Multidimensionale Skalierung

.

Skriptum zu Evaluation und Forschungsmethoden

U. Mortensen

FB Psychologie und Sportwissenschaften, Institut III
Westfälischen-Wilhelms-Universität Münster

Inhaltsverzeichnis

1 Einführung

Objekte unterscheiden sich bezüglich der Ausprägung ihrer Eigenschaften. Die übereinstimmung der Ausprägungen der Eigenschaften bei zwei Objekten drückt sich in ihrer Ähnlichkeit oder Unähnlichkeit aus. Das Ziel der multidimensionalen Skalierung ist, aus Maßen für die Ähnlichkeit bzw. Unähnlichkeit – allgemein der *Proximität*[1] der Objekte – diejenigen Eigenschaften oder "Dimensionen" zu gewinnen, die die Proximität der Objekte bestimmen. Dabei kann die globale Proximität beurteilt werden oder die Proximität bezüglich eines vorgegebenen Merkmals. Das Resultat einer multidimensionalen Skalierung liefert nicht nur ein Bild der latenten Merkmale, die die Objekte charakterisieren, sondern gleichzeitig ein Bild des Urteilsprozesses.

Die Objekte können Personen, Gegenstände, aber auch Begriffe sein:

Beispiel 1.1 Thurstone (1927) ließ Verbrechen hinsichtlich ihrer Schwere beurteilen. Hier sind die Verbrechen die Objekte, und ihre Schwere ist das interessierende Merkmal. Thurstone gelang es, eine *Intervallskala* für das Merkmal "Schwere eines Verbrechens" zu erstellen. Die beurteilenden Personen (Vpn) hatten die Aufgabe, für alle möglichen Paare einer Menge von n Verbrechen anzugeben, welches der jeweils zwei Verbrechen "schwerer" war. Aus den Häufigkeiten der "schwerer"-Urteile können dann Skalenwerte für die einzelnen Verbrechen errechnet werden. Man kann sagen, daß die Vpn ihre Urteile auf der Basis der psychologischen Distanz zwischen den Verbrechen gebildet haben. $\square$

Fraglos sind Begriffe wie "Schwere eines Verbrechens" komplex in dem Sinne, daß dieser Begriff von einer Reihe von Merkmalen der Objekte abhängen wird. Man kann vermuten, daß diese Merkmale wie latente Variable in das Urteil über die interessierende Eigenschaft des Objektes eingehen.

Ein Modell des Urteilsprozesses, das eindimensionale Skalen für das Merkmal (Schwere der Verbrechen in Beispiel 1.1 vorhersagt, ergibt sich leicht aus dem Modell der Faktorenanalyse: es sei s_i, der Skalenwert des i-ten Objektes. Dann werden die latenten Variablen wie Prädiktoren des Merkmals angeschrieben. Die Regressionsgewichte sind dabei für das betrachtete Merkmal charakteristisch, so daß

$$s_i = a_1 l_{i1} + a_2 l_{i2} + \cdots + a_r l_{ir} \tag{1}$$

gelten soll; der Einfachheit halber werden mögliche Fehler e zunächst weggelassen. Die $l_{is}, s = 1, \cdots, r$ sind die Maße des Objekts, etwa des i-ten Verbrechens, auf den r verschiedenen latenten Beurteilungsdimensionen, und die Gewichte $a_1, \cdots .a_r$ geben die Gewichtung dieser Dimensionen für die Beurteilung hinsichtlich der "Schwere" der Verbrechen an. Verschiedene Beurteiler (Vpn) treten in (1) noch nicht auf, das Modell (1) wird zunächst für nur *eine* Person betrachtet.

Nach (1) sind die Gewichte $a_s, s = 1, \cdots, r$ für alle Objekte gleich, die unterschiedlichen Beurteilungen etwa verschiedener Verbrechen resultieren aus deren unterschiedlichen Ausprägungen auf den latenten Dimensionen, wie etwa das Ausmaß an Gewalttätigkeit gegen Personen, an politischer/religiöser Motiviertheit, an Raffinesse (z.B. bei einem Versicherungsbetrug). Die a_s sind eben charakteristisch für *ein* Merkmal von Verbrechen, dessen Ausprägung sich als Linearkombination der latenten Dimensionen gewinnen läßt.

Die Form von (1) impliziert die Eindimensionalität der Beurteilungen. Die l_{is} können ja zu einer $(n \times r)$-Matrix zusammengefaßt werden, n die Anzahl der Objekte und r die Anzahl der latenten Dimensionen, die zur Beurteilung der Objekte herangezogen werden, und die Gewichte a_s können zu einem Vektor $a = (a_1, \cdots, a_r)'$ zusammengefaßt werden. Es sei $S = (s_1, \cdots, s_n)'$ der Vektor der Skalenwerte, d.h. jede Komponente von S repräsentiere den Skalenwert eines

[1] Nähe

Objekts. Dann ist Modell (1) äquivalent mit

$$S = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{pmatrix} = \begin{pmatrix} l_{11} & l_{12} & \cdots & l_{1r} \\ l_{21} & l_{22} & \cdots & l_{2r} \\ \vdots & & \cdots & \\ l_{n1} & l_{n2} & \cdots & l_{nr} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_r \end{pmatrix} = L\, a \tag{2}$$

S ist eine Linearkombination der Spaltenvektoren von L. Dem faktorenanalytischen Modell entsprechend können die Spalten von L als Basisvektoren eines n-dimensionalen Vektorraumes betrachtet werden. Damit wird angenommen, daß sie linear unabhängig sind. Der Vektor S ist dann eine Linearkombination linear unabhängiger Vektoren. Nun entspricht ein Vektor einer geraden Linie im Raum, ebenso die Koordinatenachsen, die durch die Basisvektoren, d.h. die Spaltenvektoren, von L definiert werden. Die Komponenten von S repräsentieren jeweils eines der Objekte, genau so wie die Komponenten der Spaltenvektoren von L die Objekte repräsentieren. Insbesondere können ja die Komponenten als Projektionen der Punkte betrachtet werden, die die einzelnen Objekte im n-dimensionalen Merkmalsraum (genauer gesagt: in einem r-dimenionalen Teilraum des n-dimensionalen Raums, der eben von den r n-dimensionalen Basisvektoren in L aufgespannt wird) repräsentieren. Analog dazu sind die Komponenten von S als Projektionen dieser Punkte auf eine Achse in diesem Raum zu verstehen, die das Merkmal repräsentiert, in bezug auf das die Merkmale gemessen werden; in Beispiel 1.1 eben die Schwere von Verbrechen.

Die Darstellung (2) weist damit auf eine bemerkenswerte Implikation des faktorenanalytischen Ansatzes, die Skalenwerte s_i, $i = 1, \cdots, n$ als Linearkombination latenter Dimensionen oder Variablen zu deuten, hin: ein beliebig gewähltes Merkmal der Objekte kann durch eine Achse im Raum der latenten Variablen dargestellt werden. Formal ergibt sich diese Achse als Vektor, der durch Linearkombination aus den Basisvektoren entsteht, wobei die Basisvektoren diejenigen Merkmale abbilden, die die urteilende Person als Bezugsmerkmale wählt. Es muß hier auf zwei Punkte eingegangen werden: es ist implizit angenommen worden,

1. daß die latenten Variablen, die die urteilende Person als Referenzmerkmal zur Beurteilung der Objekte hinsichtlich des vorgegebenen Merkmals benützt, durch *linear unabhängige* Vektoren repräsentiert werden,

2. daß für ein beliebig gewähltes Merkmal der Objekte, in bezug auf das sie gemessen werden sollen, ein entsprechender "Gewichts"-vektor a existiert, so daß die Meßwerte durch geeignete Gewichtung der Basisvektoren resultieren.

Es werde angenommen, daß die latenten Dimensionen *nicht* durch linear unabhängige Vektoren repräsentiert werden. Dann werden sie durch Vektoren abgebildet, die sich als Linearkombination linear unabhängiger Vektoren darstellen lassen. Aber alle Vektoren repräsentieren irgendein Merkmal, und so kann man sich ebensogut auf diejenigen Merkmale beziehen, die durch linear unabhängige Vektoren repräsentiert werden.Im faktorenanalytischen Ansatz können solche Vektoren jedenfalls stets gefunden werden. Dieser Sachverhalt ist eine Implikation des mathematischen Ansatzes, den man entweder mit einer psychologischen Axiomatik gleichsetzen will, oder der ganz einfach über die ursprüngliche Intention, nämlich ein einfaches Modell zur "Erklärung" der Skala zu haben, hinausweist: das Modell impliziert dann mehr, als man verantworten kann. Es ist in jedem Fall eine empirische Frage, ob der Ansatz *in seiner Allgemeinheit*, d.h. mit all seinen Implikationen, gültig ist.

Der zweite Punkt bezieht sich darauf, daß offenbar implizit angenommen wird, daß stets die gleichen latenten Dimensionen als Bezugsrahmen gewählt werden, denn es wird bei der Formulierung ja nichts darüber ausgesagt, *unter welchen Bedingungen* ein spezieller Bezugsrahmen gewählt wird, d.h. es wird nicht weiter spezifiziert, ob für ein gegebenes, zu beurteilendes

Merkmal wie etwa die "Schwere" eines Verbrechens auch *bestimmte*, für dieses Merkmal charakteristische Dimensionen als Rahmen gewählt werden oder nicht. Es wird nur angenommen, *daß* ein solcher Bezugsrahmen existiert. Es wird also nicht ausgeschlossen, daß für jedes Merkmal (außer der "Schwere" eines Verbrechens könnte man ja auch seine "Abscheulichkeit" messen wollen) auf den gleichen Satz von latenten Variablen bezug genommen wird. Man kann sich fragen, wieviele Merkmale denn überhaupt in bezug auf ein bestimmtes latentes System von Variablen beurteilt werden können. Man kann also die Hypothese aufstellen, daß jeder Vektor a eine bestimmte Achse im Merkmalsraum und damit ein meßbares Merkmal repräsentiert, - damit gäbe es unendlich viele. Die Frage ist, ob Menschen tatsächlich imstande sind, zwischen den im Prinzip unendlich nahe beieinander liegenden Merkmalen, die durch solche Achsen charakterisiert werden, überhaupt zu unterscheiden. Man mag nun einwenden, daß die Betrachtung solcher Implikationen gar nicht beabsichtigt war, als das Modell (2) angesetzt wurde. Gleichwohl macht die Betrachtung deutlich, daß der Ansatz (2) theoretisch deutlich unterbestimmt ist. Es ist jedenfalls einigermaßen unklar, welche Zusatzannahmen getroffen werden müssen, damit die Forderung nach unendlich feiner semantischer Differenzierung *nicht* impliziert wird. Solche Zusatzannahmen würden vermutlich nicht nur helfen, die vermutlich bizarre Forderung eines semantischen Kontinuums zu umgehen, sie würde auch die latenten Strukturen näher spezifizieren. Letzlich wird man auf die Notwendigkeit geführt, explizite Modelle über die psychologischen Prozesse zu formulieren, die zu Beurteilungen führen; Modelle der Art (2) sind, sofern sie nicht aus spezifischen psychologischen Theorien abgeleitet werden und damit auch spezifische Einschränkungen formuliert werden, *ad-hoc* und theoretisch insofern weitgehend leer; – leer, weil es aus rein mathematischen Gründen stets möglich ist, einen Vektor (die Skalenwerte definieren ja einen Vektor) als Linearkombination irgendwelcher linear unabhängiger Vektoren darzustellen. Es ist durchaus möglich, daß für begriffliche Stimuli wie Verbrechen ein Urteilsmodell, daß sich auf kontinuierliche latente Variablen bezieht, gar nicht sinnvoll ist (Tversky und Hutchinson (1986)).

Wie generell bei nahezu jedem faktorenanalytischen Modell, so sind auch im Modell (2) die verschiedenen möglichen Systeme von latenten Variablen zumindest mathematisch gleichberechtigt, und so lange es keine psychologischen Annahmen gibt, die bestimmte Systeme von Basisvektoren (und damit latenten Variablen) auszeichnen, das Modell eben keine Annahmen über sie macht, sind sie auch alle psychologisch gleichberechtigt. So kann das Merkmal "Schwere von Verbrechen", daß in (2) als abhängige Variable eingeführt wurde, auch mit einem der Vektoren in L vertauscht werden, der dann seinerseits als abhängige Variable auftaucht. Die "Schwere von Verbrechen" wird dann eine "latente" Dimension und kann dazu benützt werden, die Position der Objekte auf Achsen, die irgendwelche Merkmale repräsentieren, zu bestimmen. Vom Modell her gleichberechtigt wäre aber auch die Dimension "Abscheulichkeit", die dann zu einer Determinante der "Schwere"wird. Psychologisch plausibel ist die Gleichwertigkeit dieser beiden Dimensionen aber nicht: es fällt bei manchen Verbrechen leicht, ihre Schwere aus ihrer Abscheulichkeit herzuleiten, nicht aber ihre Abscheulichkeit aus ihrer Schwere: eine Vergewaltigung ist ein schweres Verbrechen wegen ihrer Abscheulichkeit, aber man empfindet sie nicht als abscheulich, weil sie ein schweres Verbrechen ist. Ein Versicherungsbetrug, zumindest wenn es um hinreichend hohe Beträge (Millionen DM) geht, wird von der Justiz ebenfalls als "schweres" Verbrechen geahndet, - mit gutem Grund, aber die "Abscheulichkeit" der Vergewaltigung haftet dieser Art von Betrug im allgemeinen nicht an.

Es liegt dann nahe, zu versuchen, die latenten Dimensionen direkt zu erfassen. Eine erste Möglichkeit dafür besteht darin, mehrere Personen urteilen zu lassen. Der Vektor a wird dann personenspezifisch sein, - aber möglicherweise auch die latenten Dimensionen, zumindest aber die Werte l_{is}, d.h. die Ausprägung der Objekte auf den Dimensionen. Die eben vorgebrachten Einwände gegen den faktorenanalytischen Ansatz (Gleichwertigkeit der möglichen Systeme von Basisvektoren, aber auch die postulierte Linearität) bleiben gleichwohl erhalten.

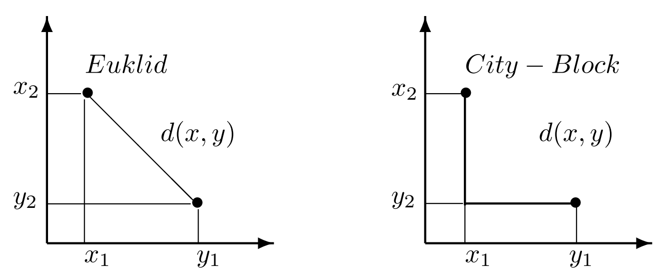
Eine zweite Möglichkeit kann in dem Versuch gesehen werden, aus den Schätzungen für die *psychologischen Distanzen* – Unähnlichkeiten – zwischen den Objekten die latenten Dimensionen zu erschließen. Denn wenn schon angenommen wird, daß die Objekte durch Ausprägungen auf verschiedenen latenten Dimensionen charakterisiert werden können, so werden sie durch Punkte in einem entsprechenden mehrdimensionalen Raum repräsentiert, und die Distanz zwischen diesen Punkten muß in irgendeiner Form die Unähnlichkeit, eben die psychologische Distanz, ausmachen.

Im faktorenanalytischen Modell werden die Objekte durch Punkte repräsentiert, die durch die für die Objekte stehenden Vektoren definiert werden. Die Charakterisierung durch Punkte ist der durch Vektoren äquivalent. Die Länge der Vektoren ist identisch mit dem Abstand (der Distanz), der Endpunkte der Vektoren vom Ursprung des Koordinatensytems, und die ist durch den Satz des Pythagoras erklärt. Dementsprechend sind auch die Distanzen zwischen allen Punkten durch den Satz des Pythagoras erklärt: sind x und y zwei Objekte mit den Koordinaten $x_1, ..., x_r$, $y_1, ..., y_r$, so ist die entsprechende Distanz zwischen den Punkten

$$d(x,y) = \sqrt{\sum_{k=1}^{r}(x_i - y_i)^2} \qquad (3)$$

Die Verbindungslinie zwischen zwei Punkten ist eine Gerade, und die ist nach Euklid die kürzeste Verbindung zwischen zwei Punkten. Dementsprechend sagt man, (3) definiere eine *euklidische Metrik*. In Abbildung 1 wird die euklidische Metrik veranschaulicht.

Abbildung 1: Distanzmaße: Euklidisch ($p = 2$) und City-Block $p = 1$



Die euklidische Metrik, d.h. die Definition des Begriffs der *Distanz*, gemäß (3), erscheint auf den ersten Blick eine natürliche Wahl zu sein, denn sie scheint sich ja auch direkt aus der Verallgemeinerung des eindimensionalen Skalenbegriffs zu ergeben. Dort werden die Objekte ja auf einer Geraden repräsentiert, und die Distanzen zwischen den Objekten sind einfach durch die Differenzen der Skalenwerte gegeben. Diese wiederum sind ein Spezialfall von (3), denn sie entsprechen dem Fall $r = 1$.

Aber es ist auch vorstellbar, daß man nur auf gekrümmten Wegen von einem Punkt zu einem anderen gelangt. Nach Einstein legt das Licht im Weltraum seinen Weg auf gekrümmten Bahnen zurück; die Gravitationskräfte großer Sterne oder ganzer Galaxien lenken es von seiner geraden Bahn ab. Vielleicht, so mag man sich überlegen, gibt es ja auch in psychologischen Räumen "Kräfte", die die Verbindungen zwischen Punkten krümmen. Man müßte dann nach einer Definition der Distanz suchen, die solche Krümmungen zuläßt und möglicherweise (3) als Spezialfall zuläßt. In Abbildung 1 wird außer der euklidischen Metrik noch der Spezialfall einer gekrümmten Verbindung zwischen zwei Punkten dargestellt, die *City-Block-Metrik*; hier ist die Distanz gerade gleich der Summe der Koordinatendifferenzen. Der Name ergibt sich aus der Analogie zur Verbindung zwischen zwei Punkten in einer Stadt wie Manhattan, in der die meisten Straßen rechtwinklig zueinander liegen.

überlegungen dieser Art führen dazu, daß man zunächst den Begriff der Metrik genauer definiert. Dann kann man psychologische Betrachtungen anstellen, die auf bestimmte Distanzbegriffe oder eben auf bestimmte Metriken führen. Es zeigt sich, daß für eine wichtige Klasse von Distanzen, die in der Tat (3) als Spezialfall enthält, mit einem gegebenen Satz von Distanzen auch die Wahl der Achsen, d.h. der latenten Variablen, festliegt. Wie im nächsten Abschnitt gezeigt wird, bleiben bei dieser Klasse die Distanzen nur dann invariant gegenüber einer Transformation der latenten Variablen, wenn die Distanzen durch die euklidische Metrik definiert sind. Was auf den ersten Blick wie eine "natürliche" Wahl erschien, erweist sich als Spezialfall mit unangenehmen Konsequenzen.

2 Metriken, Distanzen, Ähnlichkeiten

Wie in der Faktorenanalyse sollen bei der multidimensionalen Skalierung Objekte durch Punkte in einem mehrdimensionalen psychologischen Raum repräsentiert werden. Die Punkte sind durch Koordinaten definiert, die die Ausprägungen reflektieren, die ein Objekt hinsichtlich der verschiedenen Merkmalsdimensionen hat. Die Beurteilung der psychologischen Distanz ("Unähnlichkeit") zwischen irgend zwei Objekten wird jetzt von den Differenzen der Ausprägungen auf den einzelnen Dimensionen abhängen; die *Art* dieser Abhängigkeit bestimmt das Distanzurteil. Die Distanz ist ein Maß für die Entfernung zwischen zwei Punkten im psychologischen Raum, aber die psychologische Entfernung ist nicht notwendig durch die "Luftlinie", d.h. durch die Gerade, die die kürzeste Verbindung zwischen den Punkten liefert, gegeben. Die Distanz hängt vielmehr von der speziellen *Metrik* des Raumes ab. Dieser Begriff und seine Implikationen für den Urteilsprozeß sollen im folgenden erläutert werden.

Ein zunächst recht allgemeiner Ansatz besteht darin, den psychologischen Raum als einen *metrischen Raum* zu charakterisieren. Ein solcher Raum wird durch gewisse allgemeine Eigenschaften der Distanzen zwischen den Punkten im Raum beschrieben.

Definition 2.1 *Es sei Ω eine Menge von Objekten, und jedes Objekt aus Ω sei durch einen Punkt in einem Raum der Dimension $r \geq 1$ repräsentierbar. Für irgend drei Punkte $x, y, z \in \Omega$ mögen die Distanzen d zwischen ihnen die folgenden drei Eigenschaften haben:*

M1: $d(x, y) \geq 0$

M2: $d(x, y) = d(y, x)$ *(Reflexivität)*

M3: $d(x, z) \leq d(x, y) + d(y, z)$ *(Dreiecksungleichung)*

Dann heißt Ω ein metrischer Raum. *Gilt M2 nicht, so heißt der Raum auch* quasi-metrisch.

Die Forderung M1 ist unmittelbar einsichtig: Distanzen zwischen zwei Punkten können nicht negativ sein. M2 ist die Forderung der *Reflexivität.* Für viele physikalische Räume erscheint M2 ebenfalls eher trivial zu sein: so ist es bekanntlich von Münster bis Bremen so weit wie von Bremen bis Münster. Die Trivialität dieser Aussage beruht aber auf der stillschweigenden Annahme, daß man für die Reise die gleiche Autobahn oder die gleiche Bahnstrecke benutzt. Nehmen wir aber einmal an, die Bahn würde gar nicht mehr zwischen den Städten verkehren und die Autobahn sei von Bremen in Richtung Münster gesperrt, in umgekehrter Richtung aber befahrbar. Man müßte also für die Rückfahrt von Bremen nach Münster eine andere Strecke wählen, die entweder länger oder kürzer ist als die Autobahnstrecke, und M2 wäre nicht mehr erfüllt. *Soziale Distanzen* erfüllen oft ebenfalls M2 nicht: die soziale Distanz von Frankreich nach Deutschland ist bekanntlich größer als die von Deutschland nach Frankreich (d.h. Franzosen reisen weniger oft nach Deutschland als Deutsche nach Frankreich), man hat es hier also mit einer Quasi-Metrik zu tun.

Wirft man drei Sandkörner auf eine Tischplatte, so werden sie im allgemeinen die Eckpunkte eines Dreiecks bilden, nur in extrem seltenen Ausnahmefällen werden sie alle exakt auf einer Linie liegen. Wie auch immer sie liegen, die Distanzen zwischen ihnen erfüllen M3. Eine ähnliche Situation *kann* man auch dann vorfinden, wenn die Distanzen nicht einfach durch gerade Verbindungslinien gegeben sind.

Ein psychologischer Raum Ψ muß nicht notwendig ein metrischer Raum sein. Nimmt man einen metrischen Raum für Ψ an, so hat man bereits eine bestimmte Einschränkung bezüglich der Menge der möglichen Repräsentationen der Objekte aus Ω vorgenommen. Verfügt man über eine Theorie , aus der Aussagen A über die psychologischen Distanzen zwischen den Objekten aus Ω folgen, so definieren die Aussagen A, wie sich die Distanzen aus den Werten, die die Objekte auf den latenten Dimensionen haben, zusammensetzen. Um zu überprüfen, ob die Theorie mit der Annahme eines metrischen Raumes für Ψ kompatibel ist, muß man untersuchen, ob die durch A definierten Aussagen die Axiome M1 – M3 eines metrischen Raumes erfüllen.

Zur weiteren Diskussion der Repräsentation von Objekten in einem psychologischen Raum sind einige weitere Begriffe notwendig. Da Objekte durch Eigenschaften charakterisiert werden können, soll zunächst der Begriff des Maßes einer Menge eingeführt werden:

Definition 2.2 *Es seien A und B zwei beliebige, durchschnittsfremde Mengen, d.h. es soll $A \cap B = \emptyset$ gelten. m sei eine Abbildung $A \mapsto m(A)$, $B \mapsto m(B)$, $m(A)$ und $m(B)$ reelle Zahlen, derart, daß*

(i) $m(\emptyset) = 0$

(ii) $m(A) \geq 0$, $m(B) \geq 0$

(iii) $m(A \cup B) = m(A) + m(B)$

gilt; dann heißt m Maß.

m kann einfach die Anzahl der Elemente in einer Menge angeben, aber es ist auch möglich, daß m die Summe der "Gewichte" (was immer das sei) der Elemente ist, etc. Man macht sich leicht klar, daß für irgendzwei Mengen M_1 und M_2 mit $M_1 \subseteq M_2$ stets

$$m(M_1) \leq m(M_2) \tag{4}$$

gilt. Denn dann existiert eine Menge M_3 derart, daß $M_2 = M_1 \cup M_3$, $M_1 \cap M_3 = \emptyset$ so daß $m(M_2) = m(M_1) + m(M_3)$ nach (iii), und wegen $m(M_3) \geq 0$ (s. (ii)) folgt sofort (4). Weiter folgt für beliebige Mengen M_1 und M_2

$$m(M_1 \cup M_2) = m(M_1) + m(M_2) - m(M_1 \cap M_2), \tag{5}$$

wie man leicht durch Zerlegung von M_1 und M_2 in die disjunkten Teilmengen $M_1 - (M_1 \cap M_2)$, $M_2 - (M_1 \cap M_2)$ und $M_1 \cap M_2$ und Anwendung von (iii) nachweist.

Der folgende Begriff der *Aspektdistanz* wurde von Restle (1959) eingeführt:

Definition 2.3 *Es seien a und b zwei beliebige Objekte aus Ω. a und b seien durch Eigenschaftsmengen – Aspektmengen – A und B charakterisiert, auf denen ein Maß m erklärt sei. Weiter sei, wie üblich, $A - B$ die Menge der Elemente von A, die nicht auch in B enthalten sind, und $B - A$ sei analog definiert. Dann heißt*

$$d(a, b) = m(A - B) + m(B - A) \tag{6}$$

die Aspektdistanz *von a und b.*

$d(a,b)$ ist also um so größer, je größer die Maße $m(A-B)$ und $m(B-A)$ sind; dies sind die Maße auf den Aspektmengen, die für a bzw b spezifisch sind in dem Sinne, als die Eigenschaften in $A-B$ und $B-A$ *nicht* beiden Objekten a und b zukommen. Sind im Extremfall A und B identisch, so sind die Mengen $A-B$ und $B-A$ leer und wegen (i) ist die Distanz $d(a,b)$ gleich Null; Definition 2.3 erscheint also als sinnvoll. Ob die Definition auch stets *psychologisch* sinnvoll ist, ist eine andere Frage.

Es wird zunächst ein zu (2.3) äquivalenter Ausdruck für $d(a,b)$ hergeleitet. Sicherlich gilt $A = (A-B) \cup (A \cap B)$, und $(A-B) \cap (A \cap B) = \emptyset$. Also muß $m(A) = m(A-B) + m(A \cap B)$ gelten. Analog sieht man, daß $m(B) = m(B-A) + m(A \cap B)$ gelten muß.

Es gilt

$$d(a,b) = m(A) + m(B) - 2m(A \cap B). \tag{7}$$

Diese Aussage folgt sofort aus der Definition der Aspektdistanz sowie aus (5), ist aber auch leicht wie folgt zu sehen: es ist

$$m(A-B) = m(A) - m(A \cap B),\ m(B-A) = m(B) - m(A \cap B).$$

Setzt man diese Ausdrücke für $m(A-B)$ und $m(B-A)$ in (6) ein, so erhält man (7). Dieser Ausdruck hat gegenüber (6) den Vorteil, die Distanz $d(a,b)$ durch die Maße $m(A)$, $m(B)$ und $m(A \cap B)$ auszudrücken, die gelegentlich leichter herleitbar als die Maße $m(A-B)$ und $m(B-A)$ sind.

Es gilt nun

Theorem 2.1 *Ist d eine Aspektdistanz, so definiert sie einen metrischen Raum.*

Daß die Bedingung M1 erfüllt ist, folgt sofort aus der Definition des Maßes und der Tatsache, daß die Aspektdistanz eine Summe von Maßen ist. Die Reflexivität folgt ebenfalls sofort aus (7), denn $m(A) + m(B) = m(B) + m(A)$ und $A \cap B = B \cap A$. Der Nachweis der Gültigkeit der Dreiecksungleichung ist ein wenig länglich, und da die Details im Folgenden nicht gebraucht werden, findet man ihn im Anhang.

Beispiel 2.1 Die "Objekte" a und b seien Personen, die durch eine Reihe von Merkmalen charakterisiert werden können. Die psychologische Distanz zwischen diesen Personen soll über diese Merkmalsmengen ausgedrückt werden. Der Einfachheit halber sollen hier nur einfache, direkt beobachtbare Eigenschaften betrachtet werden. So sei A die Menge, die a beschreibt, und es sei $A = \{$überdurchschnittliche Körpergröße, mittelblondes Haar, Brille, männlich, zwischen 30 und 35 Jahre alt$\}$. b sei durch $B = \{$durchschnittlich groß, braunes Haar, keine Brille, weiblich, zwischen 30 und 35 Jahre alt$\}$ beschrieben.

Bei dieser Beschreibung werden einige Schwierigkeiten der Charakterisierung durch Aspektmengen deutlich: es tauchen für a und b die gleichen Merkmale, nur in verschiedenen Ausprägungen auf. Das erste Merkmal ist die Körpergröße, die einmal überdurchschnittlich, das andere Mal durchschnittlich ist. Die Körpergröße variiert sicherlich quantitativ. Das zweite Merkmal ist die Haarfarbe, die aber nicht quantitativ variiert, denn mittelblond und braun sind eigentlich nicht verschiedene quantitative Ausprägungen einer Haarfarbe. Eine Brille ist ebenfalls entweder vorhanden oder nicht. Hier kann man sagen, daß das Merkmal "Brille" vorhanden ist oder nicht. Das letzte Merkmal ist das Geschlecht. "Männlich" und "weiblich" sind sicherlich keine quantitativen Ausprägungen *eines* Merkmals. Will man nun die Aspektdistanz zwischen den Personen a und b anhand der Mengen A und B bestimmen, so muß man festlegen, wie man die Merkmale in die Distanzberechnung eingehen läßt, und das heißt, wie man die Mengen $A-B$ und $B-A$ und die zugehörigen Maße definiert. Eine Möglichkeit besteht darin, einfach die übereinstimmung des jeweiligen Merkmals zu konstatieren. $A-B$ enthält alle Eigenschaften, die nur in A, nicht auch in B enthalten sind. Man findet $A-B = \{$überdurchschnittliche Körpergröße, mittelblondes Haar, Brille, männlich$\}$, und $B-A = \{$durchschnittlich groß, braunes Haar, keine Brille, weiblich$\}$. Wählt man für die Maße dieser Mengen einfach die Anzahl der Elemente, so ist $m(A-B) = 4$, $m(B-A) = 4$ und nach (6) ist $d(a,b) = 8$. Es ist klar, daß dieser Berechnung von $d(a,b)$ eine Reihe von ad-hoc-Annahmen zugrunde liegt. So geht jedes Merkmal mit gleicher Gewichtung ein, was aber psychologisch nicht so sein muß. So kann für einen Beobachter die Haarfarbe von untergeordneter Bedeutung sein und damit die psychologische Distanz kaum beeinflussen. Der Unterschied der Körpergrößen geht gewissermaßen dichotom ein: es wird nur bewertet, ob die Körpergrößen gleich sind oder nicht, – auch ein geringer Unterschied bedeutet dann "Ungleichheit" und geht mit gleichem Gewicht ein wie etwa der Geschlechtsunterschied. Für einen bestimmten Beobachter mag aber die Körpergröße von vornherein eine geringe Bedeutung für die wahrgenommene Distanz haben, und darüber hinaus ist es plausibel, daß die Größe des Unterschiedes in das Urteil eingeht. Weiter ist es möglich, daß die Wahrnehmung der Ausprägung von Eigenschaften sowie die Gewichtung der Eigenschaften für den Wert der Distanz von Gestalteffekten abhängt, ein gegebenes Merkmal also je nach Kontext anders bewertet wird. Hieraus folgt, daß die Definition der psychologischen Distanzen durch Maße auf Aspektmengen die logische Struktur des Distanzbegriffes in sehr abstrakter Form charakterisiert und deshalb durch weitere Spezifikationen zur tatsächlichen Urteilsbildung in Beziehung gesetzt werden muß.

Für die Mengen und ihre Maße folgt, wenn man von den eben angegebenen Schwierigkeiten absieht, $m(A) = 5$, $m(B) = 5$, $m(A \cap B) = 1$, also $d(a,b) = 5+5-2\cdot 1 = 8$, in übereinstimmung mit der Berechnung gemäß $d(a,b) = m(A-B) + m(B-A)$. $\square$

Wenn ein Distanzmaß die Bedingungen einer Metrik erfüllt, so scheint es intuitiv vernünftige Bedingungen zu erfüllen. Gleichwohl ist der Begriff der Distanz in der Umgangsprache kaum enthalten. Hier spricht man eher von Ähnlichkeiten. Menschen scheinen aber sehr wohl imstande zu sein, Ähnlichkeitsurteile, an die man durch die Begrifflichkeit des Alltags gewöhnt ist, in Distanzurteile zu übersetzen. Man nimmt dabei an, daß die Distanz zwischen zwei Objekten als um so größer gesehen wird, je geringer die Ähnlichkeit der Objekte eingeschätzt wird, und umgekehrt wird die Distanz um so geringer sein, je größer die Ähnlichkeit gesehen wird. Dementsprechend liegt es nahe, nicht nur den Distanzbegriff, sondern auch den Ähnlichkeitsbegriff in einer formalisierten Form einzuführen, um von Ähnlichkeitsurteilen auf die latenten Dimensionen schließen zu können. Dies erscheint auch schon deshalb als vernünftig, weil man gelegentlich Daten vorliegen hat, die weder die Form von Distanz- noch von Ähnlichkeitsurteilen haben, sondern die Wahrscheinlichkeit von *Konfusionen* zwischen Objekten reflektieren. Wenn etwa zwei Objekte, etwa Personen, a und b häufig miteinander verwechselt werden, so wird man sagen, daß sie einander eben sehr ähnlich seien. Gelingt es, aus diesen Konfusionen ein Ähnlichkeitsmaß zu berechnen, so kann man dies wiederum auf Distanzen beziehen und von dort vielleicht zu den latenten Dimensionen vorstoßen, die die Ähnlichkeit ausmachen.

Es zeigt sich, daß der Begriff der Ähnlichkeit zwar aus der Umgangssprache jedem vertraut ist, er sich deshalb trotzdem nicht leicht so formulieren läßt, daß Deduktionen in bezug auf latente Urteilsdimensionen möglich sind. Intuitiv verbindet man mit dem Begriff, daß Ähnlichkeit ein Maß für gemeinsame Merkmale sein sollte. Es ist aber nicht klar, was dabei ein Merkmal ist, denn auch Beziehungen zwischen Merkmalen können in das Ähnlichkeitsurteil eingehen, d.h. auch Beziehungen zwischen Merkmalen sind Merkmale. Dies korrespondiert durchaus zum Gebrauch des Begriffs des Merkmals in der Logik, wo man häufig den Ausdruck "Prädikat" statt Merkmal gebraucht: Beziehungen zwischen Merkmalen sind Relationen, und die sind Prädikate "höherer Ordnung". Sollen Ähnlichkeiten in einer inversen Beziehung zu psychologischen Distanzen stehen und sollen die Distanzen eine Metrik liefern, so wird man auch fordern, daß Ähnlichkeitsurteile symmetrisch sind: drückt etwa $s(a,b)$ die Ähnlichkeit zwischen a und b aus, und ist $s(b,a)$ die Ähnlichkeit zwischen b und a, so soll $s(a,b) = s(b,a)$ gelten. Die Gültigkeit dieser Symmetrie wird durch den Gebrauch des Wortes "zwischen" suggeriert. Aber unsere Sprache erlaubt es, nicht nur von der Ähnlichkeit "zwischen" Objekten zu sprechen. Man kann auch fragen "Wie ähnlich ist das Objekt a dem Objekt b?", ohne daß man auf Unverständnis stoßen würde. Dabei zeigt sich dann, daß bei Aufgaben dieser Art durchaus Ähnlichkeiten $s(a,b)$ und $s(b,a)$ angegeben werden, die die Symmetrieforderung *nicht* erfüllen.

Einen historischen Überblick zum wissenschaftlichen Gebrauch des Begriffs der Ähnlichhkeit findet man bei Gregson (1975). An dieser Stelle soll nur auf einige Aspekte der Schwierigkeit, mit dem Ähnlichkeits- und, dazu korrrespondierend, mit dem Distanzbegriff umzugehen, hingewiesen werden, damit deutlich wird, daß Modelle des multidimensionalen Skalierens keineswegs als allgemeine Modelle der Urteilsbildung oder der Wahrnehmung gelten können, wie von einigen Vertretern dieser Forschungsrichtung behauptet wird.

Um die Ideen zu fixieren, soll zunächst eine Definition von "Ähnlichkeit" eingeführt werden, die zumindest die Symmetrieforderung erfüllt: es seien a und b zwei Objekte, die durch Eigenschaftsmengen A und B definiert seien. Weiter sei m ein Maß auf diesen Mengen. Dann sei

$$s(a,b) = \frac{m(A \cap B)}{m(A \cup B)} \tag{8}$$

ein Maß für die Ähnlichkeit von a und b.

Das Maß (8) scheint zuerst von Jaccard (1908) eingeführt worden zu sein (zitiert nach Gregson (1975)); man könnte also auch vom *Jaccard-Index* sprechen. Dem Maß zufolge ist die Ähnlichkeit nicht nur abhängig von der Menge der gemeinsamen Merkmale, wie sie durch $A \cap B$ gegeben ist, sondern sie hängt noch von der Menge der überhaupt gegebenen Merkmale ab, also von $A \cup B$. Diese Menge in das Ähnlichkeitsurteil einzubeziehen ist sicherlich vernünftig: steht man zwei Personen a und b gegenüber, die nur ein Merkmal gemeinsam zu haben scheinen, nämlich das, eine Brille zu tragen, so wird diese Gemeinsamkeit kaum zur Ähnlichkeit der beiden beitragen, denn relativ zur Gesamtmenge der wahrgenommenen Eigenschaften ist dieses Merkmal kaum von Gewicht. Liegt aber nur eine Beschreibung der beiden Personen vor und werden in dieser Beschreibung nur wenige Merkmale genannt, so kann das gemeinsame Merkmal "trägt Brille" eine große Ähnlichkeit bedeuten.

Es werde für einen Moment angenommen, Personen würden die Ähnlichkeiten zwischen Objekten gemäß (8) beurteilen, also den Wert eines Jaccard-Indexes angeben. Weiter werde angenommen, daß sie die psychologische Distanz zwischen den Objekten gemäß der Restleschen Aspektdistanz (6) angeben würden. Dann sieht man leicht, daß Ähnlichkeiten und Distanzen nicht durch eine streng monotone Beziehung miteinander verknüpft sind. Denn aus $m(A \cup B) = m(A) + m(B) - m(A \cap B)$ und (7) folgt

$$\frac{1}{s} = \frac{d - m(A \cap B)}{m(A \cap B)} = \frac{d}{m(A \cap B)} - 1$$

Betrachtet man nun eine Menge von Objekten ω_i mit jeweilig verschieden zusammengesetzten Aspektmengen A_i und B_i, so folgt aus dieser Beziehung, daß der Wert von s nicht notwendig kleiner wird, wenn der von d steigt, und umgekehrt. Es genügt, die Menge von Paaren von Objekten zu betrachten, für die $d_{ij}/m(A_i \cap B_j) = c$, c eine Konstante, ist. Denn dann variiert der Wert der Distanz, aber der der Ähnlichkeit bleibt konstant. Der Grund dafür ist, daß die Aspektdistanz nicht nur vom Durchschnitt $A \cap B$, sondern auch von den Maßen $m(A_i)$ und $m(B_j)$ abhängt.

Ganz offenbar folgen Ähnlichkeits- und Distanzurteile keinen Mechanismen, die für alle Personen für alle Situationen und alle Objekte gleichartig ablaufen. Wie eine urteilende Person konzeptualisiert, was sie als "ähnlich" oder "unähnlich" klassifiziert, wird von den situativen Randbedingungen der Urteilsbildung abhängen. Die allgegenwärtigen Gestalteffekte werden überdies dazu führen, daß die einzelnen Merkmale nicht unabhängig voneinander in die Urteile eingehen. Generell ist dieser Sachverhalt Psychologen schon lange bekannt, und insofern ist es bemerkenswert, daß er im Zusammenhang mit Fragen der Skalierung relativ spät explizit diskutiert wurde, nachdem eine Reihe bahnbrechender und einflußreicher Arbeiten bereits erschienen war und in vieler Hinsicht bereits diskutiert worden ist.

Eine solche Diskussion wurde von Tversky (1977) vorgelegt. Tverskys Arbeit kann hier nicht ausführlich diskutiert werden, es soll nur auf einige wesentliche Aspekte hingewiesen werden.

Wie Restle (1959) geht auch Tversky von den Aspektmengen A und B aus, die die Objekte a und b charakterisieren. Wie die Aspektdistanz soll die Ähnlichkeit der Objekte vom Durchschnitt $A \cap B$ sowie von den Differenzen $A - B$ und $B - A$ abhängen. Weiter sei F eine Funktion, über die zunächst nichts weiter ausgesagt wird. Jedenfalls kann dann die Ähnlichkeit ganz allgemein wie folgt definiert werden.

1. Ein ordinales Maß für die Ähnlichkeit ist

$$s(a, b) = F(A \cap B, A - B, B - A) \tag{9}$$

   Für s wird zunächst nur die Ordinalskaleneigenschaft gefordert.

2. *Monotonizität:* $s(a, b) \geq s(a, c)$ genau dann, wenn $A \cap B \supset A \cap C$ und $A - B \subset A - C$ und $B - A \subset C - A$.

3. *Unabhängigkeit:* Die Objektpaare (a, b), (c, d) einerseits und (a', b') und (c', d') andererseits seien identisch hinsichtlich der gleichen zwei Merkmale, während die Paare (a, b) und (a', b') ebenso wie die Paare (c, d) und (c', d') sich hinsichtlich der verbleibenden dritten Eigenschaft gleichen. Dann soll

$$s(a, b) \geq s(a', b') \quad \text{genau dann, wenn} \quad s(c, d) \geq s(c', d')$$

4. *Lösbarkeit*: Die Menge der Merkmale seien hinreichend reichhaltig.

5. *Invarianz:* Intervalle sollen über verschiedene Faktoren gleich bleiben.

Sind diese Bedingungen erfüllt, so läßt sich zeigen (Tversky (1977)), daß Intervallskalen S und f existieren, die die Bedingungen

$$\begin{aligned} S(a, b) &\geq S(c, d) \quad \text{genau dann, wenn} \quad s(a, b) \geq s(c, d) & (10) \\ S(a, b) &= \Theta\, f(A \cap B) - \alpha\, f(A - B) - \beta\, f(B - A) & (11) \end{aligned}$$

wobei $\Theta, \alpha, \beta \geq 0$ freie Parameter sind. Das durch die Skalen S und f definierte Ähnlichkeitsmaß nennt Tversky das *Kontrastmodell*. Die Funktion f korrespondiert zum Maß m.

Das Kontrastmodell umfaßt eine Klasse von Modellen. Für $\Theta = 1, \alpha = \beta = 0$ ergibt sich

$$S(a, b) = f(A \cap B)$$

d.h. die Ähnlichkeit hängt nur von den gemeinsamen Merkmalen von a und b ab. Andererseit sei $\Theta = 0$, $\alpha = \beta = 1$. Dann ist

$$-S(a, b) = f(A - B) + f(B - A)$$

d.h. $-S(a, b) = d(a, b)$, wobei $d(a, b)$ durch die Aspektdistanz (6) gegeben ist.

Statt des Modells (11) kann auch das Modell

$$S(a, b) = \frac{f(A \cap B)}{f(A \cap B) + \alpha f(A - B) + \beta f(B - A)} \tag{12}$$

betrachtet werden. In diesem Fall ist $S(a,b)$ auf das Intervall $0 \leq S(a,b) \leq 1$ normalisiert. Tversky nennt das Modell (12) das *Verhältnismodell* (ratio model). Für $\alpha = \beta = 1$ ergibt sich als Spezialfall der Jaccard-Index (8).

Tverksy referriert eine Reihe von empirischen Untersuchungen, um Aspekte seiner Modelle zu untersuchen. So erhob er etwa Daten zur Ähnlichkeit von Nationen. Urteile, denen zufolge Nord-Korea Ähnlichkeit zu China hat, China aber kaum Ähnlichkeit zu Nord-Korea zeigt, sind die Regel. Es folgen Untersuchungen zur Ähnlichkeit von Buchstaben, eine Analyse der Daten von Rothkopf (1957), der die Verwechlungshäufigkeit von Morsesignalen untersuchte, sowie der Ergebnisse von Rosch (1973, 75) über wahrnehmungsmäßige und semantische Kategorien; alle diese Daten erlauben eine Interpretation in Termen des Kontrast- bzw. des Verhältnismodells. Diese Modelle sind wesentlich allgemeiner als solche, die nahezu automatisch bei Untersuchungen unterstellt werden, in denen Objekte multidimensional skaliert werden. Gleichwohl muß festgehalten werden, daß auch hier eigentlich nur ein allgemeiner Rahmen bereitgestellt wird, innerhalb dessen Ähnlichkeitsurteile diskutiert werden können; man schließt gewissermaßen von einem solchen Modell auf die Prozesse. Eigentlich sollte man umgekehrt vorgehen, nämlich von einem Modell der Prozesse auf das Modell bezüglich der Urteile schließen. Es muß allerdings zugegeben werden, daß dieser Weg vorerst noch der weitaus beschwerlichere ist.

Die Welt der Ähnlichkeiten ist offenbar komplex, und deshalb kann man mit gutem Recht vermuten, daß die Welt der psychologischen Distanzen die gleiche Eigenschaft hat. Es ist deswegen klar, daß die in den folgenden Abschnitten behandelten Skalierungsmodelle nur Spezialfälle allgemeiner Wahrnehmungs- und Beurteilungsprozesse sein können, auch wenn der Anspruch auf Allgemeinheit bei ihrer ursprünglichen Formulierung diese Problematik verschleierte.

3 Zur Wahl spezieller Metriken

Beals, Krantz und Tverksy (1968) haben allgemeine Bedingungen formuliert, die erfüllt sein müssen, damit eine Menge von Objekten in einem metrischen Raum repräsentiert werden können. Es sollen hier nur die wesentlichen Punkte dieser Arbeit genannt werden, da sie verdeutlichen, daß eine solche Repräsentation keineswegs generell gerechtfertigt ist.

Eine der Grundforderungen ist, daß der psychologische Raum in guter Näherung kontinuierlich sein muß, also nicht grundsätzlich eine "körnige", also diskrete Struktur hat. Wie bei der Diskussion der Skala für die Schwere von Verbrechen bereits angedeutet, ist diese kontinuierliche Struktur keine triviale Eigenschaft, insbesondere dann nicht, wenn es um die Repräsentation semantischer Strukturen geht (wenn also etwa Wörter hinsichtlich der Ähnlichkeit ihrer Bedeutung skaliert werden sollen). Beals et al. führen aus, daß die Kontinuumseigenschaft aber nicht notwendig als Eigenschaft der abzubildenden Objekte gesehen werden müsse, sondern als Eigenschaft des *einbettenden Systems*; in der Tat kann man ja Objekte hinsichtlich ihrer Länge vermessen, ohne daß man dabei fordert, daß es für jede mögliche Länge irgendwo im Universum auch ein Objekt der betrachteten Art gibt. überdies könne der semantische Raum kontinuierlich sein, ohne daß die Wörter in diesem Raum eine kontinuierliche Menge bilden müßten; im übrigen, meinen die Autoren, solle sich der Leser seine eigene Meinung dazu bilden.

3.1 Additive Segmente und Konturen gleicher Ähnlichkeit

Beals et al. fügen zunächst den üblichen Axiomen für eine Metrik eine vierte hinzu:

M4: *Segmentelle Additivität:* Es seien x und z irgendzwei Punkte in dem Raum, in dem die in Frage stehenden Objekte repräsentiert werden sollen. Dann existiert eine Menge Y von Punkten, die umkehrbar eindeutig auf die Elemente eines Abschnitts I der Zahlengeraden abgebildet werden kann derart, daß x und y die Endpunkte dieses Intervalles sind, und t_1 und t_2 sind zwei Variablen derart, daß $y_1(t_1), y_2(t_2) \in I$, dann $d(y_1, y_2) = |t_1 - t_2|$.

Die Forderung bedeutet, daß für $y \in Y$,

$$d(x, z) = d(x, y) + d(y, z)$$

erfüllt ist. Diese Gleichung erklärt den Ausdruck *Segmentelle Additivität*. Durch M4 werden gewisse Einschränkungen für die Wahl der Metrik definiert.

Beispiel 3.1 Als Beispiel betrachte man die Punkte auf einem Kreisbogen (Beals et al., p.129). Die Abstände können einmal durch die normale euklidische Metrik erklärt werden; die Distanzen zwischen den Punkten entsprechen dann der Länge der Sehnen, die die Punkte miteinander verbinden. Eine andere, mögliche Metrik ist die Länge des kürzeren Bogens zwischen den Punkten auf dem Kreis. Je länger die Sehne, desto länger ist ebenfalls das verbindende Kreissegment; man sagt, die beiden Metriken seien *ordinal äquivalent*. Hat man allerdings statt eines Kreises eine Ellipse, so erweisen sich die beiden Metriken nicht mehr als ordinal äquivalent. Eine gegebene Menge von Distanzen läßt nur noch höchstens eine der beiden Metriken zu.

Betrachtet man nun irgenddrei Punkte auf dem Kreisumfang, so ist die Summe der Distanzen zwischen den äußeren Punkten und dem mittleren Punkt stets gleich der Distanz zwischen den äußeren Punkten. Diese Aussage überträgt sich nicht auf die Distanzen, die durch die Sehnen gegeben sind. Der Grund dafür ist, wie Beals et al. ausführen, die Tatsache, daß man hier eben nur Punkte auf dem Kreisumfang betrachtet. Punkte im Inneren des Kreises sind ja nicht "erklärt" worden. Bei den gängigen Metriken wie der oben angegebenen euklidischen Metrik wird aber davon ausgegangen, daß Punkte, die die Additivitätseigenschaft ermöglichen, stets gegeben sind. M4 is die Forderung nach der Existenz solcher Punkte. Es soll also ausgeschlossen werden, daß die Punkte nur auf bestimmten (Hyper-) Ebenen liegen können. □

Wie die Autoren weiter zeigen, ist der Begriff der *Kontur gleicher Ähnlichkeit* (isosimilarity contour, isosimilarity sphere) von zentraler Bedeutung für die Annahme der Existenz einer Metrik. Der Begriff soll deshalb ausführlich definiert werden:

Definition 3.1 *Es seien a, b, c, ... Objekte, die in einem metrischen Raum repräsentiert werden können derart, daß die Distanz zwischen den entsprechenden Punkten des Raumes die psychologische Unähnlichkeit der Objekte abbildet*[2]*. Mit* δ *werde die Unähnlichkeit bezeichnet, und die Schreibweise* $\delta(a, b) \sim \delta(a', b')$ *soll anzeigen, daß die Objekte* a, b *einerseits und* a', b' *andererseits gleich unähnlich sind. Dann ist eine* Kontur gleicher Ähnlichkeit mit dem Zentrum a und dem Radius $\delta(a, b)$ *die Menge aller Punkte (= Objekte)* a'*, für die* $\delta(a, a') \sim \delta(a, b)$ *gilt. Zwei Konturen gleicher Ähnlichkeit heißen* konzentrisch*, wenn sie das gleiche Zentrum a haben.*

Anhand der Begriffe "segmentelle Additivität" und "Kontur gleicher Ähnlichkeit" läßt sich der Begriff der Homogenität eines Raumes charakterisieren. Dazu betrachte man zwei konzentrische Konturen gleicher Ähnlichkeit um das Objekt a. Jedes Objekt c auf der äußeren Kontur definiert dann in bezug auf a eine Richtung im Raum. Weiter sei die Forderung nach segmenteller Additivität erfüllt. Auf der Verbindungslinie von a nach c liegt dann ein Objekt b derart, daß $d(a, c) = d(a, b) + d(b, c)$. Die Dreiecksungleichung impliziert dann, daß es keinen anderen Punkt als b auf der Kontur gleicher Ähnlichkeit gibt, der näher an c liegt; die Distanz $d(b, c)$ ist gerade die Differenz der Radien der beiden Konturen gleicher Ähnlichkeit. Diese Differenz sollte die gleiche sein für alle Richtungen, d.h. für *alle* Punkte c auf der äußeren Kontur.

Ist diese Richtungs- und Ortshomogenität nicht gegeben, so existiert auch keine Metrik, d.h. es können keine Distanzen zwischen den Objekten gefunden werden, die den Axiomen der Metrik

[2] Da die Objekte eindeutig jeweils einem Punkt zugeordnet werden, wird zur Vereinfachung der Sprache im folgenden nicht zwischen einem Objekt und dem es repräsentierenden Punkt unterschieden.

genügen. Umgekehrt läßt sich eine Metrik durch die Form einer Kontur gleicher Ähnlichkeit veranschaulichen; ein Beispiel wird für die Minkowski-Metrik weiter unten gegeben.

3.2 Intra- und interdimensionale Charakterisierungen

Das Ziel der multidimensionalen Skalierung besteht darin, die interessierenden Objekte in einem Koordinatensystem zu beschreiben, dessen Achsen gewisse latenten Dimensionen sind, durch die die Objekte beschrieben werden können. Um zu einer solchen Repräsentation zu gelangen, muß erklärt werden, wie die Unterschiede der Objekte bezüglich ihrer Koordinaten auf den einzelnen Dimensionen die insgesamt wahrgenommenen Unterschiede oder Unähnlichkeiten zwischen den Objekten bestimmen. Es seien a_i, b_i, c_i, ... die Werte der Koordinaten der Objekte a, b, c, ... auf der i-ten Dimension. Beals et al. stellen zwei Forderungen auf:

1. *Intradimensionale Subtraktivität*: Der Beitrag der i-ten Dimension zur GesamtunÄhnlichkeit sei durch
$$\phi_i(a_i, b_i) = \phi_i(|a_i - b_i|) \tag{13}$$
gegeben, wobei ϕ_i eine monoton wachsende Funktionen der Beträge der Differenzen $a_i - b_i$ ist.

2. *Interdimensionale Additivität* Die Distanz zwischen a und b sei durch
$$d(a, b) = F\left(\sum_{s=1}^{r} \phi_s(a_s, b_s)\right) \tag{14}$$
gegeben.

Beals et al. formulieren dann eine Reihe notwendiger und hinreichender Bedingungen, die erfüllt sein müssen, damit interdimensionale Additivität und intradimensionale Subtraktivität angenommen werden können. Ein empirischer Test dieser Bedingungen erlaubt es dann, zu entscheiden, ob diese beiden Eigenschaften angenommen werden dürfen oder nicht. Da es in den meisten praktischen Situationen sehr schwierig sein dürfte, diese Bedingungen empirisch zu überprüfen, soll auf eine detaillierte Darstellung dieser Bedingungen an dieser Stelle verzichtet werden. Statt dessen soll eine Metrik eingeführt werden, die - sind die genannten Bedingungen erfüllt - für die praktische Arbeit in Frage kommt.

3.3 Die Minkowski-Metrik

Definition 3.2 *Die Objekte werden in bezug auf r Dimensionen $D_1, \cdots, D_r$ miteinander verglichen; das Objekt $a \in \Omega$ habe die Ausprägungen $a_1, \cdots, a_r$ auf den Dimensionen $D_1, \cdots, D_r$, und $b \in \Omega$ habe die Ausprägungen $b_1, \cdots, b_r$. Es gelte*
$$d(a, b) = \left[\sum_{s=1}^{r} |a_s - b_s|^p\right]^{1/p}, \tag{15}$$
wobei die reelle Zahl $p > 0$ ein freier Parameter ist. Dann ist durch d eine Minkowski-Metrik[3] *definiert.*

Theorem 3.1 *Durch (15) wird eine Metrik definiert.*

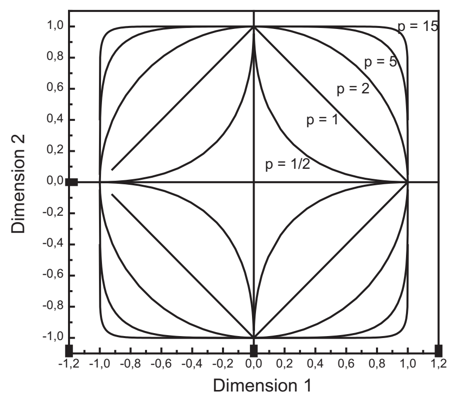
[3]Hermann Minkowski (1964 – 1909), Mathematiker und Lehrer Einsteins. Er führte diesen Distanzbegriff in Zusammenhang mit seinem Vorschlag, Zeit und Raum zu einem 4-dimensionalen Raum-Zeit-Kontinuum zusammenzufassen, ein. Einstein übernahm diesen Begriff für seine Allgemeine Relativitätstheorie.

Beweis: Die Gültigkeit von $d(a,b) \geq 0$ und $d(a,b) = d(b,a)$ ist offenkundig. Es bleibt zu zeigen, daß die Dreiecksungleichung gilt. Der Nachweis ist ein wenig länglich und soll hier übergangen werden (s. Kolmogoroff und Fomin (1970)). □

Offenbar erfüllt die Minkowski-Metrik die Bedingungen der intradimensionalen Subtraktivität und der interdimensionalen Additivität. Die Minkowski-Metrik wurde zuerst von Kruskal (1964) in die Diskussion der multidimensionalen Skalierung eingebracht.

Faßt man für einen Moment p als freien Parameter auf, so kann man sagen, daß die Minkowski-Metrik eigentlich eine Klasse von Metriken definiert, die sich eben durch den Wert von p unterscheiden; in Abb. 1 wurden die Spezialfälle $p = 1$ und $p = 2$ bereits illustriert. Die Konturen gleicher Ähnlichkeit für verschiedene Werte von p wird in Abb. 2 gezeigt. Einige wichtige Spe-

Abbildung 2: Minkowski-Metriken und Konturen gleicher Ähnlichkeit



zialfälle sollen kurz beschrieben werden.

3.3.1 Der Fall p = 1

In diesem Fall gilt

$$d(a,b) = \sum_{s=1}^{r} |a_s - b_s| \tag{16}$$

Die wahrgenommene Unähnlichkeit entspricht demnach einfach der Summe der bei den einzelnen Dimensionen wahrgenommenen Unterschieden.

Das Modell (16) wurde bereits von Housholder und Landahl (1945) vorgeschlagen; diese Autoren argumentierten, daß das Nervensystem registirierte Unterschiede aufaddiere, um die Unähnlichkeit zweier Stimuli zu repräsentieren. Attneave (1950) griff das Modell in einer Diskussion des Ähnlichkeitsbegriffs wieder auf und gab ihm den Namen "City-Block-Modell", - aus offenkundigen Gründen. Denn man gelangt von einem Ort (=Objekt) a zu einem anderen b nur, indem man "Wege" parallel den Koordinatenachsen geht. Dies gleicht der Situation einer Person in Manhattan, die von Punkt a zu Punkt b gelangen will: liegen die beiden Punkte in einer Straße, so muß man nur geradeaus gehen, liegen sie nicht in einer Straße, so muß man Straßen durchlaufen, die stets in einem rechten Winkel zueinander liegen.

Das City-Block-Modell ergibt sich u.a. dann, wenn die Objekte hinsichtlich verschiedener, sich nicht überschneidender (disjunkter) Eigenschaftsmengen beurteilt werden.

Diese Interpretation korrespondiert zu der von Shepard (1964), der argumentierte, daß Metriken mit p-werten in der Nachbarschaft von 1 dann zu erwarten seien, wenn die Aufmerksamkeit der Beurteiler zwischen verschiedenen Dimensionen, hinsichtlich derer sich die Objekte unterscheiden, flukutiert.

3.3.2 Der Fall $p = 2$

In diesem Fall ergibt sich

$$d(a,b) = \left[\sum_{s=1}^{r}(a_s - b_s)^2\right]^{1/2} \tag{17}$$

also die bereits bekannte euklidische Metrik.

Es wird oft argumentiert, daß die euklidische Metrik dann den Urteilen über die Unähnlichkeit von Objekten unterliegt, wenn die Objekte in einem "diffusen" Merkmalsraum definiert sind. Gemeint sind damit Merkmale, die wie Tonhöhe, Helligkeit, Klangfarbe oder Farbton etc. kontinuierlich variieren (Torgerson (1958, 1965), Shepard (1964), Hyman and Well (1967)). Der folgende Satz mag zur Plausibilität dieser Aussage beitragen:

Theorem 3.2 *Es sei Ω eine Menge von Objekten, die durch Punkte in einem Raum repräsentiert werden können, der durch eine Minkowski-Metrik mit dem Parameter p charakterisiert sei. Der Wert ist nur dann invariant gegenüber Rotationen der Achsen des Raumes, wenn die Metrik euklidisch ist, d.h. wenn $p = 2$.*

Beweis: Der Einfachheit halber werde der Fall $r = 2$ angenommen. Es werde $x_i := a_i - b_i$ gesetzt. $d(a,b)$ kann als Wert einer Funktion der $r = 2$ Variablen x_1, x_2 aufgefaßt werden. Die Rotation des Achsensystems bedeutet, daß die x_i-Werte in bestimmte, vom Rotationswinkel α abhängige y_i-Werte übergehen. Die Transformationsformeln sind bekanntlich durch

$$\begin{aligned} y_1 &= x_1 \cos\alpha + x_2 \sin\alpha \\ y_2 &= x_2 \cos\alpha - x_1 \sin\alpha \end{aligned} \tag{18}$$

gegeben. Nach der Rotation ist dann die Minkowski-Distanz durch

$$d_\alpha(a,b) = (y_1^p + y_2^p)^{1/p}$$

gegeben, wobei der Index α bei d anzeigen soll, daß es sich um die Distanz nach der Rotation des Achsensystems um α handeln soll. Setzt man für y_i die Ausdrücke aus (18) ein, so erhält man

$$d_\alpha(a,b) = [(x_1 \cos\alpha - x_2 \sin\alpha)^p + (x_2 \cos\alpha - x_1 \sin\alpha)^p]^{1/p}.$$

Die Distanz zwischen zwei Objekten soll durch die Rotation nicht geändert werden, d.h. es wird $d_\alpha(a,b) = const$ für alle α gefordert. Dann muß die Ableitung $d_\alpha(a,b)/d\alpha = 0$ sein für alle α. Führt man die Differentiation nach α durch und vereinfacht die entstehenden Ausdrücke, so sieht man, daß $d_\alpha(a,b)/d\alpha = 0$ nur dann, wenn die Gleichung $(x_1 \cos\alpha + x_2 \sin\alpha)^{p-2} = (x_2 \cos\alpha - x_1 \sin\alpha)^{p-2}$ für alle x_1, x_2 und für alle α erfüllt ist; dies ist offenbar nur dann der Fall, wenn $p = 2$. $\square$

Dieser Satz impliziert, daß es nicht *bestimmte* Dimensionen (= Rotationen) sind, die die Distanz zwischen zwei Objekten festlegen, wenn $p = 2$. Für $p \neq 2$ gilt dann gerade die Umkehrung; in diesem Fall sind es bestimmte Dimensionen, die den Unterschied zwischen den Objekten erzeugen. Zur Illustration sei noch einmal an die Thurstonesche Skalierung von Verbrechen hinsichtlich ihrer Schwere erinnert. Nimmt man an, daß die Urteile aufgrund der Bewertung hinsichtlich einiger latenter Dimensionen zustande kommen, und nimmt man weiter an, daß diese Dimensionen wie unabhängige Prädiktoren in einer multiplen Regression wirken, so erscheint ja der Vektor der Skalenwerte als Linearkombination der Vektoren, die die latenten Dimensionen abbilden. Für den "latenten" Raum ist dann eine euklidische Metrik definiert. Die prinzipielle Austauschbarkeit der Basisdimensionen hat dabei aber zu begrifflichen Schwierigkeiten geführt; die Annahme einer nicht-euklidischen Metrik für den semantischen Raum könnte diese Schwierigkeiten auflösen.

In einer euklidischen Metrik können die Objekte durch Vektoren $v_a = (a_1, \cdots, a_r)'$, $v_b = (b_1, \cdots, b_r)'$ etc. abgebildet werden. Die Distanz zwischen den Objekten ist dann durch die Distanz zwischen den Endpunkten der Vektoren gegeben; insbesondere ist

$$d^2(a,b) = \sum_{s=1}^{r}(a_s - b_s)^2 \tag{19}$$

Ausmultipliziert ergibt sich

$$d^2(a,b) = \sum_{s=1}^{r}(a_s^2 + b_s^2 - 2a_s b_s) = \sum_{s=1}^{r} a_s^2 + \sum_{s=1}^{r} b_s^2 - 2\sum_{s=1}^{r} a_s b_s$$

Dies ist aber gleichbedeutend mit

$$d^2(a,b) = |v_a|^2 + |v_b|^2 - 2v_a^{'} v_b \tag{20}$$

Dabei sind $|v_a|$ und $|v_b|$ die Längen der Vektoren und $v_a^{'} v_b$ ist das Skalarprodukt zwischen ihnen.

3.3.3 Die Supremumsmetrik, $p \to \infty$

Es sei

$$|a_0 - b_0| = \max_s |a_s - b_s| \tag{21}$$

d.h. $|a_0 - b_0|$ sei der Absolutbetrag der größten Koordinatendifferenz. Betrachtet werde

$$\left(\frac{d(a,b)}{|a_0 - b_0|}\right)^p = \sum_{s=1}^{r}\left(\frac{|a_s - b_s|}{|a_0 - b_0|}\right)^p$$

Offenbar gilt

$$\frac{|a_s - b_s|}{|a_0 - b_0|} \leq 1, \quad \text{für alle } s$$

Dann folgt[4]

$$\lim_{p\to\infty}\left(\frac{|a_s - b_s|}{|a_0 - b_0|}\right)^p = 0, \quad s \neq 0$$

und demnach

$$\lim_{p\to\infty}\left(\frac{d(a,b)}{|a_0 - b_0|}\right)^p \to 1,$$

also

$$d(a,b) \to |a_0 - b_0|. \tag{22}$$

[4]Sei $y \geq 0$; für $y \geq 1$, $\log y \geq 0$, und für $y < 1$ ist $\log y < 0$; insbesondere gilt $\log y \to -\infty$ für $y \to 0$. Nun sei $y = a^p$, mit $a < 1$, $p > 0$. Dann folgt $a^p \to 0$ für $p \to \infty$, denn $\log y = p \log a < 0$, da $\log a < 0$, und dementsprechend $\log y = p \log a \to -\infty$ für $p \to \infty$. Aber $\log y \to -\infty$ bedeutet, wie gerade gezeigt, $y = a^p \to 0$.

Für einen großen Wert des Parameters p ist die Distanz zwischen den Objekten a und b also im wesentlichen durch diejenge Differenz der Koordinatenwerte bestimmt, die von allen Diferenzen die größte ist. Gilt

$$d(a, b) = |a_0 - b_0|, \tag{23}$$

so spricht man auch von der *Supremumsmetrik*. Natürlich ist ein großer p-Wert nicht die einzige Bedingung, unter der die Supremumsmetrik gilt; es ist möglich, daß die Beurteiler einfach ihre Urteile nach der maximalen Diffferenz ausrichten. Dieser Fall ist dann von dem eines großen p-Wertes nicht zu unterscheiden.

3.3.4 Der Fall $p \to 0$

Der Fall $p = 0$ ist natürlich nicht definiert, da $1/p$ dann nicht erklärt ist. Aber der Fall $p \to 0, p \neq 0$ kann betrachtet werden.

Sicherlich gilt für $p \to 0$.

$$|a_s - b_s|^p \to 1 \quad \text{für alle } s$$

Dann folgt

$$d(a, b) \to r \tag{24}$$

d.h. es gilt $d(a, b) \approx r$ für alle a und b. Die Unähnlichkeit ist einfach durch die Anzahl der beurteilten Dimensionen gegeben und für alle Paare von Objekten gleich groß; der Wert der Differenzen auf den einzelnen Dimensionen geht in die Urteile nicht mehr ein.

3.3.5 Der allgemeine Fall

Eine interessante Interpretation der Minkowski-Metrik wurde von Wender (1968) vorgeschlagen. Es sei $v_a = (a_1, \cdots, a_s)'$ der Vektor der Koordinaten des Punktes a in einem Raum mit einer Minkowski-Metrik mit dem Parameter p. Die Länge dieses Vektors ist gleich dem Abstand, d.h. der Distanz, zwischen dem Ursprung des Koordinatensystems und dem Punkt a, und gemäß der allgemeinen Minkowski-Metrik durch

$$d(0, a) = |v_a| = \left(\sum_{s=1}^{r} |a_s|^p \right)^{1/p} \tag{25}$$

gegeben. Dann ist

$$|v_a|^p = \sum_{s=1}^{r} |a_s|^p.$$

Aber es ist sicherlich $|v_a|^p = |v_a|^{p-1}|v_a|$ und analog dazu $|a_s|^p = |a_s|^{p-1}|a_s|$, so daß

$$d(0, a) = |v_a| = \sum_{s=1}^{r} \left(\frac{|a_s|}{|v_a|} \right)^{p-1} |a_s|$$

Setzt man nun

$$g_s := \left(\frac{|a_s|}{|v_a|} \right)^{p-1} \tag{26}$$

so erhält man

$$d(0, a) = \sum_{s=1}^{r} g_s |a_s|. \tag{27}$$

Diese Gleichung besagt, daß sich der Abstand eines Punktes vom Nullpunkt durch eine *gewogene Summe* der Absolutbeträge der Koordinaten auffassen läßt. Eine analoge Betrachtung gilt für beliebige Distanzen $d(a,b)$. Aus (26) folgert man, daß für kleine Werte von a_s das Gewicht g_s klein wird, wenn p große Werte hat. Die Distanzen $d(0,a)$ oder allgemein $d(a,b)$ sind dann im wesentlichen durch die großen Differenzen $|a_s|$ bzw. $|a_s - b_s|$ bestimmt. Die großen Differenzen ziehen gewissermaßen den größten Teil der Aufmerksamkeit auf sich, die kleinen werden vernachlässigt. Die Wendersche Interpretation deckt sich also mit derjenigen, derzufolge der Wert von p den Grad der Aufmerksamkeit widerspiegelt, der den Dimensionen zukommt, und diese Aufmerksamkeit richtet sich nach der Größe der Unterschiede zwischen den Koordinaten.

Es sollen noch die Gewichte (26) für kleine Werte von p betrachtet werden. Der erste Fall sei $p \to 2$; aus (26) folgt dann sofort $g_s \to |a_s|/|v_a|$ für alle s, so daß das Gewicht einer Dimension proportional zur Ausprägung (bzw. zur Differenz im Falle $d(a,b)$) eines Merkmals ist. Für $p \to 1$ folgt $g_s \to 1$, d.h. alle Dimensionen haben das gleiche Gewicht. Für $p < 1$ schließlich folgt

$$g_s = \left(\frac{|a_s|}{|v_a|}\right)^{p-1} = \left(\frac{|v_a|}{|a_s|}\right)^{1-p}$$

d.h. die Gewichte nehmen Werte umgekehrt proportional den Ausprägungen $|a_s|$ an, $a_s \neq 0$ vorausgesetzt.

Was diese Aussagen über die Gewichte g_s bedeuten, kann nur anhand der Objekte diskutiert werden, die man bei einer speziellen Untersuchung betrachtet; die vorangegangene Diskussion der Minkowski-Metrik kann nur allgemeine Linien der Interpretation andeuten. Die Ziele der speziellen Untersuchung lassen dann auch eine genauere Untersuchung der Frage zu, wieweit eine mehrdimensionale Skalierung von Objekten überhaupt Einblick in die Wahrnehmungsstruktur bzw. in die kognitiven Prozesse der Beurteilung liefert. Vielleicht liefert eine mehrdimensionale Skalierung eine Beschreibung, die bestimmte Aspekte der Urteilsbildung aufscheinen läßt; ohne eine unabhängig von der Skalierung formulierte Theorie der Prozesse oder ohne gezielte Fragestellungen etwa bezüglich des Wertes von p wird man kaum viel Gewinn aus einer mehrdimensionalen Skalierung ziehen.

Auf spezifische Probleme der Repräsentation der Struktur psychologischer "Objekte" weisen Tversky und Hutchinson (1986) hin. Die Autoren zeigen, daß geometrische Modelle eine für das spezielle Modell spezifische obere Grenze der Anzahl der Punkte definieren, die *den gleichen nächsten Nachbar* haben. Der nächste Nachbar des i-ten Objekts ist das j-te Objekt, wenn gilt

$$d(i,j) \leq d(i,k) \quad \text{für alle} \quad k \neq j$$

Das j-te Objekt ist dann dem i-ten am ähnlichsten, oder wird am häufigsten mit ihm verwechselt, etc.. Die Relation "ist nächster Nachbar" ist nicht symmetrisch: für das i-te Objekt kann das j-te der nächste Nachbar sein, aber das i-te Objekt muß nicht der nächste Nachbar für das j-te Objekt sein. Tversky et al. weisen darauf hin, daß in einer Dimension ein Punkt höchstens zwei nächste Nachbarn haben kann, in zwei Dimensionen höchstens fünf, in drei Dimensionen höchstens 11, etc.. Sie zeigen anhand verschiedener Datensätze, daß insbesondere bei begrifflichen Stimuli (= "Objekten") ihre geometrische Repräsentation erhebliche Verzerrungen aufweist: eine Ursache dafür ist, daß diese Art von Objekten oft eine hierarchische Struktur hat, etwa Frucht → Südfrucht → (Orange, Banane, ...), Frucht → Apfel, etc.. Die Problematik tritt kaum auf bei Objekten der unmittelbaren Wahrnehmung, wie Farben, Klängen etc.. Diese Befunde zeigen, daß die oben angeratene Vorsicht sich nicht nur aus epistemologischer Moral ergibt.

4 Die Skalierung von Distanzen

Bei der multidimensionalen Skalierung werden latente Dimensionen zur Erklärung von Distanzmaßen bestimmt. Solche Maße können auf verschiedene Weise zustande kommen:

1. Durch direkte Schätzung von Unähnlichkeiten δ_{ij} (*psychologischen Distanzen*) für Paare (ω_i, ω_j) von Objekten.

2. Durch direkte Schätzung von Ähnlichkeiten σ_{ij} für Paare (ω_i, ω_j) von Objekten. Diese Schätzungen σ_{ij} müssen in Distanzmaße transformiert werden; dazu wird eine monoton fallende Funktion f mit $f(\sigma_{ij}) = \delta_{ij}$ angenommen.

3. Durch indirekte Schätzung von Ähnlichkeiten oder Distanzen über bedingte Häufigkeiten $h(x_i|y_j)$ und $h_(y_j|x_i)$ (*Konfusionen*). So kann y_j ein bestimmter Stimulus sein, und die Aufgabe der Vp kann darin bestehen, den Stimulus zu identifizieren, wobei sie ihn mit anderen aus einer vorgegebenen Menge, etwa mit x_i, verwechseln kann. In einem anderen Fall können die y_j und x_i Personen sein, und $h(x_i|y_j)$ ist die Häufigkeit, mit der y_j sich an x_i wendet.

4. Durch Urteile über das Ausmaß, in dem ein Merkmal in einem anderen "enthalten" ist.

4.1 Torgersons Methode der vollständigen Triaden

Varianten der direkten Schätzung – etwa auf Ratingskalen – sind Methoden wie die der *vollständigen Triaden* (Torgerson (1958)), bei der die Vp Aussagen der Form "Stimulus ω_i ist dem Stimulus ω_j ähnlicher als ω_k" entweder zustimmt oder nicht. Aus der relativen Häufigkeit der "Ja"-Antworten lassen sich dann Distanzmasse d_{ij}, d_{ik} und d_{jk} herleiten. Denn wenn die Unterschiede zwischen den Objekten nicht allzu groß sind, werden die genannten Urteile nicht mit Sicherheit gefällt, so daß die Urteile nur zu einem Teil $\hat{p}^i_{jk}$ von der obengenannten Form sind, und zu einem komplementären Teil $\hat{p}^i_{kj}$ von der Form "Stimulus ω_i ist dem Stimulus ω_k ähnlicher als ω_j". Dann wird angenommen, daß etwa

$$\hat{p}^i_{jk} = \Phi(\hat{d}_{ij} - \hat{d}_{ik}) \tag{28}$$

ist, wobei Φ eine Wahrscheinlichkeitsfunktion ist; speziell wird die *Normalverteilung* angenommen, d.h.

$$\hat{p}^i_{jk} = \Phi(\hat{d}_{ij} - \hat{d}_{ik}) = \frac{1}{2\pi}\int_{-\infty}^{\hat{d}_{ij}-\hat{d}_{ik}} \exp(-\frac{z^2}{2})\, dz.$$

Die Beziehung (28) erlaubt die Bestimmung der Differenzen $\hat{d}_{ij} - \hat{d}_{ik}$; es ist

$$\Phi^{-1}(\hat{p}^i_{jk}) = \hat{d}_{ij} - \hat{d}_{ik} \tag{29}$$

Die Details der Methode findet man in Torgerson (1958). Der Nachteil dieses Ansatzes besteht darin, daß eben nur die Differenzen (29) von Distanzen geschätzt werden können, was bedeutet, daß die Distanzen nur in ihrer relativen Größe zueinander geschätzt werden können, nicht aber ihr absoluter Wert, d.h. die Schätzung der Distanzen hat nur Intervallskalenqualität. Torgerson spricht dementsprechend von (29) als von den *komparativen Distanzen*. Darauf wird weiter unten noch eingegangen. Ein weiterer Nachteil besteht in der großen Zahl von Vergleichen, die eine Person durchführen muß, denn für n Objekte gibt es $n(n-1)(n-2)/6$ mögliche Dreiergruppen von Objekten (Triaden); für $n = 7$ etwa muß die Vp zwar erst 35 Urteile abgeben, für $n = 10$ sind es dann aber bereits 120. Die unvermeidlichen Aufmerksamkeitsfluktuationen sind dann eine nicht zu unterschätzende Fehlerquelle.

4.2 Bedingte Häufigkeiten und Konfusionen

Liegen nur Häufigkeiten von Konfusionen zwischen Objekten vor, so muß von ihnen zu Schätzungen für Distanzen übergegangen werden. Hierzu gibt es verschiedene Ansätze. Luce (1963) schlug das Modell

$$h(y_j|x_i) = \frac{\sigma(x_i, y_j)b(y_j)}{\sum_r \sigma(x_i, y_r)b(y_r)} \tag{30}$$

vor; hierin ist $\sigma(x_i, y_j)$ die Ähnlichkeit zwischen x_i und y_j und $b(y_j)$ ist einfach ein Bias, mit dem die Vp sich für y_j entscheidet. Die Asymmetrie zwischen $h(y_j|x_i)$ und $h(x_i|y_j)$ wird also durch Einführen des Bias erklärt, denn das Maß σ wird als symmetrisches Maß angenommen; vermutlich, weil der angenommene Begriff der Ähnlichkeit die Annahme einer solchen Symmetrie nahelegt, nicht, weil sich die Symmetrie der Ähnlichkeit aus einem Modell über Wahrnehmungsfunktionen oder kognitive Prozesse ergäbe. Darüber hinaus wird angenommen, daß

$$-\log \sigma(x_i, y_j) = d(x_i, y_j) \tag{31}$$

ein Distanzmaß ist, das die Bedingungen einer Metrik erfüllt. In anderen Worten, es soll

$$\sigma(x_i, y_j) = e^{-d(x_i, y_j)} \tag{32}$$

gelten. Für eine weiterführende Diskussion dieses Ansatzes muß auf Luce (1963) verwiesen werden.

Shepard (1958) gelangt über Betrachtungen der Reizgeneralisierung zu einem Ausdruck, in dem die bedingten Häufigkeiten für Entscheidungen für ein bestimmtes Objekt ebenfalls über die Exponentialfunktion auf die Distanz bezogen werden. Der Einfachheit halber werde

$$h_{ij} = h(y_j|x_i), \quad h_{ji} = h(x_i|y_j)$$

gesetzt. Shepard (1958) zeigt dann, daß unter bestimmten Bedingungen

$$\sqrt{\frac{h_{ij}h_{ji}}{h_{ii}h_{jj}}} = e^{-k\, d_{ij}} \tag{33}$$

wobei k eine Konstante und d_{ij} die symmetrische Distanz zwischen x_i und y_j ist. Man bemerke, daß die linke Seite von (33) im wesentlichen ein geometrisches Mittel der Verwechslungshäufigkeiten und deshalb ein (symmetrisches) Maß für die Ähnlichkeit der Objekte ω_i und ω_j ist (auf die Reihenfolge der Zahlen bei der Produktbildung kommt es nicht an!). Im übrigen muß man sagen, daß die Shepardsche Herleitung von (33) auf empirisch kaum überprüfbaren Annahmen beruht, so daß man (33) auch gleich einfach als *Annahme* einführen kann.

4.3 Urteile des "Enthaltenseins"

Für eine Reihe von Merkmalen können Aussagen der Art "Das Merkmal a ist zum Anteil p im Merkmal b enthalten" gemacht werden, oder Aussagen der Form "Ist x_i gegeben, dann ist mit der Wahrscheinlichkeit p auch y_j gegeben". Solche Aussagen können ebenfalls zu einem Distanz- oder Ähnlichkeitsmaß für die Merkmale a und b führen.

4.3.1 Der Ansatz von Hays (1958)

Es seien $a_1, \cdots, a_n$ Objekte oder Merkmale. Aussagen der obigen Form können dann als Angaben der Form $p(a_j|a_i)$ oder $p(a_i|a_j)$ gedeutet werden, wobei h durch p ersetzt wurde, um den umgangssprachlichen Gebrauch des Wortes "Wahrscheinlichkeit"anzudeuten (natürlich können auch wirkliche Wahrscheinlichkeiten eingesetzt werden). Weiter sei A_i die Aspektmenge des Objekts a_i. Hays setzt dann

$$p(a_j|a_i) = \frac{m(A_i \cap A_j)}{m(A_j)} \tag{34}$$

und $p(a_i|a_j)$ ist natürlich analog definiert; m ist ein Maß auf den Mengen A_i. Aus (34) folgt sofort

$$\frac{p(a_j|a_i)}{p(a_i|a_j)} = \frac{m(A_j)}{m(A_i)} \tag{35}$$

Summiert man diese Ausdrücke über alle i, so erhält man

$$\sum_{i=1}^{n} \frac{p(a_j|a_i)}{p(a_i|a_j)} = \sum_{i=1}^{n} \frac{m(A_j)}{m(A_i)} = m(A_j) \sum_{i=1}^{n} \frac{1}{m(A_i)} \tag{36}$$

Dies gilt für ein beliebiges Objekt A_j; man setzt nun

$$c = \sum_{i=1}^{n} \frac{1}{m(A_i)} \tag{37}$$

c ist eine Konstante für alle $m(A_j)$, und damit erhält man aus den $p(A_i|A_j)$ eine Abschätzung des Maßes $m(A_j)$, für alle j. Wegen (34) folgt dann aber auch $c\,m(A_i \cap A_j) = c\,p(a_j|a_i)\,m(A_i)$. Aufgrund der Definition der Aspektdistanz erhält man dann die Schätzung

$$\hat{d}_{ij} = c\,d_{ij} = c\,m(A_i) + c\,m(A_j) - 2c\,m(A_i \cap A_j) \tag{38}$$

Der unbekannte Faktor wird dabei als für alle Distanzen gleich angenommen, so daß die Distanzen zwischen a_i und a_j bis auf einen Skalenfaktor (c) bestimmt sind.

Hays (1958) interpretiert die Distanzen als euklidische Distanzen. Die Ableitung von (38) impliziert aber an keiner Stelle, daß diese Interpretation notwendig ist. Vermutlich hat Hays die Möglichkeit, andere als euklidische Distanzen zu betrachten, nicht gesehen; Kruskal (1964) war wahrscheinlich der erste, der die allgemeine Minkowski-Distanz in die Diskussion brachte.

4.3.2 Ekman und Lindmans (1961) Modell

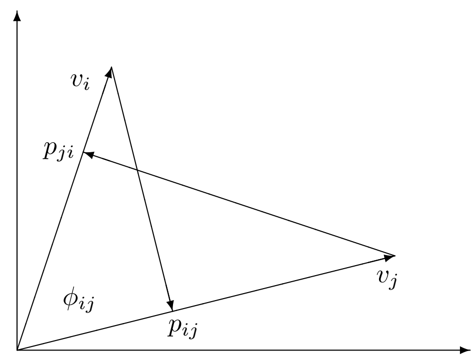
Mehr der Vollständigkeit halber und um die Möglichkeiten des Modellierens aufzuzeigen sei noch kurz auf das Modell von Ekman und Lindman (1961) eingegangen. Es wird angenommen, daß die zu beurteilenden Stimuli in einem mehrdimensionalen Raum mit einer euklidischen Metrik liegen. Wie im faktorenanalytischen Modell werden die Stimuli durch Vektoren $v_i, v_j, i, j = 1, \cdots, n$ repräsentiert. Die Ähnlichkeit zwischen zwei Stimuli ist immer abhängig von der Menge der Merkmale oder Aspekte, die die Stimuli gemeinsam haben, also davon, wieweit die Menge des einen Stimuli in der des anderen enthalten ist; darüber hinaus werden die Maße der Gesamtmengen eine Rolle spielen. Ekman et al. machen nun die Annahme, daß ein Maß für das Enthaltensein durch die Projektionen der Vektoren v_i und v_j, die die Stimuli σ_i und σ_j repräsentieren, gegeben sind sind (vergl. Abb. 3). Die Vektoren v_i und v_j bilden den Winkel ϕ_{ij}. Es sei p_{ij} die Projektion des Vektors v_i auf den Vektor v_j, und p_{ji} sei die Projektion des Vektors v_j auf den Vektor v_i. Aus der Definition des Kosinus folgt sofort

$$\cos\phi_{ij} = \frac{p_{ij}}{|v_i|} = \frac{p_{ji}}{|v_j|} \tag{39}$$

Die Versuchsperson (Vp) wird nun gebeten, Urteile der Form "Stimulus σ_i ist in Stimulus σ_j im Ausmaß q_{ij} enthalten." abzugeben. Der Wert q_{ij} wird dabei direkt als Prozentsatz geschätzt. Ebenso gibt die Vp das Urteil q_{ji} ab, also eine Schätzung des Ausmaßes, in dem der Stimulus σ_j im Stimulus σ_i enthalten ist. Formal entsprechen diese Aussagen den Gleichungen

$$q_{ij} = \frac{p_{ij}}{|v_j|}, \quad q_{ji} = \frac{p_{ji}}{|v_i|} \tag{40}$$

Abbildung 3: Vektorprojektionen



Damit ist eine direkte Beziehung zu den Kosinus in (39) gegeben: es ist

$$q_{ij}q_{ji} = \frac{p_{ij}}{|v_i|}\frac{p_{ji}}{|v_j|} = \frac{p_{ij}}{|v_j|}\frac{p_{ji}}{|v_i|}$$

d.h.

$$q_{ij}q_{ji} = \cos^2(\phi_{ij}) \tag{41}$$

bzw.

$$\cos\phi_{ij} = \sqrt{q_{ij}q_{ji}} \tag{42}$$

Die Kosinus $\cos\phi_{ij}$ entsprechen aber Korrelationskoeffizienten bis auf einen Faktor $(1/n)$ (vergl. Anhang) und können damit direkt faktorenanalysiert werden.

Die Herleitung von Beziehungen zwischen Ähnlichkeiten, Distanzen und Konfusionsmaßen $h(y_j|x_i)$ bzw. $p(a_j|a_i)$ sind alle plausibel, aber kaum eine basiert auf einem expliziten psychologischen Modell. Für viele Zwecke erscheint daher die direkte Schätzung von Distanzen als Unähnlichkeiten auch der direkteste Weg, zu Maßen zu gelangen, die als Ausgangspunkt für eine multidimensionale Skalierung genommen werden können.

4.4 Torgersons Modell

Es werde angenommen, daß die Distanzen euklidisch seien, und daß für gegebene n Reize die $\binom{n}{2}$ Distanzen vorliegen. Die Einbettung der Reizkonfiguration in einen Raum Ψ mit minimaler Dimensionalität kann dann mit einer von Torgerson (1962) vorgeschlagenen Methode gefunden werden.

Die Methode besteht im wesentlichen in einer Anwendung der Hauptachsentransformation, bei der ein durch latente Variablen definierter Raum über die Analyse von Korrelationen zwischen gemessenen Variablen bestimmt wird. Den Korrelationen entsprechen Skalarprodukte von Vektoren, und Torgerson geht davon aus, daß für irgend drei Punkte stets ein Skalarprodukt berechnet werden kann. Dieser Zusammenhang wird zuerst erläutert.

4.4.1 Korrelationen und Skalarprodukte

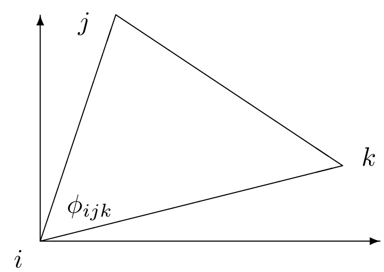
Gegeben seien n Reize, die durch n Punkte in einem r-dimensionalen Raum mit euklidischer Metrik repräsentierbar seien, wobei; dabei sei $r < n$. Die Distanzen zwischen den Punkten seien also euklidisch. Insbesondere betrachten wir irgend drei Objekte a_i, a_j und a_k, die durch die Punkte i, j und k repräsentiert seien. Die Distanzen zwischen den Punkten seien $d(i, j) = d_{ij}$,

$d(i,k) = d_{ik}$ und $d(j,k) = d_{jk}$. Für den Moment fassen wir die d_{ij}, d_{ik} und d_{jk} als Vektoren auf : d_{ij} sei der von i nach j zeigende Vektor (mit der Länge d_{ij}), d_{ik} sei der von i nach k zeigende Vektor mit der Länge d_{ik}, und d_{jk} sei der von j nach k zeigende Vektor mit der Länge d_{ik}. Der Winkel zwischen d_{ij} und d_{ik} sei ϕ_{ijk}. Nach dem Kosinussatz gilt

$$d_{jk}^2 = d_{ij}^2 + d_{ik}^2 - 2d_{ij}d_{ik}cos\ \phi_{ijk}, \tag{43}$$

In Abb. 4 ist die Beziehung zwischen den Vektoren für den 2-dimensionalen Fall dargestellt.

Abbildung 4: Kosinussatz



Der Punkt i ist durch die Koordinaten a_{i1}, a_{i2}, der Punkt j durch die Koordinaten a_{j1}, a_{j2} und k ist durch die Kordinaten a_{k1}, a_{k2} bestimmt. Nach dem Satz des Pythagoras gilt nun

$$\begin{aligned} d_{ij} &= [(a_{i1} - a_{j1})^2 + (a_{i2} - a_{j2})^2]^{1/2} \\ d_{ik} &= [(a_{i1} - a_{k1})^2 + (a_{j2} - a_{k2})^2]^{1/2} \\ d_{jk} &= [(a_{j1} - a_{k1})^2 + (a_{j2} - a_{k2})^2]^{1/2} \end{aligned} \tag{44}$$

Dann ist $d_{ij}^2 = a_{i1}^2 + a_{j1}^2 - 2a_{i1}a_{j1} + a_{i2}^2 + a_{j2}^2 - 2a_{i2}a_{j2}$, etc.. Setzt man die entsprechenden Ausdrücke in (43) ein und vereinfacht, so erhält man schließlich die Beziehung

$$a_{j1}a_{k1} + a_{j2}a_{k2} = d_{ij}d_{ik}cos\phi_{ijk}. \tag{45}$$

Darin ist aber $a_{j1}a_{k1} + a_{j2}a_{k2}$ nichts weiter als das Skalarprodukt der Vektoren (a_{j1}, a_{j2}) und (a_{k1}, a_{k2})! Für den allgemeinen Fall mit r Dimensionen erhält man dementsprechend als Verallgemeinerung von (45)

$$\sum_{s=1}^{r} a_{js}a_{ks} = d_{ij}d_{ik}\cos\phi_{ijk} \tag{46}$$

Man vergleiche in diesem Zusammenhang die Gleichung (43) mit dem Ausdruck (7). Die Gleichungen (45) und (46) besagen, daß $d_{ij}d_{ik}\cos\phi_{ijk}$ gleich dem Skalarprodukt der Vektoren ist, die die Punkte i, j und k im psychologischen Raum Ψ repräsentieren. Aus der Faktorenanalyse weiß man, daß die linke Seite von (46) gleich der Korrelation r_{jk} der gemessenen Variablen X_j und X_k ist. In der Tat läuft die Berechnung der "Ladungen" a_{is} auf eine Anwendung der Hauptkomponentenanalyse hinaus.

4.4.2 Bestimmung des Raumes Ψ

Aus (43) folgt

$$d_{ij}d_{ik}\cos\phi_{ijk} = \frac{1}{2}(d_{ij}^2 + d_{ik}^2 - d_{jk}^2) =: b_{i(jk)}; \tag{47}$$

wegen (46) ist bekannt, daß $b_{i(jk)}$ gleich dem Skalarprodukt der von i nach j bzw. k zeigenden Vektoren ist. Skalarprodukte entsprechen den Korrelationen zwischen den entsprechenden Variablen. Die $b_{i(jk)}$ können nun zu einer Matrix

$$B_i = (b_{i(jk)}) \tag{48}$$

zusammengefaßt werden, d.h. für jeden Punkt i können wir die Skalarprodukte der von i zu irgendzwei anderen Punkten weisenden Vektoren berechnen. Da, wie man sich leicht überlegt, $b_{i(jk)} = b_{i(kj)}$ ist, muß B_i eine symmetrische Matrix sein. Dann existiert aber eine Matrix A_i derart, daß

$$B_i = A_i A_i' \tag{49}$$

gilt. Die Matrix A_i enthält die a_{ij} für alle j; die a_{ij} können als Faktorladungen interpretiert werden und dementsprechend über die Faktorenanalyse ermittelt werden. Als Ursprung des Koordinatensystems hat man dann den Punkt (Reiz) i. Im Falle fehlerbehafteter Daten werden aber die so ermittelten Punktekonfigurationen für jedes i leicht verschieden sein, und man wählt deshalb, um die Fehler auszugleichen, den Schwerpunkt der Konfiguration als Ursprung des Koordinatensystems.

Um die Notation einzuführen, läßt man zunächst in (49) den Index i weg. Man hat dann für $B = AA'$ mit $B = (b_{jk})$

$$b_{jk} = \sum_{s=1}^{r} a_{js} a_{ks} \tag{50}$$

Die Verschiebung der Konfiguration derart, daß der Koordinatenursprung mit dem Ursprung des Koordinatensystems zusammenfällt, ist eine *Translation*. Für den Mittelpunkt gelte

$$\begin{aligned} B^* &= A^* A^{*'} \\ a_{js}^* &= a_{js} - c_s \end{aligned} \tag{51}$$

wobei c_s die Verschiebungskonstante auf der s-ten Dimension ist. Es muß gelten

$$c_s = \frac{1}{n} \sum_{j=1}^{n} a_{js} = \frac{1}{n} \sum_{k=1}^{n} a_{ks}, \quad b_{jk}^* = \sum_{s=1}^{n} a_{js}^* a_{jk}^* \tag{52}$$

Setzt man hier die Ausdrücke aus (51) ein, so erhält man

$$b_{jk}^* = \sum_{s=1}^{r} (a_{js} - c_s)(a_{ks} - c_s). \tag{53}$$

Es gilt nun der

Theorem 4.1 *Die Skalarprodukte b_{jk}^* für die Paare der Vektoren vom Mittelpunkt der Konfiguration zu den Punkten, die die Reize repräsentieren, lassen sich über die empirisch gewonnenen Distanzen zwischen den Reizen gemäß*

$$b_{jk}^* = \frac{1}{2} \left(\frac{1}{n} \sum_{j=1}^{n} d_{jk}^2 + \frac{1}{n} \sum_{k=1}^{n} d_{jk}^2 - \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} d_{jk}^2 - d_{jk}^2 \right) \tag{54}$$

Beweis: Den Beweis des Satzes findet man in Torgerson (1962), p. 259.

Aus den empirisch gewonnenen Distanzen können also direkt die Centroid-Skalarprodukte b_{jk}^* gewonnen werden. Die b_{jk}^* können als Korrelationen aufgefaßt werden und demgemäß faktorenanalysiert werden; diese Analyse liefert eine $r \times n$-Matrix von Ausprägungen der n Reize auf r orthogonalen Dimensionen. Die Dimensionen sind eindeutig bis auf Rotation.

Werden die Distanzen als *komparative Distanzen*, etwa nach der oben genannten Methode der Triaden, bestimmt, so muß berücksichtigt werden, daß eben nur Differenzen von Distanzen bestimmt werden; die Distanzen ergeben sich erst durch Addition einer additiven Konstanten, die aus den Daten geschätzt werden muß. Eine auf Messick und Abelson (1956) zurückgehende Schätzung wird ebenfalls in Torgerson (1962) dargestellt.

Bei Torgersons Skalierung wird im allgemeinen über die Urteile von Vpn gemittelt. Dies setzt voraus, daß die Vpn im Prinzip die Stimuli in der gleichen Weise sehen; Unterschiede in den Urteilen beruhen auf statistischen Fluktuationen. Es ist klar, daß Torgersons Ansatz zu einer mehrdimensionalen Skala führen kann, die nicht der Skala auch nur einer einzigen Vp entspricht.

Bei der Torgerson-Skalierung gibt es nur eine Art, individuelle Differenzen in der Sichtweise der Stimuli zu erfassen: man muß die Daten für jede befragte Person (Vp) getrennt skalieren. Dies setzt voraus, daß man von einer einzelnen Person genügend viele Urteile bekommt, um eine statistisch stabile Abschätzung der Distanzen zu erhalten. Mindestens ebenso schwerwiegend wie dieses Erhebungsproblem ist aber die Frage, wie unterschiedlich denn die Faktorenlösungen einzelner Vpn sein müssen, um einen systematischen Unterschied der Sichtweisen nahezulegen. Weiter muß man sich fragen, worin ein solcher Unterschied besteht, denn man kann ja partiell in der Sichtweise übereinstimmen, und partiell nicht übereinstimmen. Einen Ansatz zur Beantwortung dieser Fragen haben Tucker und Messick (1963) geliefert.

4.5 Die Verallgemeinerung von Tucker und Messick

Wie bei der Torgerson-Skalierung gehen wir von empirisch gewonnenen Distanzen zwischen den Stimuli aus. Hat man n Stimuli, so gibt es pro Vp $(n(n-1)/2)$ Distanzurteile (natürlich kann jede Vp die Paare mehrfach beurteilen, und die hier betrachteten Distanzen sind dann individuelle Mittelwerte). Es gebe N Vpn. Wir können dann eine Datenmatrix X aufstellen: X habe N Spalten und $m = n(n-1)/2$ Zeilen. Jede Zeile enthält die Distanzurteile der N Vpn für ein bestimmtes Reizpaar (j,k). Das Tucker-Messick-Modell besteht im Wesentlichen in einer Anwendung des Satzes über die Basisstruktur einer gegebenen Matrix. Ist X eine beliebige $m \times N$-Matrix mit Rang $rg(X) = r \leq \min(m, N)$ so existieren stets r linear unabhängige, m-dimensionale Vektoren derart, daß für den j-ten Spaltenvektor X_j von X die Darstellung

$$X_j = p_{j1}A_1 + \cdots + p_{jr}A_r. \tag{55}$$

gilt. Dabei sind die p_{jk}, $k = 1, \cdots, r$ "Gewichte", mit denen die Vektoren A_k in die Darstellung von X_j eingehen. Die A_k werden in einer Matrix A zusammengefaßt, $A = [A_1, \cdots, A_r]$, und die Koeffizienten p_{jk} in einer Matrix P. Dann kann (55) in der Form

$$X = AP' \tag{56}$$

ausgedrückt werrden, wobei $X = [X_1, \cdots, X_n]$ ist. Die Vektoren A_k müssen linear unabhängig sein, darüber hinaus ist man in ihrer Wahl frei. Deswegen ist es keine Einschränkung der Allgemeinheit, sie insbesondere orthogonal zu wählen. Dann erhält man sofort eine Identifikation der Matrix P. Denn bildet man das Produkt $X'X$, so ist nach (56)

$$X'X = PA'AP'. \tag{57}$$

Da die A_k als orthogonal vorausgesetzt werden, muß $A'A = \Lambda$ eine Diagonalmatrix sein, da sie ja die Skalarprodukte $A_k'A_s$ enthält, und $A_k'A_s = 0$ für $k \neq s$ und $A_k'A_s = \lambda_k$ für $s = k$, λ_k das Quadrat der Länge von A_k. Damit ist (57) äquivalent zu

$$X'X = P\Lambda P', \; \Lambda = \text{diag}(\lambda_1, \cdots, \lambda_r), \tag{58}$$

und dieser Ausdruck wiederum besagt, daß P die Matrix der Eigenvektoren und Λ die Diagonalmatrix der Eigenwerte von $X'X$ ist. Die Eigenwerte λ_k von $X'X$ entsprechen also den Quadraten der Längen der Basisvektoren A_k. Die Eigenvektoren in P können als normiert betrachtet werden; damit gilt $P'P = I$. Wären die P_k nicht normiert, so hätten sie Längen π_k und es wäre $P'P = \Pi = \text{diag}(\pi_1^2, \cdots, \pi_k^2)$. Ein Eigenvektor wird normiert, indem seine Komponenten mit $1/\pi_k$ multipliziert werden. In Matrixform erhält man die normierten Eigenvektoren durch Bildung des Produkts $\tilde{P} = P\Pi^{-1/2}$, so daß $P = \tilde{P}\Pi^{1/2}$ und $X = A\Pi^{1/2}\tilde{P}'$, $X'X = \tilde{P}\Pi^{1/2}A'A\tilde{\Pi}^{1/2}\tilde{P}' = \tilde{P}\Pi^{1/2}\Lambda\Pi^{1/2}\tilde{P}'$; darin ist $\Pi^{1/2}\Lambda\Pi^{1/2}$ wieder eine Diagonalmatrix $\text{diag}(\pi_1^2\lambda_1, \cdots, \pi_r^2\lambda_r)$. Die Betrachtung der normierten Eigenvektoren von $X'X$ ist also gleichbedeutend damit, daß man annimmt, daß die Vektoren A_k die quadrierten Längen $\pi_k^2\lambda_k$ haben. Die Länge der A_k ist aber für unsere Betrachtungen unwesentlich; man schreibt also wieder P statt $\tilde{P}$ für die normierten Vektoren und λ_k statt $\pi_k^2\lambda_k$. Gilt also $P'P = I$, so folgt aus (56) sofort $XP = A$, d.h. man kann die Matrix A ausrechnen, wenn man die Matrix P der Eigenvektoren von $X'X$ bestimmt hat. Die A_k haben nun irgendeine Länge, und man kann sie wiederum auf die Länge 1 normieren. Wie bei der Normierung der Eigenvektoren geschieht dies durch Multiplikation mit $\Lambda^{-1/2}$. Man bildet also die Matrix Q der normierten Basisvektoren gemäß

$$Q = A\Lambda^{-1/2}.$$

Dann folgt $A = Q\Lambda^{1/2}$ und (56) ist gleichbedeutend mit

$$X = Q\Lambda^{1/2}P'; \tag{59}$$

dies ist die *Grundstruktur* der Matrix X (vergl. Kap. 10, Satz 10.1). Bildet man die Matrix XX', so erhält man

$$XX' = Q\Lambda^{1/2}P'P\Lambda^{1/2}Q' = Q\Lambda Q', \tag{60}$$

d.h. Q ist die Matrix der Eigenvektoren der Matrix XX', und offenbar hat XX' die gleichen Eigenwerte wie $X'X$.

Es sei jetzt speziell X die Matrix der psychologischen Distanzen zwischen den Stimuli; X_j ist der $(n(n-1)/2)$-dimensionale Spaltenvektor der Distanzen der j-ten Vp. Die Darstellung (59) gilt für *jede* reelle Matrix X, also auch für diese Matrix von Distanzen. Die orthogonalen Vektoren $A_k = \sqrt{\lambda_k}Q_k$, Q_k der k-te, normierte Eigenvektor von XX', repräsentieren hier linear unabhängige Vektoren, aus denen sich die Spaltenvektoren X_j der Distanzen für die einzelnen Vpn durch Linearkombination darstellen lassen. Sinngemäß sind also die Komponenten der A_k ebenfalls als Distanzen zwischen den Stimuli aufzufassen. Damit repräsentieren die A_k gewissermaßen Sichtweisen (*viewpoints* bei Messick und Tucker) der Stimuli. Die lineare Unabhängigkeit der A_k bedeutet, daß sich keine dieser Sichtweisen A_k aus den anderen Sichtweisen A_s, $s \neq k$, herleiten läßt. Bestimmte Komponenten von A_k und A_s mögen vielleicht gleich sein (d.h. die Sichtweisen stimmen eventuell in bezug auf bestimmte Reize überein), aber insgesamt ergibt sich keine systematische Beziehung zwischen den Distanzen. Dabei kann es sein, daß A_1 einen 3-dimensionalen Raum, A_2 einen 4-dimensionalen Raum, A_3 einen 1-dimensionalen Raum repräsentiert, etc.. Die Konfigurationen der Punkte, die die Stimuli repräsentieren, können von einem Raum (einem Vektor A_k) zum anderen (A_s) völlig verschieden sein.

Die Darstellung der Distanzen für die j-te Vp, $X_j = p_{j1}A_1 + \cdots + p_{jr}A_r$ bedeutet nun, daß die Sichtweise der j-ten Vp sich aus diesen unabhängigen Sichtweisen $A_1, \cdots, A_r$ additiv zusammensetzt, und zwar mit einer für die j-te Vp charakteristischen Gewichtung der einzelnen Sichtweisen A_k, die durch den Vektor $(p_{j1}, \cdots, p_{jr})'$ gegeben ist. Dabei ist es natürlich möglich, daß X_j mit einer bestimmten Sichtweise A_k identisch ist; dann ist $p_{jk} \neq 0$ und $p_{js} = 0$ für alle $s \neq k$.

Wie bereits gesagt, charakterisieren die Vektoren $(p_{j1}, \cdots, p_{jr})'$, $j = 1, \cdots, N$ die Vpn, und p_{jk} ist das Gewicht, mit dem die k-te Sichtweise in die Urteile, d.h. in die Wahrnehmng, der j-ten Vp eingeht. Wir können die $p_{j1}, \cdots, p_{jr}$ als Koordinaten zur Repräsentation der Vpn durch Punkte im Raum der Sichtweisen interpretieren; auf diese Weise kann man Gruppen von Vpn mit ähnlichen Sichtweisen bestimmen. Die "Lösungen" A_k sind bekanntlich nur eindeutig bis auf Rotation; wir können die Matrix A also insbesondere so transformieren (= rotieren), daß einer der Vektoren genau durch den Mittelpunkt einer Gruppe läuft; die verschiedenen Sichtweisen werden dann in bezug auf die Sichtweise dieser speziellen Gruppe definiert. Ein solches Vorgehen ist von Interesse, wenn man die Sichtweisen bestimmter Gruppen miteinander vergleichen möchte; ebenso kann man die Veränderungen der Sichtweise einer Gruppe nach einer Intervention (Trainingsprogramm, Therapie, politische Ereignisse, etc.) untersuchen.

Die Frage, wie individuelle Unterschiede bei der mehrdimensionalen Skalierung berücksichtigt werden können, ist von anderen Autoren ebenfalls aufgenommen worden, etwa von Carroll und Chang (1970). Eine Zusammenfassung findet man bei Carroll (1972), wo auch eine Reihe von Beispielen gegeben werden. Die Modelle beschränken sich im wesentlichen ebenfalls auf die euklidische Metrik.

4.6 Nichtmetrische Skalierung

Der in vieler Hinsicht interessanteste Ansatz, Objekte durch Punkte in einem Raum mit minimal notwendiger Dimension einzubetten, geht auf Kruskal (1964a, 1964b) zurück: der Ansatz ermöglicht es, nicht nur die Anzahl der Dimensionen und die Projektionen der Punkte auf diese Dimensionen, sondern darüber hinaus den Parameter p der Minkowski-Metrik abzuschätzen.

Die Grundidee des Kruskalschen Verfahrens geht auf einen Ansatz Shepards (1962a) zurück, der sich allerdings auf die "Einbettung" der Punkte in einen r-dimensionalen Raum mit euklidischer Metrik beschränkte. Shepard ging von empirisch gegebenen Unähnlichkeiten oder Proximitäten δ_{ij} zwischen Objekten ω_i, ω_j, $i, j = 1, \cdots, n$ aus, wobei die Symmetrie

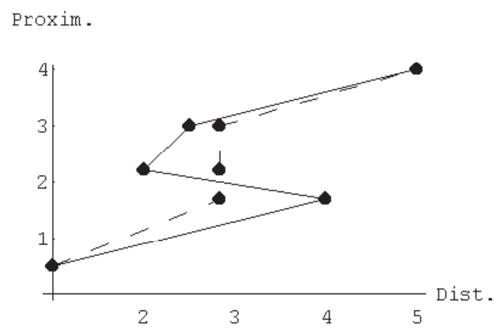
$$\delta_{ij} = \delta_{ji}$$

vorausgesetzt wird. Bei Maßen für die Unähnlichkeit (oder Ähnlichkeit) kann die Frage nach der Skalenqualität oft schwer zu beantworten sein, mit Sicherheit kann aber zumindest Ordinalskalenqualität vorausgesetzt werden. Shepards Abschätzung der Anzahl r der Dimensionen sowie der Koordinaten (Projektionen) der Punkte auf die Achse geht von nicht mehr als der Ordinalskalenqualität der Maße δ_{ij} aus; deswegen ist beim Shepardschen und dementsprechend beim Kruskalschen Ansatz auch von der *nichtmetrischen multidimensionalen Skalierung* die Rede. Dazu werden die δ_{ij} in eine Rangreihe gebracht: für n Objekte gibt es gerade $N = n(n-1)/2$ verschiedene Proximitäten δ_{ij}, so daß man die Rangreihe

$$\delta_{i_1,j_1} < \delta_{i_2,j_2} < \cdots < \delta_{i_N,j_N}$$

erhält. Ausgehend von einer im Prinzip beliebig - im übrigen aber möglichst günstig - gewählten Ausgangskonfiguration der repräsentierenden Punkte werden dann die Distanzen d_{ij} berechnet. Der Kern des Verfahrens besteht nun darin, die d_{ij} so lange in einer noch zu beschreibenden Weise zu verändern, daß sie bestmöglich mit den δ_{ij} übereinstimmen. Dazu muß natürlich ein Maß für die übereinstimmung definiert werden. Um das Verfahren etwas spezifischer zu beschreiben und um das Maß der übereinstimmung einführen zu können, soll zunächst das von Kruskal so genannte *Scattergram* eingeführt werden. Dazu werden die δ_{ij} gemäß ihrer Rangordnung vom kleinsten bis zum größten Wert auf der y-Achse eines Koordinatensystems aufgetragen. Für jeden Wert δ_{ij} wird der zugehörige Wert für d_{ij} auf der x-Achse aufgetragen. Verbindet man die Punkte miteinander, beginnend vom Punkt für das kleinste Proximitätsmaß δ bis zu dem mit dem größten Wert für δ, so ist eine notwendige Bedingung für eine gute Anpassung der d_{ij} an die δ_{ij}, daß die Verbindungslinie monoton steigt. Für die zunächst gewählte Konfiguration, d.h. für die zuerst berechneten d_{ij} wird das kaum je der Fall sein; man betrachte dazu die durchgezogene Linie in Abb. 5. Jede Nichtmonotonität weist darauf hin, daß die d_{ij} noch verändert werden müssen, damit die berechnete Konfiguration derjenigen entspricht, die implizit durch die δ_{ij} gegeben ist. Man betrachte in der Abb. 5 die Punkte (4, 1.7) und (2, 2.2). Die

Abbildung 5: Scattergramm



Verbindungslinie springt von (4, 1.7) zurück zu dem Punkt (2, 2.2). Dieser Rücksprung bildet eine Nichtmonotonität, denn $d_{ij} = 4$ ist im Vergleich zu $\delta_{ij} = 1.7$ zu groß, bzw. $d_{ij} = 2$ ist im Vergleich zu $\delta_{ij} = 2.2$ zu klein. Erst der Punkt (5, 4) ist im Vergleich zu (4, 1.7) wieder monoton ansteigend. Zumindest die Punkte (4, 1.7), (2, 2.2) und (2.5, 3) müssen durch eine geeignete Veränderung der d-Werte, also von 4, 2.2 und 2.5, verändert werden. Diese Veränderung muß so sein, daß insgesamt vom ersten Punkt - (1, .5) - bis zum letzten Punkt - (5, 4) - eine monotone Beziehung zwischen den δ- und den d-Werten entsteht. Diese d-Werte müssen also ersetzt werden durch andere, die zu einer monotonen Beziehung führen. Diese neuen Werte sollen, für eine gegebene Menge von d-Werten, mit $\hat{d}$ bezeichnet werden. Die Idee ist jetzt, für die d_{ij}-Werte, die eine Nichtmonotonität bilden, *einen* $\hat{d}$-Wert zu finden; dieser $\hat{d}$-Wert entspricht den drei übereinanderliegenden Punkten in 5, - sie haben ja den gleichen Abzissenwert, eben $\hat{d}$; darüber hinaus transformieren sie die Kurve in eine monotone.

$\hat{d}$ muß aus den d-Werten, die die Nichtmonotonität bilden, geschätzt werden. Die Schätzung muß zwei Forderungen genügen: (i) sie muß die Kurve (Verbindungslinie der Punkte) so transformieren, daß sie monoton wird, (ii) sie soll die die Nichtmonotonität erzeugenden Distanzen d_{ij} so gut wie möglich repräsentieren. Punkt (i) ist unmittelbar klar, er ist ja die Motivation für die Einführung des Wertes $\hat{d}$. Punkt (ii) wird klar, wenn man bedenkt, daß man die d_{ij} natürlich so wenig wie möglich verändern möchte, um so schnell wie möglich zu einer akzeptablen Lösung, d.h. Schätzung der d-Werte, zu gelangen. Dementsprechend fordert Kruskal (1964), daß die Größe

$$S = \sum_{i<j} (d_{ij} - \hat{d}_{ij})^2 \tag{61}$$

minimal werden soll. Die Summation bezieht sich dabei nur auf die Distanzen, die die Nichtmonotonität erzeugen. (61) hat die Form einer Varianz, und die wird bekanntlich minimal, wenn $\hat{d}$ das arithmetische Mittel der d_{ij}-Werte in (61) ist; mithin

$$\hat{d} = \frac{1}{m} \sum_{i<j} d_{ij}, \tag{62}$$

wobei m die Anzahl der Distanzen ist, die die Nichtmonotonität erzeugen.

Natürlich ist es möglich, daß die Abbildung der δ_{ij} gegen die d_{ij} mehr als nur eine Nichtmonotonität zeigt. Für jede Menge von d_{ij}-Werten, die im eben erklärten Sinn eine Nichtmonotonität erzeugen, wird gemäß (61) ein entsprechender $\hat{d}$-Wert bestimmt. Die Gesamtgüte der Anpassung der durch die d_{ij}-Werte definierten Konfiguration an die implizit durch die δ_{ij}-Werte gegebene Konfiguration wird dann durch die gemäß (61) erklärte Gesamtsumme über alle Nichtmonotonitäten ausgedrückt:

$$S^* = \sum_{i<j}(d_{ij} - \hat{d}_{ij})^2 \tag{63}$$

wobei die Indizierung der $\hat{d}$-Werte jetzt anzeigen soll, daß jedem d_{ij}-Wert, der eine Nichtmonotonität erzeugt, ein bestimmter $\hat{d}$-Wert entspricht.

Der Wert von S^* hängt noch von Dehnungen oder Stauchungen der durch die d_{ij} gegebenen Konfiguration ab. Dieser Sachverhalt stört, denn es kommt ja nur darauf an, daß die Relationen zwischen den d_{ij} stimmen. Die Division von S^* durch $\sum_{i<j} d_{ij}^2$ bedeutet eine Normierung von S^*, durch die das Gütemaß unabhängig von Dehnungen oder Stauchungen der Konfiguration wird. Dementsprechend hat man die

Definition 4.1 *Die Größe*

$$S = \sqrt{\frac{\sum_{i<j}(d_{ij} - \hat{d}_{ij})^2}{\sum_{i<j} d_{ij}^2}} \tag{64}$$

heißt Stress *der durch die* d_{ij} *definierten Konfiguration.*

Entspricht die berechnete Konfiguration der tatsächlich vorliegenden perfekt, so werden die d_{ij}- und die $\hat{d}_{ij}$-Werte identisch und der Stress wird Null. Man wird also versuchen, für gegebenen r- und p-Wert die d_{ij} so zu bestimmen, daß S minimal wird.

Die hierzu verwendete Methode ist aus der numerischen Mathematik als *Gradientenabstiegsmethode* bekannt. Das Verfahren soll nur kurz angedeutet werden, Details findet man in Lehrbüchern der numerischen Mathematik. Da entsprechende Computer-Programme existieren, kommt es nur darauf an, die Grundidee zu kennen.

Der Wert von S hängt von den d_{ij} ab, die wiederum durch die angenommenen Koordinaten der Punkte gegeben sind. Die Veränderung der d_{ij} zur Minimalisierung von S wird also durch eine Veränderung der Koordinaten x_{is} der $i = 1, \cdots, n$ Punkte auf den $s = 1, \cdots, r$ Dimensionen bewirkt. Die Koordinaten der n Punkte können als Komponenten *eines* Vektors x angeschrieben werden:

$$x = (x_{11}, \cdots, x_{1r}, \cdots, x_{n1}, \cdots, x_{nr})'$$

und S kann als Funktion von x aufgefaßt werden,

$$S = S(x).$$

Dann kann der Vektor der partiellen Ableitungen von S nach den x_{is} gebildet werden:

$$\partial S = \left(-\frac{\partial S}{\partial x_{11}}, \cdots, -\frac{\partial S}{\partial x_{nr}} \right)$$

Dies ist der *Gradient* von S; er zeigt stets in die Richtung der größten Veränderung von S. Die x_{is}-Werte werden dieser Richtung entsprechend verändert. Dieser Prozess wird so oft wiederholt, bis S sich nicht mehr verändert; man hat dann ein Minimum für S erreicht. Insgesamt wird dieses Vorgehen für verschiedene r- und p-Werte wiederholt; man entschließt sich dann für diejenige Konfiguration, für die sich insgesamt der kleinste Wert für S ergeben hat.

Eine Reihe von Beispielen findet man in Romney, Shepard und Nerlove (1972).

Ein generelles Problem bei dem Versuch, eine allgemeine Minkowski-Metrik zu finden, besteht darin, daß Algorithmen wie der Kruskalsche in lokale Minima laufen, und daß das Kruskalsche Stress-Maß die Güte der Anpassung für verschiedene Werte des Minkowski-Exponenten

p nicht gleichmäßig abbildet. Eine Möglichkeit, dieses Problem zu verringern, besteht darin, zunächst eine euklidische Metrik an die Daten anzupassen und die so gefundene Konfiguration als Startkonfiguration für die Suche nach einer allgemeinen Minkowski-Metrik zu wählen.

4.7 Abschließende Bemerkungen

Wie bei der Faktorenanalyse stellt sich die Frage nach der Natur der Dimensionen, die dem Modell der multidimensionalen Skalierung zufolge als latente Variable die Beziehungen zwischen den Objekten definieren. Grundsätzlich läßt sich sagen, daß diese Dimensionen nicht notwendig psychologisch real sind. Denn sie ergeben sich zunächst nur als Resultat einer Art zusammenfassender Beschreibung der psychologischen Distanzen, Ähnlichkeiten oder Konfusionshäufigkeiten. Darüber hinaus *können* sie tatsächlich als Beurteilungsdimensionen existieren. Die Axiomatik von Beals, Krantz und Tversky (1968) unterstellt diese Art von Existenz und ist insofern ein spezielles Modell der Generierung solcher Urteile.

Die Kernannahmen des Modells von Beals, Krantz und Tversky (1968) sind (i) die " Zerlegbarkeit" der Beschreibung der Stimuli (Objekte) in Aspekte oder eben "Dimensionen", (ii) intradimensionale Subtraktivität und (iii) interdimensionale Additivität. Zerlegbarkeit bedeutet, daß es keine Interaktionen zwischen den Dimensionen gibt, - sie werden unabhängig voneinander beurteilt. Die Minkowski-Metrik drückt diese Eigenschaft direkt aus, denn es treten z.B. keine Produktterme mit Koordinaten verschiedener Dimensionen auf, alle Dimensionen gehen nur additiv in das Distanzurteil ein. Es ist nicht klar, warum es im subjektiven Raum keine Interaktionen zwischen verschiedenen Dimensionen geben soll. Die Analyse nach dem Shepard-Kruskalschen Verfahren etwa liefert stets eine Repräsentation der Objekte in einem mehrdimensionalen Raum mit einer bestimmten Minkowski-Metrik, d.h. einem bestimmten Metrik-Parameter p, die Frage ist aber, ob dieser Befund dann bedeutet, daß die beurteilenden Personen auch nach diesem Modell gehandelt oder wahrgenommen haben. Die im Zusammenhang mit der Faktorenanalyse (vergl. Kap. 10) diskutierten Arbeiten von Overall (1964) und Armstrong (1967) legen die Vermutung nahe, daß dieser Schluß nicht notwendig gerechtfertigt ist.

Wender (1971) ist die Frage, ob das Modell von Beals et al. korrekt ist, direkt angegangen. Dazu zeigte er seinen Vpn Paare von Rechtecken, die sich hinsichtlich der Dimensionen (a) Größe der Fläche, (b) Form unterschieden, wobei "Form" als das Verhältnis von Länge und Breite definiert war. Die Aufgabe bestand darin, für ein gegebenes Paar die Distanz (Unterschiedlichkeit) auf einer Rating-Skala anzugeben ("0" für "kein Unterschied" und "8" für "völlig verschieden"). In einem zweiten Experiment wurden jeweils zwei Paare gezeigt und die Aufgabe bestand darin, anzugeben, welches der beiden Paare als unterschiedlicher erschien. Wender wandte nun kein Programm an, um die Dimensionalität und den Minkowski-Exponenten p zu schätzen, sondern er untersuchte die Urteile direkt. Denn die Stimuluspaare waren so konstruiert, daß sich etwa die Ratings für bestimmte Teilklassen von Paaren nur zufällig unterscheiden sollten, wenn die Bedingungen der Zerlegbarkeit und der intradimensionalen Subtraktivität gelten. In ähnlicher Weise konnten die Paarvergleiche getestet werden. Es zeigte sich, daß beim Paarvergleich die Bedingungen bei fast allen Vpn verletzt waren, und bei den Ratings immer noch bei mehr als der Hälfte der Vpn.

Für Psychologen, die sich mit der visuellen Wahrnehmung auskennen, ist das Ergebnis nicht überraschend, kennen sie doch die allgegenwärtigen Gestalteffekte. Die Frage ist, warum solche Effekte bei anderen "Objekten" nicht auftreten sollen. Die Frage ist angesichts der Tatsache, daß die multidimensionale Skalierung durchaus venünftige Resultate liefert (man betrachte etwa die Skalierung von Nationen von Wish, Deutsch und Biener (1972), die von Adjektivkombinationen (Cliff (1972)) oder die von semantischen Strukturen (Rapoport und Fillenbaum (1972)). Insbesondere die euklidische Repräsentation erscheint hier als eine sinnvolle Beschreibung der Objekte. Vermutlich darf man nicht schließen, daß die Dimensionen in sich nicht weiter zerlegbare Einheiten sind; eher dürfte es sich um irgendwelche nichtlineare Funktionen elementarer Größen handeln. Die Spekulation über solche Größen anhand einer Skalierung ist aber vermutlich müßig, ein direkterer psychologischer Ansatz dürfte für eine diesbezügliche Theorienbildung fruchtbarer sein.

Die Frage nach der Symmetrie der Ähnlichkeits- oder Distanzurteile ist natürlich ebenfalls kritisch. Die Arbeit Tverskys (1977) und die daran anknüpfenden Untersuchungen zeigten, daß keinesfalls selbstverständlich von der Symmetrie dieser Größen ausgegangen werden kann. Es erwies sich aber lange als einigermaßen schwierig, den Tverskyschen Ansatz in einen Algorithmus umzusetzen, der die "Features" und ihre Maße für eine gegebene Menge von Objekten lieferte. Carroll und Arabie (1980) verweisen auf einige frühe Arbeiten, etwa Krumhansl (1978). Ein neuerer Ansatz wurde von DeSarbo et al. (1992) vorgelegt. Diese Arbeit enthält neben einem Modell, nicht-symmetrische psychologische Distanzen zu skalieren, einen überblick über die wesentlichen Arbeiten zum Modell von Tversky (1977). Aus Platzgründen kann an dieser Stelle nicht weiter darauf eingegangen werden. In Carroll und Arabie (1980) finden sich ebenfalls Hinweise auf Arbeiten bezüglich anderer Metriken, z.B. bezüglich der Riemann-Metriken. Weiteren Einblick in Modelle der multidimensionalen Skalierung findet man in Borg (1981) und Borg und Lingoes (1987).