilung der beiden „Generationen“ auf die sozialen Kategorien eine Immobilitätsquote von 45 % zu erwarten, d.h. 55 % der Population müßte mobil sein in dem Sinne, daß sie sich in anderen sozialen Kategorien befindet als ihre Herkunftsfamilien. Daß tatsächlich die Mobilität erheblich geringer ist (Immobilenquote von 63 %) zeugt davon, daß Schweden noch weit vom Zustand „perfekter Mobilität“ (in dem oben definierten technischen und keinem anderen Sinne!) entfernt ist. Welche Rolle spielen nun die Ausbildungsinstitutionen in dieser Situation relativer Rigidität des Schichtsystems?

1. Die Differenz der Werte in der ersten und zweiten Zeile der Tabelle zeigt, daß die Ausbildungsinstitutionen bei der Weitergabe sozialer Ungleichheit eine geringe Rolle spielen. Da in der Matrix B die Ungleichheiten in den Chancen des Zuganges zu den Bildungsinstitutionen

und in K die Zusammenhänge zwischen Bildung und beruflichem Erfolg vollständig eingehen, wäre trotz dieser empirischen Zusammenhänge fast eine Situation der perfekten Mobilität zu erwarten, wenn die Herkunft nur über die Determination der Ausbildung den beruflichen Erfolg beeinflussen würde, wenn es also keine Wirkungen der Herkunft unabhängig von der Bildung gäbe.

2. Der Vergleich der Werte der zweiten und dritten Zeile gibt jedoch einen Eindruck von der Relevanz der Wirkung der Herkunft unabhängig von den durch die Ausbildung vermittelten Effekten. Gleichzeitig dürfte klar werden, wie wenig man in diesem Fall im Sinne eines Abbaus von Schichtstrukturen erreicht hätte, wenn es gelingen sollte, eine Gleichheit in den Chancen des Zugangs zu den Bildungsinstitutionen zu realisieren. Zwar wäre dann die durch die Bildung vermittelte Wirkung der Herkunft gleich Null. Die Existenz von starken „delayed effects“ der Herkunft schlägt jedoch in dem Gesamtzusammenhang zwischen Herkunft und beruflichem Erfolg in solcher Weise zu Buche, daß sich dieser Zusammenhang insgesamt auch bei Egalisierung der Bildungschancen kaum lockern würde, d.h. selbst dann würde insgesamt das Ausmaß der Transmission sozialer Positionen in der Generationenfolge fast unverändert groß bleiben.

Man vergleiche mit diesen Ergebnissen die Situation in Stockholm im Jahre 1949, worauf wir nicht mehr eingehen wollen, da uns in unserem Zusammenhang die methodologischen Probleme mehr interessieren als substantielle Fragen. Es sei nur darauf hingewiesen, daß man anhand dieser Daten Spekulationen über die langfristige Entwicklung des nationalen Schichtsystems anstellen kann, wenn der Prozeß der Urbanisierung und Industrialisierung fortschreitet.

Kommen wir abschließend noch zu zwei weiteren Spezialisierungen des Drei-Attribute-Falls, wie er in obigem Graphen dargestellt wurde. Bis jetzt hatten wir angenommen, daß das Attribut (allgemeiner die Polytomie) B den Status einer intervenierenden Variablen hat. Es ist nun auch möglich, daß zwischen A und B und zwischen B und C empirische Zusammenhänge bestehen, nicht jedoch zwischen A und C, daß aber B sowohl A als auch C in einem kausalen oder zeitlichen Sinne vorangeht. Dann hat B den Status einer antezedierenden Variablen, die zwischen A und C einen Scheinzusammenhang verursacht, analog zum Fall der „spurious correlation“ bei metrischen Variablen. Auch diese Struktur wird dadurch beschrieben, daß wir sagen, in den Ereignissen B_1 und B_0 sind die Ereignisse A und C stochastisch unabhängig voneinander. Wir kommen damit ebenfalls zu (Th 2). Ob eine Variable B den Status einer intervenierenden oder einer antezedierenden Variablen hat, ist anhand der deduzierbaren Konsequenzen nicht zu unterscheiden – genau so wenig wie im Falle der Korrelationen von metrischen Variablen. In beiden speziellen Wirkungsstrukturen gilt – und das ist

an sich das für unser Problem wichtigere Ergebnis —, daß sich gemäß (Th 2) die Werte bestimmter bedingter Wahrscheinlichkeiten aus denen anderer logisch deduzieren lassen.

In einer dritten speziellen Struktur schließlich bestehen zwischen A und B keinerlei Zusammenhänge — was nicht heißen muß, daß die bedingte Wahrscheinlichkeit von B_i falls A_i gleich Null ist, sondern nur, daß sie gleich der unbedingten Wahrscheinlichkeit von B_i ist —; A und B jedoch „wirken“ auf C. Diese Situation ist von G.L. WILBER (1967) „*konvergierender Einfluß*“ genannt worden.

In dieser Situation besteht eine *generelle* stochastische Unabhängigkeit von A und B (und nicht nur in den Ereignissen C_1 und C_0), was wiederum formalisiert wird durch

$$(6) P(A_i \wedge B_j) = P(A_i) \cdot P(B_j)$$

Eine solche Aussage — aus der übrigens die oben erwähnte Gleichheit von bedingter und unbedingter Wahrscheinlichkeit, und zwar sowohl für A als auch für B, folgt — ist empirisch testbar. Da die Situation des konvergierenden Einflusses von A und B auf C im wesentlichen durch die bedingten Wahrscheinlichkeiten $P(C_k|B_j)$ und $P(C_k|A_i)$ beschrieben wird, könnte man sich die Frage stellen, ob es möglich ist, aufgrund dieser Kenntnis etwas über die Wirkung des gemeinsamen Auftretens von A und B auf C auszusagen, also etwas über die *bedingte Wahrscheinlichkeit* $P(C_k|A_i \wedge B_j)$ zu folgern. Wie wir jedoch wissen, ist dies nicht möglich, da es sich hier um die schon besprochene Nicht-Konjunktivität handelt.

3. Probabilistische Je-Desto-Aussagen

Wir können im folgenden nicht sämtliche Varianten von monotonen probabilistischen Beziehungen zwischen Ordnungsbegriffen diskutieren. Deshalb wollen wir uns auf die Fälle beschränken, in denen solche Aussagen durch die Werte von Korrelationskoeffizienten dargestellt werden, eine in den Sozialwissenschaften nicht unübliche Vorgehensweise. Die Charakterisierung eines Zusammen-

hangs zwischen zwei Variablen durch Korrelationskoeffizienten setzt voraus, daß die Variablen die Struktur mindestens einer Intervallskala besitzen, und daß die Beziehung durch eine lineare Funktion approximierbar ist, aber wir hatten in Abschnitt II 1.1 schon gesehen, daß in vielen Fällen bestimmte Voraussetzungen notwendig sind, um überhaupt Aussagen formulieren zu können. Wir wollen deshalb im folgenden annehmen, daß die Beziehungen linear sind, daß jedoch nur ihre Richtung durch das Vorzeichen des Korrelationskoeffizienten spezifiziert wird, und daß sie probabilistisch sind, insofern die jeweils explikativen Variablen nicht hinreichend sind zur genauen Prognose der jeweils abhängigen Variablen.

Für Aussagen dieser Art ist das Problem der Transitivität ausführlich diskutiert worden. Wir möchten uns deshalb auf die Wiedergabe der wesentlichen Punkte, sofern diese in unserem Argumentationszusammenhang interessant sind, beschränken und ansonsten auf die relevante Literatur verweisen (H.L. COSTNER und R.K. LEIK 1964; J. GALTUNG 1967: 466 ff.; H.M. BLALOCK 1969: Kap. 2; H.L. COSTNER 1969; K.D. BAILEY 1970). Betrachten wir eine „Miniatur“-Wirkungsstruktur, bestehend aus den Variablen X_1 , X_2 und X_3 , und gehen von den beiden theoretischen Aussagen (genauer: Satzformeln) aus¹⁶:

(1) Je größer X_1 , desto größer X_2

(2) Je größer X_2 , desto größer X_3

Als numerische Ausdrücke für diese beiden monotonen Beziehungen seien die beiden Korrelationen

$$(1') \text{ Corr}(X_1, X_2) = r_{12}$$

$$(2') \text{ Corr}(X_2, X_3) = r_{23}$$

gegeben.

Kann man nun aus den beiden „Aussagen“ (1) und (2) die „Aussage“

(3) Je größer X_1 , desto größer X_3

16 Solange die Bedeutung der Variablen X_i nicht angegeben wird, haben die sprachlichen Ausdrücke der Art (1) bzw. (2) nur die Form von Sätzen oder Aussagen, ohne solche zu sein. Um sie als Aussagen (Sätze) zu verstehen, ist es erforderlich, alle auftretenden X_i in bestimmter Weise zu deuten. — Man beachte weiterhin, daß im Unterschied zu II.1 alle Satzformeln dieses Abschnitts probabilistisch zu interpretieren sind. Dementsprechend müßte z. B. (1) genauer lauten: „Für alle Paare (i, j) von Objekten gilt: wenn der X_1 -Wert von i größer ist als der X_1 -Wert von j , dann besteht eine bestimmte Wahrscheinlichkeit (z. B. größer als 0.5 oder nahe 1) dafür, daß auch der X_2 -Wert von i größer ist als der X_2 -Wert von j “.

deduzieren und etwas über die absolute Größe und das Vorzeichen von

$$(3') \text{Corr}(X_1, X_3) = r_{13}$$

ausagen? Wir fragen also nach einer Deduktionsfunktion für r_{13} in Abhängigkeit von r_{12} und r_{23} , die es uns in der schwachen Form einer Vorzeichenregel erlauben würde, (3) zu behaupten.

Zunächst einmal ist zu bemerken, daß es nicht ohne weiteres nur aufgrund der Zuordnungen (1) – (1') und (2) – (2') klar ist, inwiefern die formalen Eigenschaften der Korrelationsbeziehung denen der theoretischen Beziehungen gleich sind. Die Korrelationsbeziehung ist symmetrisch, d.h. es gilt

$$\text{Corr}(X_1, X_2) = \text{Corr}(X_2, X_1)$$

Wenn nun (1) durch (1') ohne Bedeutungsverlust wiedergegeben wird, dann muß (1) äquivalent sein der „Aussage“

$$(1_a) \text{ Je größer } X_2, \text{ desto größer } X_1$$

da (1_a) dann ebenfalls (1') äquivalent wäre. Dann aber müßte gelten: falls aus (1) und (2) die „Aussage“ (3) ableitbar ist, ist (3) auch aus (1_a) und (2) ableitbar. Da wir jedoch i.a. nicht bereit sein werden, trotz als wahr vorausgesetzter Prämissen diese Konsequenz zu akzeptieren, muß entweder die (bis jetzt implizit gelassene, aber in Analogie zum Aussagenkalkül verwandte) Ableitungsregel oder die Annahme aufgegeben werden, daß eine Äquivalenz zwischen Korrelationsbeziehungen und monotonen Beziehungen der Art (1), (2) oder (3) besteht.

Welche Aussagen kann man nun über den Wert von $\text{Corr}(X_1, X_3)$ bei gegebenen r_{12} und r_{23} formulieren, ohne die Art der Beziehung (1) und (2) weiter zu präzisieren? Gemäß der Formel für die Partial-Korrelation für $r_{13.2}$ („Korrelation von X_1 mit X_3 nach Elimination der linearen Effekte von X_2 “) gilt nach Umformung

$$(PK) \quad r_{13} = r_{12} \cdot r_{23} + r_{13.2} \sqrt{1 - r_{12}^2} \sqrt{1 - r_{23}^2}$$

Hieraus folgt, daß

$$\text{Corr}(X_1, X_3) = \text{Corr}(X_1, X_2) \cdot \text{Corr}(X_2, X_3)$$

genau dann gilt; wenn die Partial-Korrelation $r_{13.2} = 0$ ist. D.h. (PK) wäre die gesuchte Deduktionsfunktion, wenn man das Verschwinden von $r_{13.2}$ voraussetzen könnte. Insbesondere wür-

de dann auch die Vorzeichenregel gelten (H.L. COSTNER und R.K. LEIK 1964): *Das Vorzeichen der deduzierten Beziehung ist gleich dem algebraischen Produkt der Vorzeichen der postulierten Beziehungen.*

$$(VR) \quad \text{sign } r_{13} = \text{sign } r_{12} \cdot \text{sign } r_{23}$$

Nun ist die Annahme über das Verschwinden der Partialkorrelation nicht in den Aussagen (1) und (2) enthalten; sie muß also zusätzlich vorausgesetzt werden. (2) besagt nämlich nur, daß X_2 eine erklärende Variable für X_3 ist, nicht jedoch die einzige¹⁷. Insbesondere ist nach (2) zugelassen, daß auch X_1 eine erklärende Variable von X_3 ist. Das Verschwinden der Partialkorrelation aber bedeutet, daß X_1 nur über X_2 auf X_3 einwirkt, also keine direkte erklärende Variable von X_3 ist. Diese Annahme entspricht einer ganz bestimmten Wirkungsstruktur, welche folgendermaßen darstellbar ist¹⁸:

Nun besteht die Möglichkeit, komplexe Wirkungsstrukturen mit Hilfe sogenannter *linearer Kausalstrukturen* zu analysieren. Hier werden die Beziehungen zwischen den Variablen durch Systeme linearer Gleichungen beschrieben, wobei jede Gleichung jeweils eine abhängige Variable in Abhängigkeit von anderen Variablen und einem zusätzlichen Term, in dem die Wirkungen sämtlicher implizit gelassener Faktoren zusammenge-

17 Setzt man dafür, daß eine Variable X_2 als im probabilistischen Sinne erklärend für X_3 bezeichnet wird, als notwendig voraus, daß die Verteilung von X_3 unter der Bedingung von X_2 keine konstante Funktion von X_2 ist, dann schließt die in (2) formulierte Bedingung, daß die bedingten Verteilungen von X_3 mit den X_2 -Werten variieren, es nicht aus, daß sie in Abhängigkeit von weiteren Variablen variieren, also z. B. auch X_1 .

18 Mit e_i sind sämtliche Faktoren gemeint, die nicht explizit Berücksichtigung finden (sog. implizite Faktoren; in der Korrelationsrechnung: Residuen, IrrtumsvARIABLE, error terms). Da X_1 als exogene Variable behandelt wird, deren Zustandekommen im Gegensatz zu den endogenen Variablen X_2 und X_3 nicht zur Diskussion steht, muß man hier keine impliziten Faktoren auführen. Weiterhin kommt im Diagramm die noch zu erwähnende Annahme zum Ausdruck, daß die direkten Wirkungen der e_i spezifisch sind: alle impliziten Faktoren, welche direkt auf X_2 wirken, haben keinen direkten Einfluß auf X_3 und umgekehrt.

faßt werden, darstellt. Ziel der Analyse ist es, aus gegebenen Korrelationskoeffizienten die Koeffizienten des Gleichungssystems zu identifizieren, in welchen Richtung und Größe der Abhängigkeiten zum Ausdruck kommen. Falls es sich um sogen. rekursive Strukturen handelt, bei denen die „Kausalbeziehungen“ asymmetrisch sind, sind die Koeffizienten unter der Annahme, daß die Wirkungen der impliziten Faktoren spezifisch sind, ohne große technische Schwierigkeiten zu identifizieren¹⁹.

Das Resultat einer solchen *Dependenz-* oder *Pfadanalyse* besteht in einer Menge von Dependenz- bzw. Pfadkoeffizienten, welche gewissermaßen die direkten Wirkungen zwischen den Variablen ausdrücken. Sie können aufgrund des sogenannten *Grundtheorems der Pfadanalyse* berechnet werden, welches im Rahmen der rekursiven linearen Kausalstrukturen beweisbar ist:

$$(Th) \quad r_{ij} = d_{ij} + \sum_k r_{jk} \cdot d_{ik}$$

D.h. jede Korrelation r_{ij} zwischen zwei Variablen X_i und X_j läßt sich darstellen als Summe des direkten Einflusses von X_j nach X_i – gemessen durch d_{ij} – und der indirekten, über andere Variable X_k vermittelten Einflüsse, wobei jeder direkte Einfluß d_{ik} von X_k auf die abhängige Variable X_i gewichtet wird mit der Korrelation r_{jk} dieser Variablen X_k mit der betrachteten erkläativen Variablen X_j . Durch rekursive Verwendung des Grundtheorems der Pfadanalyse ist es möglich, alle Korrelationskoeffizienten durch Dependenzkoeffizienten auszudrücken, bzw. auch umgekehrt Dependenzkoeffizienten aus Korrelationen zu berechnen.

Aus einer speziellen rekursiven linearen Kausalstruktur, die durch eine bestimmte Matrix der Dependenzkoeffi-

zienten definiert ist, kann man deshalb folgende Arten von Konsequenzen deduzieren:

1. Für beliebige Variablenpaare sind die numerischen Werte der Korrelationskoeffizienten berechenbar.
2. Jeder Korrelationskoeffizient als Ausdruck des Bruttozusammenhangs ist zerlegbar in einzelne Anteile als Ausdruck der spezifischen Wirkungen, der indirekten Effekte und der konfundierten Effekte aller der exogenen Variablen, deren Korrelationen nicht zerlegt werden.
3. Es sind Beziehungen zwischen den Korrelationskoeffizienten deduzierbar. Dies ist besonders dann interessant, wenn bestimmte Eingänge der Matrix der Dependenzkoeffizienten über das Erfordernis der Rekursivität hinaus gleich Null sind. In diesen Fällen ist die Variablenstruktur unvollständig²⁰.

Allerdings ist zu beachten, daß zur Bestimmung des numerischen Wertes eines bestimmten Korrelationskoeffizienten die Werte der Dependenzkoeffizienten gegeben sein müssen. Falls diese aufgrund externer Informationen bekannt sind, ist dies nicht weiter problematisch, wie z.B. in der Genetik, wo sie ein Bestandteil der Gesetzesannahmen über die Vererbungsmechanismen sind (s. z.B. C.C. LI 1955). Anders jedoch, wenn die Dependenzkoeffizienten erst aus empirischen Korrelationen geschätzt werden. Hier ist die Unterscheidung zwischen vollständigen und unvollständigen Strukturen wichtig²¹. Denn nur in unvollständigen Strukturen sind in den linearen Kausalstrukturen Konsequenzen über die Werte von Korrelationskoeffizienten ableitbar aus Dependenzkoeffizienten, zu deren Bestimmung man die abzuleitenden Korrelationskoeffizienten nicht benötigt. Bei vollständigen Strukturen hingegen benötigt man zur Bestimmung der Dependenzkoeffizienten sämtliche empirischen Korrelationen.

19 In der Ökonometrie und Genetik ist der der Theorie linearer Kausalstrukturen zugrundeliegende Kalkül schon lange bekannt. RAYMOND BOUDON (1965, 1967) und OTIS D. DUNCAN (1966) kommt das Verdienst zu, ihn in die Soziologie eingeführt zu haben. Inzwischen wird die den Soziologen zugängliche und verständliche Literatur immer umfangreicher (Übersichten über die Probleme und einzelnen Techniken z. B. bei R. BOUDON 1968; K.C. LAND 1968; D.R. HEISE 1968; R. ZIEGLER 1972; Kap. 4); gleiches gilt für die empirischen Arbeiten, in denen z.B. Pfadanalysen Verwendung finden (in deutscher Sprache z.B. W. MÜLLER 1972; s. auch E. WEDE 1970). Generell ist festzustellen, daß sich das Denken in „Kausalmodellen“ steigender Beliebtheit erfreut, wozu vor allem die Arbeiten von HUBERT M. BLALOCK beigetragen haben (z.B. 1961a, 1968, 1969).

20 In rekursiven Kausalstrukturen ist zirkuläre Kausalität ausgeschlossen: Keine der endogenen Variablen kann (direkt oder indirekt) sowohl Ursache als auch Wirkung einer anderen endogenen Variablen sein. Dabei heißt „ X_i ist direkte Ursache von X_j “, daß es in dem die Wirkungsstruktur beschreibenden Gleichungssystem eine Gleichung für X_j gibt, in der X_i als erklärende Variable mit einem von Null verschiedenen Koeffizienten α_{ij} auftritt. Vermittelt der Vorschrift, daß ein Pfeil von X_j nach X_i genau dann verläuft, wenn X_j direkt X_i verursacht, werden Wirkungsstrukturen durch Graphen dargestellt; indirekte Verursachung wird dann durch die Existenz eines Pfades beschrieben. Graphen rekursiver Systeme enthalten daher keine Zyklen: es gibt keinen Pfad, so daß man von einer Variablen zu ihr „zurückkommen“ kann. Aus der Nicht-Zirkularität folgt insbesondere die Asymmetrie der Relation der direkten Verursachung: $\alpha_{ij} \neq 0$ impliziert $\alpha_{ji} = 0$. Daher sind rekursive Systeme durch eine durch das Vertauschen der Gleichungen ermöglichte Anordnung aller Koeffizienten in Form einer Dreiecksmatrix definierbar.

21 Vollständige rekursive Kausalstrukturen sind dadurch charakterisiert, daß für alle Variablenpaare genau eine direkte Kausalbeziehung gilt. Damit wird eine streng lineare Anordnung aller expliziten Variablen möglich. In unvollständigen Kausalstrukturen gibt es hingegen Variablenpaare, zwischen denen zwar eine indirekte aber keine Kausalbeziehung existiert.

Unter der Voraussetzung, daß die aus den explizit berücksichtigten Variablen X_1 , X_2 und X_3 und den impliziten Faktoren e_2 und e_3 bestehende Menge ein geschlossenes System bildet, läßt sich die eingangs erwähnte Aussage (2) im Rahmen einer linearen Kausalstruktur auf zweierlei Weise darstellen, nämlich durch

$$(II) \quad X_3 = \alpha_{32} X_2 + e_3$$

$$(IIa) \quad X_3 = \alpha_{31} X_1 + \alpha_{32} X_2 + e_3 \quad \text{oder durch}$$

Die Darstellung (IIa) ist allgemeiner, da in ihr im Vergleich zu (II) ein möglicherweise von Null verschiedener Koeffizient α_{31} auftritt. Hierbei sind die α -Koeffizienten unstandardisierte Dependenzkoeffizienten, die den Grad der asymmetrischen Einflußbeziehungen von X_2 auf X_3 bzw. von X_1 und X_2 auf X_3 ausdrücken. Drückt man (1) ebenfalls in der Sprache der linearen Kausalstrukturen aus, erhält man

$$(I) \quad X_2 = \alpha_{21} X_1 + e_2$$

Unter Verwendung des Pfadtheorems ergeben sich folgende Zerlegungen der Korrelationskoeffizienten in standardisierte Dependenzkoeffizienten

$$(4) \quad r_{12} = d_{21}$$

$$(5) \quad r_{23} = d_{32} + d_{31} d_{21}$$

$$(6) \quad r_{13} = d_{31} + d_{32} \cdot d_{21}$$

Falls man davon ausgehen kann, daß der relevante Ausschnitt der Wirkungsstruktur durch (II) und nicht (IIa) beschrieben wird – d.h. daß die standardisierten und unstandardisierten Dependenzkoeffizienten für die direkte Wirkung von X_1 auf X_3 gleich Null sind ($\alpha_{31} = d_{31} = 0$) – erhält man in der linearen Kausalstruktur also

$$r_{13} = d_{32} \cdot d_{21}$$

und da dann

$$r_{23} = d_{32} \cdot d_{21} = r_{12} \quad \text{gilt}$$

schließlich

$$(7) \quad r_{13} = r_{12} \cdot r_{23}$$

Dies ist die *gesuchte Deduktionsfunktion* für die Korrelationskoeffizienten für die spezielle Wirkungsstruktur, in der X_1 keine direkten Wirkungen auf X_3 ausübt.

Die Annahme einer speziellen Wirkungsstruktur in einer linearen Kausalstruktur führte also zum gleichen Ergebnis, wie die des Verschwindens der Partialkorrelation. Tatsächlich sind die Bedingungen $\alpha_{31} = d_{31} = 0$ und $r_{13,2} = 0$ äquivalent, falls man Partialkorrelationen im Rahmen von linearen Kausalstrukturen interpretiert²².

Halten wir fest, daß wir für die uns interessierenden Korrelationen mit Hilfe des Pfadtheorems eine spezielle Deduktionsfunktion gefunden haben, wenn wir davon ausgehen können, daß für das Variablensystem (II) gilt. In diesem Fall besitzt auch die Vorzeichenregel Gültigkeit, und es kann konstatiert werden, daß aus (1) und (2), wenn sie in der angegebenen Weise als (I) und (II) verstanden werden, (3) logisch folgt. Das *Pfadtheorem* selbst kann in seinen spezialisierten Versionen der Form (4) bis (6) als *generelle Deduktionsfunktion* betrachtet werden.

Falls man die Prämissenmenge (1) und (2) nicht in der angegebenen Weise reformuliert, insbesondere also zuläßt, daß (IIa) ebenfalls eine mögliche Darstellung der „Aussage“ (2) in einer linearen Kausalstruktur ist, was bedeutet, daß $r_{13,2} \neq 0$ zugelassen ist, kann man die Vorzeichenregel für (3') nur dann anwenden und damit „Aussage“ (3) deduzieren, wenn man als zusätzliche Annahme die Bedingung einführt, daß das Vorzeichen der Partialkorrelation gleich dem Vorzeichen des Produkts der beiden einfachen Korrelationen ist. Inhaltlich bedeutet dies, daß X_1 unabhängig von X_2 einen direkten Einfluß auf X_3 ausübt, der mit dem indirekten über X_2 vermittelten Einfluß gleichgerichtet ist.

Will man selbst diese Annahme nicht machen, d.h. soll erlaubt sein, daß direkte und indirekte Einflüsse von X_1 auf X_3 gegenläufig sind, kann man die Vorzeichenregel nur genau dann verwenden – und (3) aus (1) und (2) deduzieren –, wenn gilt²³

$$(8) \quad r_{12}^2 + r_{23}^2 > 1$$

Wir kommen auch hier wiederum zu dem Ergebnis, daß im Gegensatz zur D-Implikation von At-

22 Diese Behauptung wird im einzelnen in einer noch unveröffentlichten Arbeit präzisiert und bewiesen: H.J. HUMMELL, Komplexe Auswertungsverfahren nicht-experimenteller Daten. Eine Einführung in die Mehrvariablenanalyse, verv. MS, Köln 1971 (erscheint als „Kausalanalyse“ demnächst bei R. Oldenbourg München).

23 Beweis ebenfalls in der in der vorigen Anm. erwähnten Arbeit.

tributen die Korrelationsbeziehung nicht ohne weiteres die Bedingung der Transitivität erfüllt; sie ist nur bei sehr hohen Werten der Korrelationskoeffizienten transitiv. Wenn man aus Prämissen der Art (1) und (2) eine „Aussage“ der Art (3) deduzieren möchte und diese „Aussagen“ nur so weit spezifiziert, als sich dies in entsprechenden Korrelationskoeffizienten niederschlägt – Korrelationsaussagen und theoretische Aussagen in diesem Zusammenhang also als äquivalent betrachtet –, so wird man implizit als zusätzliche Prämisse immer Bedingungen der Art (8) als erfüllt ansehen müssen. Man muß jedoch damit rechnen, daß bei längeren Deduktionsketten die Korrelationskoeffizienten immer kleiner werden. Zusätzliche Annahmen über numerische Werte der Korrelationskoeffizienten sind dann nicht erforderlich, wenn man die theoretischen Aussagen im Rahmen einer linearen Kausalstruktur ausdrückt. Im Unterschied zur „Aussage“ (2), bei der es offen bleibt, welche anderen Variablen neben X_2 ebenfalls auf X_3 wirken, treten in der (2) entsprechenden linearen Gleichung (II) oder (IIa) *alle* für X_3 relevanten Variablen auf. Die Geschlossenheit des Systems wird formal durch Berücksichtigung der impliziten Faktoren erreicht. Und da es sich in diesem formalen Sinn um ein geschlossenes System handelt, ist für jedes beliebige Variablenpaar der entsprechende Korrelationskoeffizient deduzierbar.

III.

Die vorangegangenen Ausführungen sollten zeigen, daß man in vielen Fällen damit rechnen muß, aus sozialwissenschaftlichen Aussagen keine eindeutigen Konsequenzen ziehen zu können. Andererseits ist Argumentieren ohne Deduzieren nicht möglich.

Gleichzeitig dürfte deutlich geworden sein, daß die einzige Chance, diese mißliche Lage zu beheben, in der Übersetzung der verbal formulierten theoretischen Aussagen in die Sprache spezieller Kalküle besteht. So wurde an den verschiedenen Stellen die Verwendung der Theorie differenzierbarer Funktionen, der Wahrscheinlichkeitstheorie oder der Theorie linearer Gleichungen vorgeschlagen. Solche Vorgehensweisen können wir vorläufig als „Formalisierung“ bezeichnen.

Ohne uns zu sehr in wissenschaftstheoretische Erörterungen zu verlieren und eine allen Ansprüchen genügende Explikation der Begriffe „Kalkül“,

„Formalisierung“, „formales System“ vorzunehmen (s. z.B. R. CARNAP 1939; A. TARSKI 1946: Kap. 6; P. SUPPES 1957: Kap. 12; M. BRODBECK 1959; H. HOCHBERG 1959), wollen wir uns abschließend mit der Frage beschäftigen, inwiefern die Formalisierung sozialwissenschaftlicher Aussagen ein Weg zur Lösung der Deduktionsproblematik sein kann. Es liegt uns fern, die Formalisierung gewissermaßen als *deus ex machina* zu präsentieren, wodurch plötzlich sämtliche Probleme zum Verschwinden gebracht werden, aber wir wollen erörtern, ob Formalisierung sich überhaupt als Kandidat für die Lösung spezieller deduktionstechnischer Probleme zu qualifizieren vermag und damit als Strategie der Theoriekonstruktion in Frage kommt.

1. Betrachten wir dazu unsere Vorgehensweise in Abschnitt II.3. Hier wurde den Aussageformeln (1) und (2), die wir bei entsprechender Interpretation der Variablen X_1 , X_2 und X_3 als Bestandteil einer möglichen verbalen Theorie auffassen konnten, die Satzformeln (I) und (II) im „Kalkül“ der linearen Kausalstrukturen zugeordnet. Dabei sollte die Zuordnung die Bedingung erfüllen, daß sich die Bedeutungen der zugeordneten Aussagen, die bei gleicher Interpretation der Zeichen entstehen, in großen Bereichen überschneiden, wenn sie nicht sogar gleich sind. Falls diese relative Bedeutungsinvarianz gegeben ist, kann man auch davon sprechen, daß die Aussagen aus einer Sprache in eine andere übersetzt wurden.

Natürlich ist die angegebene „Übersetzung“ nicht völlig bedeutungserhaltend: durch die Formulierung der Aussagen in der Sprache der linearen Kausalstrukturen werden die ursprünglichen Aussagen der verbalen Theorie modifiziert und präzisiert; beispielsweise wird Asymmetrie der Variablenbeziehungen angenommen, worüber vor der „Übersetzung“ nichts ausgesagt wurde.

Durch die Übersetzung der verbalen Theorie in eine spezielle Sprache, die als Kalkül ausgearbeitet vorliegt, wird es möglich, von dem Reichtum an formalen Strukturen dieser Sprache zu profitieren²⁴. So ist im Kalkül der linearen Kausalstruk-

24 Versucht man den unserer „normalen“ Wortsprache zugrundeliegenden logischen Apparat zu rekonstruieren, erhält man im wesentlichen den Kalkül der elementaren Logik. Sprachen, in denen neben der elementaren Logik weitere formale Ausdrucksmittel zur Verfügung stehen, bezeichnen wir hier als Sprachen mit „reicherer“ Struktur. Dementsprechend beziehen sich Formulierungen wie „verbale Theorien“ u.ä. immer auf Aussagen, die in einer Sprache formuliert

turen das Grundtheorem der Pfadanalyse beweisbar, welche eine Zerlegung der empirischen Korrelationen in Summen von Produkten von Dependenz- oder Pfadkoeffizienten gibt. Durch dieses Theorem erhielten wir eine Deduktionsfunktion für die Korrelationen, d.h. wir konnten bei gegebenen Pfadkoeffizienten beliebige Korrelationen berechnen. Damit war dann schließlich entscheidbar, ob (3) eine gültige Deduktion darstellte oder nicht.

Diese Situation, die wir hier anhand des dritten Falles beschrieben haben, lag auch in den anderen Fällen vor: Die partiellen Lösungen der behandelten spezifischen Deduktionsprobleme bestanden darin, daß man versuchte, gewisse Formeln zu verwenden, die in bestimmten Kalkülen beweisbar sind, so die Kettenregel bzw. das Theorem der totalen Wahrscheinlichkeit. Dies wiederum setzte voraus, daß die verbal formulierten Aussagen in die Sprache der jeweiligen Kalküle übersetzt wurden.

werden, in der ausschließlich von der elementaren Logik Gebrauch gemacht wird. – Wenn wir uns hier hauptsächlich mit dem formalen Apparat von Sprachen beschäftigen (da das Deduktionsproblem nur hiervon abhängt und damit auch nicht durch die Formulierung bestimmter inhaltlicher Aussagen zu lösen ist), so impliziert das nicht, daß die deskriptiven Bestandteile der Wortsprache nicht besonders vielfältig sein könnten. Man betrachte dazu beispielsweise den technischen Jargon der Soziologen. Eine gewisse negative Korrelation der Originalität in der Erfindung immer neuer deskriptiver Terme und Bedeutungen einerseits mit dem Mangel an Bereitschaft, vorhandene formale Strukturen zu verwenden, wenn dies an sich notwendig wäre, bzw. dem Unvermögen, selbst solche zu konstruieren andererseits, ist jedoch nicht zu leugnen. – Auch soll hier nicht behauptet werden, daß es für das Argumentieren in der Wortsprache nicht Grade in der selbstkritischen und verantwortungs- (weil sich der Konsequenzen) bewußten Verwendung des Apparates der elementaren Logik gibt. Man kann hier vielmehr von der Pedanterie bis zur Schlamperei alle Stile des Umgangs mit Logik antreffen. Um Mißverständnisse zu vermeiden, sei schließlich ausdrücklich darauf hingewiesen, daß es uns fern liegt, durch den Nachweis der Inadäquanz verbalen Razonnierens für relevante Probleme einer neuen Schludrigkeit das Wort zu reden. Wir wollen nicht bestreiten, daß in vielen Fällen in der Präzisierung vorgelegter „Aussagen“ mit den Mitteln der elementaren Logik (etwa in Form explizit formulierter Wenn-Dann- oder Je-Desto-Aussagen) ein Fortschritt gesehen werden muß. Das Herausarbeiten der in einem Argumentationszusammenhang relevanten Variablen, die Explikation ihrer Bedeutung, das Klären ihrer Beziehungen untereinander – alles das sind wichtige Schritte, auch wenn wir aufgrund der hier behandelten Deduktionsproblematik geneigt sind, solche Schritte nur als „erste“ zu betrachten, denen die „eigentliche“ Arbeit noch folgen muß.

Wir wollen in erster Approximation mit *Formalisierung* diesen Sachverhalt der *Übersetzung verbal formulierter Theorien in die Sprache spezieller Kalküle* bezeichnen. Was damit im einzelnen gemeint ist und inwiefern sich dieser Vorschlag von anderen Verwendungsweisen des Terms „Formalisierung“ unterscheidet, soll durch eine Betrachtung verwandter Begriffe noch etwas genauer erörtert werden.

2. Wenn man ein System empirischer Aussagen zu axiomatisieren versucht, wird die Gesamtmenge der Aussagen in zwei sich ausschließende Teilmengen zerlegt, in die Menge der Axiome und in die der Theoreme. Hinzu kommt eine analoge Zerlegung des gesamten deskriptiven Vokabulars in undefinierte und definierte Terme. Die Zerlegungen sollen die Bedingung erfüllen, daß sämtliche Theoreme aus den Axiomen einschließlich der Definitionen deduzierbar sind. Nun ist die Beziehung der Ableitbarkeit eine formale Eigenschaft von Aussagen, d.h. zur Ableitung ist es irrelevant, welche Bedeutung die Terme haben bzw. ob die Sätze bei bestimmten Bedeutungen der in ihnen auftretenden Terme wahr oder falsch sind.

Deshalb ist es möglich, in axiomatischen Systemen die Irrelevanz der Bedeutung der Terme dadurch zu kennzeichnen, daß man die deskriptiven Terme durch logische Zeichen ersetzt. Das Ergebnis ist ein *abstraktes Axiomensystem* oder *Kalkül*²⁵, in welchem allein die syntaktische Struktur der Aussagen zum Ausdruck kommt. Da in abstrakten Axiomensystemen die Terme keine Bedeutung haben, sind die Zeichenfolgen keine Sätze sondern Satzformeln: sie sind weder wahr noch falsch. Sie werden erst zu wahren oder falschen Sätzen bei einer entsprechenden Interpretation. Nur: falls bei einer bestimmten Interpretation des abstrakten Axiomensystems die Grundsätze oder Axiome wahr sind, dann sind auch die abgeleiteten Sätze oder Theoreme notwendigerweise wahr. D.h. dem syntaktischen Ableitungsbegriff entspricht ein semantischer Folgerungsbegriff. Beide sind in dem Sinne miteinander verträglich, als die Ableitbarkeit der Theoreme gerade besagt, daß, wenn die Axiome bei einer bestimmten Interpretation der Grundterme wahr sind, auch die Theoreme wahr sind. Insofern kann man sagen: *Kalküle* sind

25 Genau genommen kann man von einem Kalkül erst dann sprechen, wenn sämtliche syntaktischen Regeln für die entsprechende Sprache explizit angegeben werden, d.h. die Sprache als syntaktisches System konstruiert (oder rekonstruiert) wird (vgl. hierzu R. CARNAP 1939, 1960: §§ 20ff. und § 42).

Mengen von logisch wahren Wenn-Dann-„Aussagen“, wobei die Wahrheit der Komponenten nicht zur Diskussion steht; sie sind Mengen von Formeln, die in den betreffenden Kalkülen bewiesen werden können, und die deshalb dazu dienen können, unter der Annahme der Wahrheit der Wenn-Komponenten die Dann-Komponenten als wahr zu behaupten.

Wenn man aus der Satzformel eines Kalküls eine Aussage gewinnt, indem man eine Deutung vornimmt, gibt man bestimmte Designata an, auf die sich die Zeichen des Kalküls beziehen, d.h. man spezifiziert einen Bereich von Objekten und Eigenschaften dieser Objekte. Falls die Interpretation so gear- tet ist, daß die erhaltenen Aussagen wahr sind, sagt man auch, der Bereich von Objekten „erfüllt“ den Kalkül oder der Objektbereich ist ein „Modell“ des Kalküls.

Je nach der Interpretation kann der gleiche Kalkül zu verschiedenen Systemen von Aussagen führen; und damit kann es (falls die jeweiligen Aussagen wahr sind) auch verschiedene Modelle des gleichen Kalküls geben. Im Falle der empirischen Interpretation haben wir es dann mit verschiedenen empirischen Theorien zu tun, die verschiedene Gegenstandsbereiche beschreiben. Die Theorien haben alle die gleiche syntaktische Struktur: sie sind isomorph. In ähnlicher Weise haben die verschiedenen realen Modelle gewisse strukturelle Eigenschaften gemeinsam, nämlich genau die, welche in den verschiedenen syntaktisch isomorphen Theorien beschrieben werden.

Neben der *empirischen Interpretation durch Angabe eines realen Modells* ist u. U. auch eine *logisch-mathematische Interpretation durch Angabe logischer Modelle* für den gleichen Kalkül möglich. Bekannt sind die Axiomensysteme der Geometrie, die einerseits, als physikalische Geometrien interpretiert, empirische Gesetze über Eigenschaften von festen Körpern im Raum und andererseits, als analytische Geometrien, im Rahmen der Arithmetik beweisbare Aussagen über n -tupeln von Zahlen formulieren. Wir waren auf einen solchen Fall der logischen Interpretation eines Kalküls in II 2.2 eingegangen.

Dort wurde eine Deutung des Axiomensystems der Wahrscheinlichkeitstheorie mit Hilfe von relativen Häufigkeiten in endlichen Mengen vorgenommen. Das Ergebnis dieser Interpretation waren Aussagen, die notwendigerweise wahr sind: alle W-Axiome sind dann mengentheoretisch beweisbar, was heißt, alle endlichen empirischen Bereiche erfüllen bei den angegebenen Deutungsregeln das abstrakte Axiomensystem und damit sämtliche Konsequenzen, die aus den Axiomen

folgen, insbesondere also das von uns benutzte spezielle Theorem.

Anders wiederum war die Situation im Fall der linearen Kausalstrukturen in III 3. Hier liegt der Kalkül in Form einer Menge linearer Gleichungen zwischen Variablen vor, die zu einer Menge von wahren oder falschen empirischen Aussagen werden, wenn den Variablen ein spezieller empirischer Gehalt zugewiesen wird. Die resultierenden Aussagen bilden dann die Theorie eines bestimmten Gegenstandsbereiches, auf den sich die Variablen beziehen; allerdings ist die Theorie schon in einer speziellen Sprache – nämlich der mathematischen linearer Gleichungssysteme – und nicht in der üblichen Wortsprache formuliert. Da für den gleichen Kalkül unterschiedliche Interpretationen möglich sind, es damit auch unterschiedliche reale Modelle geben kann, kann man auch sagen, daß die linearen Kausalstrukturen die syntaktische Struktur inhaltlich verschiedener empirischer Theorien darstellen.

Wahrheit bzw. empirische Gültigkeit sind keine Eigenschaften linearer Kausalstrukturen sondern der entsprechenden Theorien, die man durch Interpretation erhält. Ob eine lineare Struktur in dem Sinne anwendbar ist, daß eine spezielle Interpretation zu wahren Aussagen führt oder nicht, ist nicht eine Frage, die durch Analysen des Kalküls entscheidbar ist – auch nicht durch eine Analyse der Deutungsregeln –, sondern dies ist eine Frage nach der Existenz eines realen Modells, d.h. einer Menge empirischer Objekte, die die Formeln des Kalküls erfüllt²⁶.

Fassen wir unsere Überlegungen zusammen, so können wir sagen, daß wir bei der Formalisierung mit Kalkülen zu tun haben, die in unterschiedlicher Weise zu empirischen Theorien interpretiert werden können. Diese Theorien beschreiben – empirische Gültigkeit vorausgesetzt – Ausschnitte der realen Welt, welche selbst wieder-

26 Wenn der Kalkül „anwendbar“ sein soll, muß die Frage nach der Existenz eines realen Modells als schon entschieden vorausgesetzt werden. Mit anderen Worten: es muß aufgrund von dem Kalkül externen Kriterien geprüft werden, ob eine Menge von Gegenständen den Kalkül erfüllt, d.h. ob eine entsprechende empirische Theorie wahr ist. Allerdings gehört dazu insbesondere, daß auch sämtliche Konsequenzen, die im Kalkül behauptet werden können, unter der betreffenden Deutung sich als wahr erweisen, was wiederum dazu dienen kann, die empirische Anwendbarkeit des Kalküls anhand der Konsequenzen einem weiteren Test zu unterziehen. Im Beispiel der linearen Kausalstrukturen können solche Tests aufgrund der Konsequenzen nur dann vorgenommen werden, wenn die Wirkungsstrukturen unvollständig sind.

rum als reale Modelle des Kalküls aufgefaßt werden können, die ihn erfüllen. Verschiedene empirische Theorien sind, insofern sie als Interpretationen des gleichen Kalküls verstanden werden können, isomorph.

3. Häufig wird unter Formalisierung speziell die inverse Beziehung zur Interpretation eines Kalküls verstanden. „Eine Theorie kann zur Interpretation eines Kalküls verwendet werden. Wenn dies geschieht, ist der uninterpretierte Kalkül gewissermaßen die formale Struktur einer Theorie, d. h. der uninterpretierte Kalkül enthält allein das syntaktische System der Theorie. Ein solches syntaktisches System bezeichnen wir als eine *Formalisierung* einer Theorie“ (K.D. OPP 1970: 226). Und bei MAY BRODBECK heißt es in ihrem in einem „Symposium on Sociological Theory“ erschienenen Artikel „Any theory, quantified or nonquantified, may be formalized . . . The result is called a *formalization* of the theory. These expressions exhibit form but no content, since the letter variables have not been given descriptive meaning“ (1959: 391). Gängig sind solche Formulierungen, daß die Formalisierung empirisch gehaltvoller Theorien darin besteht, die deskriptiven Zeichen durch logische zu ersetzen, um die syntaktische Struktur der Aussagen deutlich hervortreten zu lassen. Wir wollen zu dieser Auffassung Stellung nehmen, indem wir von der *Funktion von Formalisierung* ausgehen. In unseren obigen Erörterungen verschiedener Arten sozialwissenschaftlicher Aussagen wurde zur Formalisierung immer dann gegriffen, wenn es nicht mehr möglich war, aus den nicht-formalisierten Aussagen bestimmte logische Konsequenzen zu ziehen. Unter dem hier behandelten Aspekt können wir deshalb solche Formalisierungen von Aussagensystemen als sinnvoll betrachten, deren Resultat in dem Übergang zu Kalkülen besteht, in denen eine Vielzahl von deduktiven Konsequenzen formuliert werden können. Dies kann in zweierlei Hinsicht möglich sein, indem nämlich entweder die durch Formalisierung der empirischen Aussagen erhaltenen Satzformeln selbst Axiome von ausgearbeiteten abstrakten Axiomensystemen sind oder die erhaltenen Satzformeln als zusätzliche Prämissen in abstrakten Axiomensystemen Verwendung finden können. Je mehr Aussagen in den betreffenden Kalkülen bewiesen oder unter Zuhilfenahme der zusätzlichen Prämissen abgeleitet werden können, desto eher kann man das Ziel der Formalisierung als realisiert ansehen.

Formalisierung ist also nicht bloß gleich Ersatz deskriptiver Terme durch logische, sondern Darstellung der Aussagen durch Axiome oder zusätzliche Prämissen in einem an Ausdrucksmöglichkeiten reichen speziellen Kalkül, wozu allerdings die erwähnte Ersetzung der Terme Voraussetzung ist. Damit sollte man Formalisierung in diesem Sinne von ihrer Funktion her u. a. von der *Axiomatisierung* (a) oder der *Formalisierung im Sinne der Ersetzung der Zeichen* (b) abgrenzen.

(a) Auch verbal formulierte Theorien können axiomatisiert werden. Ein Beispiel sind die im Gefolge der Ausführungen von HANS L. ZETTERBERG in der Soziologie vorgeschlagenen „axiomatisch-deduktiven Systeme“ (z.B. H.L. ZETTERBERG 1954; J.P. GIBBS u. W.T. MARTIN 1962; K.P. SCHWIRIAN u. J.W. PREHN 1962; T.K. HOPKINS 1964; J. HAGE 1965). In den vorliegenden Systemen sind jedoch nur sehr wenig Deduktionen möglich – falls man nicht überhaupt alle „Theoreme“ als aus den „Axiomen“ nicht folgend bezeichnen muß. Natürlich spielen bei der Formalisierung Axiomensysteme eine Rolle, da der Kalkül ein abstraktes Axiomensystem ist. Aber entscheidend ist nicht so sehr die Axiomatisierung der Aussagen als vielmehr die Existenz des strukturreichen Kalküls, der es erlaubt, die betreffenden Aussagen mit Hilfe seiner Zeichen (approximativ) auszudrücken.

(b) Die Ersetzung deskriptiver Terme durch logische Zeichen führt nicht unbedingt auf einen Kalkül, in dem eine Vielzahl von Konsequenzen formulierbar sind. Wenn wir beispielsweise in der Aussage „Je höher der Rang einer Person, desto größer sind ihre Kontrollmöglichkeiten“ die deskriptiven Begriffe „Rang“ und „Kontrollmöglichkeiten“ durch uninterpretierte Zeichen ersetzen und entsprechend bei allen anderen Aussagen eines gegebenen Systems verfahren, sind wir bezüglich der Deduktionsproblematik keinen Schritt weiter gekommen. Dies liegt daran, daß wir die syntaktische Struktur unserer Sprache nicht angetastet haben. Umgekehrt ist es bezüglich der deduktionstechnischen Möglichkeiten völlig irrelevant, ob man sich beispielsweise bei Aussagen in der Sprache der linearen Kausalstrukturen die Variablen X_1 als uninterpretierte Zeichen vorstellt oder sich gewissermaßen noch bestimmte Deutungsregeln hinzudenkt. Ob eine lineare Kausalstruktur *ohne* Deutungsregeln für die Variablen die sogenannte formalisierte Fassung der linearen Kausalstruktur *mit* Deutungsregeln darstellt, die man dann empirische Theorie nennt, ist ziemlich unerheblich. Entscheidend ist, daß sich in unserem Beispiel die syntaktische Struktur der Formeln (I) und (II) (mit oder ohne Deutungsregeln für die Zeichen) von der syntaktischen Struktur der Sätze (1) und (2) unterscheidet, und zwar so, daß im ersten Falle Deduktionen durchgeführt werden können, die in der verbal formulierten Theorie unmöglich sind. Worum es also geht, ist die Zuordnung der Sätze (1) und (2) zu den Formeln (I) und (II) und nicht, ob man sich bei letzteren die Deutungsregeln noch hinzudenken soll oder nicht, d. h. ob man (I) und (II) als formales und semantisches System oder nur als formales System auffassen soll.

Es geht also bei der Formalisierung verbaler Aussagen um das Problem der Übersetzung von Aussagen der Wortsprache in Aussagen mit speziellen

syntaktischen Strukturen²⁷. Während sich ein Begriff, der Formalisierung als Ersatz deskriptiver Zeichen durch logische bestimmt, auf einen Prozeß bezieht, der sich *innerhalb* einer Sprache mit gegebener syntaktischer Struktur abspielt, wird von uns ein Prozeß gemeint, der verschiedene Sprachen mit unterschiedlichen syntaktischen Strukturen verbindet. *Formalisierung bedeutet nicht das präzise Herausarbeiten der logischen Struktur von gegebenen Aussagen, sondern vielmehr die Übersetzung dieser Aussagen in eine strukturreichere Sprache*²⁸.

Wir sind uns der Tatsache bewußt, daß eine Abgrenzung des Formalisierungsbegriffes in der hier vorgeschlagenen Richtung erheblich weniger präzise sein wird als die oben erwähnte und von uns kritisierte Definition. Wenn man Formalisierung bezieht auf interpretierte oder uninterpretierte Ausdrücke eines Sprachsystems mit gleicher syntaktischer Struktur, so hat man ein ziemlich eindeutiges Kriterium dafür, ob eine bestimmte Formel die formalisierte Version einer empirisch gehaltenen Aussage ist oder nicht. Allerdings wird es dann kaum möglich, Formalisierungen von ihrer Funktion her zu unterscheiden, bzw. gerade die interessanten „Formalisierungen“ fallen überhaupt nicht unter den so präzisierten Begriff. Bestimmt man jedoch Formalisierung von der Funktion her, dann ist es sehr schwer zu entscheiden, wann tatsächlich eine Übersetzung vorliegt, die die Bedingung der Bedeutungsinvarianz erfüllt. Man sollte aber überlegen, ob die Frage der Übersetzungs-

adäquanz nicht in die Heuristik gehört. Damit wäre die Frage der Übersetzung einer Theorie in ähnlicher Weise zu behandeln wie die der Begriffsexplikation. Entscheidendes Kriterium einer adäquaten Begriffsexplikation oder adäquaten Formalisierung ist die theoretische Fruchtbarkeit und nicht so sehr die Treue in der Wiedergabe aller Bedeutungen des zu explizierenden Begriffs bzw. der zu formalisierenden Theorie. Das Resultat des Übersetzungsprozesses, die formalisierte Theorie, hat einen Anspruch darauf, im Hinblick auf die Leistungsfähigkeit betrachtet zu werden und nicht im Hinblick auf ihre Abstammung aus einer nicht-formalisierten Theorie.

Diese These, die die Übersetzungsadäquanz in die Heuristik verweist, dürfte akzeptabel erscheinen, wenn man bedenkt, daß die hier kritisierte Auffassung von Formalisierung häufig mit der Vorstellung einhergeht, daß es *zunächst theoretische Aussagen* geben müsse, die man dann *anschließend* vielleicht auch noch *formalisieren* kann. Zwar wollen wir zugestehen, daß es möglich sei, isolierte Je-Desto-Aussagen oder probabilistische Wenn-Dann-Aussagen zu formulieren; wir möchten aber nochmals unserer Skepsis Ausdruck verleihen, daß es immer möglich sei, Theorien als *Systeme* solcher Aussagen, die in logischen Beziehungen zu einander stehen, zu konstruieren. In vielen Fällen dürfte es eine Illusion sein, nichtformalisierte „Theorien“ überhaupt aufstellen zu können, die man dann der logischen Präzision wegen formalisiert, d. h. Formalisierung als saubere Reinschrift einer ansonsten fertig formulierten verbalen Theorie auffaßt (R. ZIEGLER 1972: 9). Im Gegenteil: in vielen Fällen besteht in der Konstruktion formalisierter Aussagensysteme die einzige Chance, theoretische Erörterungen überhaupt durchführen zu können. Insofern ist gerade in den Sozialwissenschaften *Formalisierung als Strategie der Theoriebildung* weiter zu fassen und meint nicht nur das Übersetzen gegebener verbaler Theorien, sondern generell die *Formulierung von Aussagensystemen in strukturreichen Sprachen, die als spezielle Kalküle ausgearbeitet vorliegen oder ausgearbeitet werden*. Dabei ist es erst in zweiter Linie relevant, ob den formalisierten Aussagen eine verbal formulierte Fassung vorangeht, also Formalisierung zusätzlich auch noch Übersetzung ist.

Von der hier herausgestellten Funktion der Formalisierung her ergibt sich auch eine Bewertung einer Vielzahl von (halb-)formalisierten Darstellungen theoretischer Ansätze, wie sie beispielsweise in dem Band „Sociological Theories in Progress“ zu finden sind (J. BERGER u. a. 1966). Hier ist oft zu konstatieren, daß eine Formalisierung in der Weise vorgenommen wurde, die keine

27 Gemäß dieser Begriffsbestimmung würde auch die *Simulation von Theorien* bzw. die *Konstruktion von Simulations-,modellen* zur Formalisierung zählen, da in Simulationen die syntaktischen Eigenschaften von Theorien physikalisch realisiert werden. Man vgl. hierzu beispielsweise den Aufsatz von S. LINDENBERG (1971), in welchem ebenfalls auf den Aspekt, daß Simulation Ableitung ermöglichen soll, hingewiesen wird.

28 Wenn J. BERGER u. a. definieren: „By ‚formalization‘ we refer to the general process of making explicit the logical structure of a set of assertions. We include in this process the activity of translating a set of statements about empirical phenomena into a particular formal language“ (1961: 3), so betonen wir hier besonders den zweiten Aspekt. Dabei muß nicht mehr betont werden, daß die formalisierte Fassung einer Theorie im allgemeinen eine Präzisierung und Modifizierung der letzteren darstellt: die durch Interpretation eines Kalküls gewonnene empirische Theorie stimmt nur approximativ mit der verbalen Fassung überein. Auch ist Formalisierung kein mechanisches Verfahren, wie das gängige Definitionen nahelegen; sondern als Spezialfall der „Übersetzung“ von Aussagen, die in verschiedenen Sprachen formuliert werden, ist Formalisierung durchaus kreativ.

Deduktionen erlaubt. Dies wird besonders deutlich, wenn die gesamte „Theorie“ nur in Form von Axiomen und Definitionen präsentiert wird, ohne daß ein abgeleitetes Theorem auftritt. Solche Formalisierungen ohne Deduktionen sind sicher geeignet zur logischen Analyse der Struktur der Beziehungen zwischen den Begriffen und zur logischen Struktur der Begriffe – falls diese durch Definitionen miteinander verknüpft und nicht jeder Begriff einfach durch einen Term repräsentiert wird. Von JOSEPH BERGER u.a. werden solche formalisierten Aussagensysteme, die vor allem der Präzisierung der Beziehungen und der Klärung der Begriffe dienen, als „explanational models“ bezeichnet (1962: Kap. 2). Aber wir sollten doch nicht vergessen: *Formalisierung von Aussagensystemen soll Deduktionen ermöglichen*. „No calculus without calculation“²⁹.

Literatur

- BAILEY, Kenneth D., 1970: Evaluating Axiomatic Theories. In: *Sociological Methodology 1970*, herg. von E.F. Borgatta und G.W. Bohrnstedt. San Francisco: Jossey-Bass.
- BERGER, Joseph, Bernard P. COHEN, J. Laurie SNELL und Morris ZELDITCH, Jr., 1962: *Types of Formalization in Small-Group Research*. Boston: Houghton Mifflin.
- BERGER, Joseph, Morris ZELDITCH, Jr., und Bo ANDERSON, 1966: *Sociological Theories in Progress*, Band 1. Boston: Houghton Mifflin.
- BLALOCK, Hubert M., 1961a: *Causal Inferences in Non-experimental Research*. Chapel Hill, N.C.: University of North Carolina Press.
- BLALOCK, Hubert M., 1961b: Correlation and Causality: The Multivariate Case. *Social Forces* 39, 246–251.
- BLALOCK, Hubert M., 1962: Four-variable causal models and partial correlations. *American Journal of Sociology* 68, 182–194.
- BLALOCK, Hubert M., 1968: Theory Building and Causal Inferences. In: *Methodology in Social Research*, herg. von H.M. Blalock und A.B. Blalock. New York: McGraw Hill.
- BLALOCK, Hubert M., 1969: *Theory Construction. From Verbal to Mathematical Formulations*. Englewood Cliffs, N.J.: Prentice Hall.
- BOUDON, Raymond, 1965: A Method of Linear Causal Analysis: Dependence Analysis. *American Sociological Review* 30, 365–374.
- BOUDON, Raymond, 1967: *L'analyse mathématique des faits sociaux*. Paris: Plon.
- BOUDON, Raymond, 1968: A New Look at Correlation Analysis. In: *Methodology in Social Research*, herg. von H.M. Blalock und A.B. Blalock. New York: McGraw Hill; rev. Fassung in: R. Boudon, *La crise de la Sociologie*. Genève-Paris: Droz 1971.
- BRAITHWAITE, Richard Bevan, 1960: *Scientific Explanation. A Study of the Function of Theory, Probability and Law in Science*. New York: Harper Torchbooks (zuerst 1953).
- BRODBECK, May, 1959: Models, Meaning, and Theories. In: *Symposium on Sociological Theory*, herg. von L. Gross. New York: Harper & Row.
- CARLSSON, Gösta, 1969: *Social Mobility and Class Structure*. Lund: Gleerup.
- CARNAP, Rudolf, 1939: *Foundations of Logic and Mathematics*. In: *International Encyclopedia of Unified Science*, Band 1, herg. von O. Neurath. Chicago: University of Chicago Press.
- CARNAP, Rudolf, 1950: *Logical Foundations of Probability*. Chicago: University of Chicago Press (2. rev. Aufl. 1962).
- CARNAP, Rudolf, 1959: *Induktive Logik und Wahrscheinlichkeit*. Bearbeitet von W. Stegmüller. Wien: Springer.
- CARNAP, Rudolf, 1960: Einführung in die symbolische Logik mit besonderer Berücksichtigung ihrer Anwendungen. 2. erw. Aufl. Wien: Springer.
- COSTNER, Herbert L., 1969: Theory, Deduction, and Rules of Correspondence. *American Journal of Sociology* 75, 245–263.
- COSTNER, Herbert L., und Robert K. LEIK, 1964: Deductions from „Axiomatic Theory“. *American Sociological Review* 29, 819–835.
- DUNCAN, Otis D., 1966: Path Analysis: Sociological Examples. *American Journal of Sociology* 72, 1–15.
- DURKHEIM, Emile, 1960: *Le suicide*. Nouv. édition. Paris: Presses Universitaires de France (zuerst 1897).
- GALTUNG, Johan, 1967: *Theory and Methods of Social Research*. Oslo: Universitetsforlaget.
- GIBBS, Jack P., und Walter T. MARTIN, 1962: Urbanization, Technology, and the Division of Labor: International Patterns. *American Sociological Review* 27, 667–677.
- HAGE, Jerald, 1965: An Axiomatic Theory of Organizations. *Administrative Science Quarterly* 10, 289–320.
- HEISE, David R., 1969: Problems in Path Analysis and Causal Inference. In: *Sociological Methodology 1969*, herg. von E.F. Borgatta und G.W. Bohrnstedt. San Francisco: Jossey-Bass.
- HEMPEL, Carl G., 1960: Inductive Inconsistencies. *Synthese* 12, 439–469; auch in: C.G. Hempel 1965.
- HEMPEL, Carl G., 1962: Deductive-Nomological versus Statistical Explanation. In: *Minnesota Studies in the Philosophy of Science*, Band 3, herg. von H. Feigl und G. Maxwell. Minneapolis: University of Minnesota Press.
- HEMPEL, Carl G., 1965: Aspects of Scientific Explanation and other Essays in the Philosophy of Science. New York-London: Free Press-Collier-Macmillan.
- HOCHBERG, Herbert, 1959: Axiomatic Systems, Formalization and Scientific Theories. In: *Symposium on Sociological Theory*, herg. von L. Gross. New York: Harper & Row.
- HOPKINS, Terence K., 1964: *The Exercise of Influence in Small Groups*. Totowa, N.J.: Bedminster Press.
- LAND, Kenneth C., 1969: Principles of Path Analysis. In: *Sociological Methodology 1969*, herg. von E.F. Borgatta und G.W. Bohrnstedt. San Francisco: Jossey-Bass.

29 Wir entnehmen diese Formulierung einer Anmerkung des Buches „Scientific Explanation“ von RICHARD BEVAN BRAITHWAITE (1966: 366), das im übrigen beispielhaft für die Art ist, wie man zur Behandlung bestimmter (hier: wissenschaftstheoretischer) Probleme mit Nutzen formale Kalküle verwenden kann.

- LAZARSFELD, Paul F., 1961: The Algebra of Dichotomous Systems. In: Studies in Item Analysis and Prediction, herg. von H. Solomon. Stanford, Cal.: Stanford University Press.
- LAZARSFELD, Paul F., 1968: The Analysis of Attribute Data. In: International Encyclopedia of the Social Sciences, Band 15. New York: Macmillan & Free Press.
- LI, C.C., 1955: Population Genetics. Chicago: University of Chicago Press.
- LINDENBERG, Siegwart, 1971: Simulation und Theoriebildung. In: Sozialtheorie und soziale Praxis. Eduard Baumgarten zum 70. Geburtstag, herg. von H. Albert. Meisenheim a.G.: Hain.
- MÜLLER, Walter, 1972: Bildung und Mobilitätsprozeß. Eine Anwendung der Pfadanalyse. Zeitschrift für Soziologie 1, 65–84.
- OPP, Karl-Dieter, 1970: Methodologie der Sozialwissenschaften. Eine Einführung in Probleme ihrer Theoriebildung. Reinbek b. Hamburg: Rowohlt (rde 339–341).
- SCHWIRIAN, K.P., und J.W. PREHN 1962: An Axiomatic Theory of Urbanization. American Sociological Review 27, 812–825.
- SIMON, Herbert, 1954: Spurious Correlation: A Causal Interpretation. Journal of the American Statistical Association 49; auch in: Ders. 1957: Models of Man – Social and Rational. New York: Wiley.
- STEGMÜLLER, Wolfgang, 1969: Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie. Band I: Wissenschaftliche Erklärung und Begründung. Berlin-Heidelberg-New York: Springer.
- SUPPES, Patrick, 1957: Introduction to Logic. Princeton, N.J.: Van Nostrand.
- TARSKI, Alfred, 1946: Introduction to Logic and to the Methodology of Deductive Sciences. 2. rev. Aufl. New York: Oxford University Press.
- WEEDE, Erich, 1970: Zur Methode der kausalen Abhängigkeitsanalyse (Pfadanalyse) in der nicht-experimentellen Forschung. Kölner Zeitschrift für Soziologie und Sozialpsychologie 22, 532–550.
- WILBER, George L., 1967: Causal Models and Probability. Social Forces 46, 81–89.
- ZETTERBERG, Hans L., 1965: On Theory and Verification in Sociology. 3. erw. Aufl. Totowa, N.J.: Bedminster Press (zuerst 1954).
- ZIEGLER, Rolf, 1972: Theorie und Modell. Der Beitrag der Formalisierung zur soziologischen Theoriebildung. München: Oldenbourg.

Anschr. des Verfassers: Dr. H.J. HUMMELL
503 Hürth, Bonnstraße 180