

# UNIVERSITÄT TRIER

FORSCHUNGSBERICHTE

zum  
MARKETING

Nr. 8  
Trier 2009

**Robert Hörstrup – Daniel Mühlhaus**

**Der Einfluss von Untersuchungsrahmen und  
Messmethodik auf die Güte von MDS-Konfigurationen**

Herausgeber: Univ.-Prof. Dr. Rolf Weiber

## **Forschungsbericht Nr. 8:**

Hörstrup, R./Mühlhaus, D. (2009):

Der Einfluss von Untersuchungsrahmen und Messmethodik auf die Güte von MDS-Konfigurationen, Forschungsbericht Nr. 8, hrsg. von Rolf Weiber, Trier.

## **Über die Autoren:**

Dipl.-Kfm. Robert Hörstrup und Dipl.-Volksw. Dipl.-Kfm. Daniel Mühlhaus sind wissenschaftliche Mitarbeiter an der Professur von Univ.-Prof. Dr. Rolf Weiber für Betriebswirtschaftslehre, insb. Marketing und Innovation ([www.innovation.uni-trier.de](http://www.innovation.uni-trier.de)) an der Universität Trier.

## **Kontaktadresse**

Univ.-Prof. Dr. Rolf Weiber  
Universität Trier  
Lehrstuhl für Marketing, Innovation und E-Business  
Fachbereich IV: BWL-AMK  
Universitätsring 15  
D-54286 Trier  
Tel.: 0049-201-2619  
Fax: 0049-201-3910  
E-Mail: [marketing@uni-trier.de](mailto:marketing@uni-trier.de)  
Internet: [www.innovation.uni-trier.de](http://www.innovation.uni-trier.de)

Copyright: Eigenverlag der Professur für Marketing, Innovation und E-Business  
an der Universität Trier  
Univ.-Prof. Dr. Rolf Weiber  
Trier 2009

ISBN 3-930230-27-5

## Vorwort des Herausgebers

Die Multidimensionale Skalierung (MDS) zählt zu den Standardinstrumenten bei der Analyse von Wahrnehmungen und besitzt im Marketing vor allem bei der Positionierungsanalyse eine herausragende Bedeutung. Der vorliegende Forschungsbericht konzentriert sich auf die Analyse der **Gestaltung des Untersuchungsdesigns** einer MDS und geht dabei der Frage nach, inwieweit bereits die Ausgestaltung des Untersuchungsrahmens und die Wahl der Messmethodik die Validität von MDS-Konfigurationen beeinflussen können.

Zu diesem Zweck werden zunächst die in der Literatur diskutierten Ansätze zur Steigerung der Datenqualität bei der Erhebung von Ähnlichkeitsdaten systematisiert. Dabei wird deutlich, dass vor allem die Wahl der Skalierung, die Berücksichtigung von Antwortunsicherheit und die Vorabnennung der Beurteilungsobjekte wesentlichen Einfluss auf die späteren MDS-Ergebnisse nehmen können. Entsprechend werden diese Aspekte in einer anschließenden zweistufigen empirischen Studie mit 132 Studierenden der Universität Trier einer genaueren Betrachtung unterzogen: In der ersten Erhebung wurden bei den Studierenden Ähnlichkeitsurteile zu unterschiedlichen Objektgruppen mit **Variation der Skalierungsmethodik** (5-, 7-, 9-stufige Skalen und Magnitudeskalierung) und **Variation des Untersuchungsrahmens** (mit und ohne Vorabnennung der Untersuchungsobjekte sowie Variation der Objektpositionen) erhoben. Auf dieser Basis wurden dann die individuellen MDS-Konfigurationen ermittelt und diese denselben Probanden in der zweiten Erhebung zur **Evaluation** vorgelegt. Aufgrund dieser zweistufigen Vorgehensweise ist es möglich, eine **Validitätsprüfung der MDS-Konfigurationen** über die übliche Verwendung von Stress-Maßen und Korrelationen hinaus zu erreichen, wodurch wertvolle Hinweise für die Gestaltung von MDS-Untersuchungsdesigns gewonnen werden konnten. Insgesamt zeigte sich, dass die **Magnitudeskalierung** eine besondere Eignung für die Erhebung von Ähnlichkeitsurteilen aufweist und die Erhebung paarweiser Ähnlichkeitsdaten als relativ robust anzusehen ist. Sie weist weder Positions-, noch Reihenfolgeeffekte auf. Auch zeigt die Beurteilungsunsicherheit hier keine negativen Effekte auf die Untersuchungsergebnisse.

Ein besonderer Dank gilt an dieser Stelle den Studierenden der Universität Trier, die über ihre engagierte Mitarbeit und das "zeitaufwändige" Bearbeiten der iteriert vorgelegten Fragebögen eine empirische Prüfung der dargestellten Ansätze und Konzepte überhaupt erst ermöglichten. Darüber hinaus gilt ein Dank aber auch den wissenschaftlichen Hilfskräften der Professur für Marketing, Innovation und E-Business, Frau cand. rer. pol. Katharina Ferreira und Herrn cand. rer. pol. Philip Wegmann, die einzelne Textfassungen durchgearbeitet, konstruktive Verbesserungsvorschläge unterbreitet und bei der Erstellung einer druckfähigen Fassung dieses Forschungsberichtes mitgewirkt haben.

Dem Forschungsbericht selbst wünsche ich nicht nur in der Scientific Community, sondern auch in der Marktforschungspraxis eine weite Verbreitung, da bei Beachtung der hier gewonnenen Erkenntnisse die Validität von mit der metrischen MDS gewonnenen Ergebnissen deutlich gesteigert werden kann.

Trier, im März 2009

Rolf Weiber

## Inhaltsverzeichnis:

1	Bedeutung der Multidimensionalen Skalierung zur Analyse von Wahrnehmungen und Gestaltungsbereiche bei der Erstellung von MDS-Untersuchungsdesigns..	3
2	Struktur des verwendeten Untersuchungsdesigns .....	6
2.1	Notwendigkeit und Aufbau des zweistufigen Untersuchungsdesigns.....	6
2.2	Eignung einer probandenbezogenen Ergebnisevaluation .....	8
3	Güte unterschiedlicher Messmethoden zur Erhebung von Ähnlichkeitsurteilen	11
3.1	Aufbau der Untersuchung und Auswahl der Messmethodik.....	11
3.2	Eignungsprüfung der verwendeten Messmethoden .....	13
3.3	Empfundene Angemessenheit verschiedener Messmethoden und Rekonstruktionsgüte der MDS-Konfigurationen .....	14
4	Gestaltung des Untersuchungsrahmens und Ansatzpunkte zur Steigerung der Datenqualität .....	19
4.1	Forschungsfragen und Hypothesen zum Untersuchungsrahmen .....	19
4.2	Untersuchungsergebnisse zur Gestaltung des Untersuchungsrahmens....	21
5	Zusammenfassung und kritische Würdigung der erzielten Untersuchungsergebnisse.....	26

## Abbildungsverzeichnis:

Abb. 1: <i>Struktur der Untersuchung</i> .....	7
Abb. 2: <i>Auswirkungen von 5er Ratingskalen und Magnitudeskalierung auf die MDS-Konfiguration</i> .....	17
Abb. 3: <i>Korrelationsbeziehungen zwischen den verschiedenen Skalen</i> .....	18

## Tabellenverzeichnis:

Tab. 1: <i>Statistische Güte der erzielten zweidimensionalen individuellen MDS-Konfigurationen</i> .....	9
Tab. 2: <i>Probandenbezogene Ergebnisevaluation</i> .....	10
Tab. 3: <i>Reliabilitätsprüfung der Messmethoden mittels QMC und Cronbach´s Alpha der standardisierten Mittelwerte</i> .....	14
Tab. 4: <i>Angemessenheit unterschiedlicher Messmethoden</i> .....	15
Tab. 5: <i>Rekonstruktionsgüte unterschiedlicher Messmethoden</i> .....	15
Tab. 6: <i>Anzahl durchschnittlich genutzter Skalenabstufungen</i> .....	18
Tab. 7: <i>Effekte der Vorabnennung aller Objekte</i> .....	23
Tab. 8: <i>Fragestellungen im Rahmen der probandenbezogenen Befragung</i> .....	23
Tab. 9: <i>Effekte der Objektpositionen</i> .....	25
Tab. 10: <i>Effekte der Beurteilungsunsicherheit</i> .....	26
Tab. 11: <i>Empfohlene Gestaltungsmaßnahmen bei MDS-Datenerhebungen</i> .....	28

## 1 Bedeutung der Multidimensionalen Skalierung zur Analyse von Wahrnehmungen und Gestaltungsbereiche bei der Erstellung von MDS-Untersuchungsdesigns

Unter dem Begriff der Multidimensionalen Skalierung (MDS) wird eine Reihe unterschiedlicher Ansätze und Techniken der multivariaten Datenanalyse subsumiert. Allen gemein ist dabei das Ziel, eine Repräsentation von Objekten, wie etwa Marken oder Produkteigenschaften anhand paarweise gegebener Ähnlichkeitsrelationen in einem metrischen Raum zu erzeugen, so dass die hiernach resultierenden Distanzen die Ähnlichkeitsstruktur der Daten möglichst gut approximiert (vgl. Mathar 1997, S. 9). Nach Young/Hamer (1978, S. 67) sollte die zu erzielende Punktekonfiguration dabei dergestalt sein, dass die Struktur der Daten einfach und verständlich vom untrainierten menschlichen Auge nachvollzogen werden kann.

Den Ausgangspunkt einer jeden MDS stellt eine Distanz- oder Ähnlichkeitsmatrix dar, die Ähnlichkeiten für alle Objektpaare enthält, wobei diese entweder via direkter Abfrage (*Wie ähnlich sind die Objekte X und Y?*) oder indirekt über die Errechnung verschiedener Beurteilungskriterien aufgestellt werden kann, die metrisch oder nichtmetrisch skaliert sind. Die direkte Abfrage stellt die klassische Vorgehensweise im Rahmen der MDS dar und bietet dabei den Vorteil, dass die Objekte gesamtheitlich betrachtet werden, wodurch Variablen- und Haloeffekte vermieden werden können. Aufbauend auf einer Distanzmatrix wird mit Anwendung der MDS das Ziel verfolgt, die enthaltene Ähnlichkeitsstruktur in einem geometrischen und zur besseren Interpretation möglichst niedrigdimensionalen Raum abzubilden, sodass die Modelldistanzen der entsprechenden Objektkonfiguration zwischen allen  $i$  bzw.  $j$  Objekten  $d(i,j)$  die empirischen Ähnlichkeiten  $\delta(i,j)$  möglichst gut widerspiegeln.

Die theoretischen Grundlagen der MDS gehen auf frühe Arbeiten der mathematischen Psychologie zurück, die bereits Ende der 30er und Mitte der 40er Jahre von Young/Householder (1938), Richardson (1938) und Gulliksen (1946) veröffentlicht wurden. Maßgeblich vorangetrieben wurden die Forschungsbemühungen in den 50er und 60er Jahren durch die wegweisenden Arbeiten von Torgerson (1952; 1965), Kruskal (1964a; 1964b), Shepard (1963; 1964; 1966; 1974) und Coombs (1950; 1964). Die Entwicklung und mittlerweile große Bedeutung der MDS in Theorie und Praxis kann insbesondere anhand der großen Zahl an Publikationen in entsprechen-

den Fachzeitschriften, wie *Psychometrika*, *Journal of Marketing Research* oder dem *Journal of Mathematical Psychology* nachvollzogen werden.

Einhergehend mit der steigenden Verfügbarkeit von Softwarepaketen zur statistischen Datenanalyse wie SAS, SPSS, GAUSS oder R und der Implementierung von Algorithmen zur Anwendung der MDS, wie bspw. TORSCA, ALSCAL oder INDSCAL hat die MDS eine wachsende Bedeutung in der sozial- als auch wirtschaftswissenschaftlichen Praxis erfahren. Sowohl im Bereich der sozialwissenschaftlich-psychologischen Forschung als auch im Marketing kann die MDS heute als „das Standardverfahren“ zur Analyse von Wahrnehmungsdimensionen bezeichnet werden (vgl. DeSarbo/Young/Rangaswamy 1997, S. 499). Neben der Anwendung in den Sozialwissenschaften (z. B. Huber 2000; Wunder 2001; Riketta/Vonjahr 1999) finden sich vor allem im Marketing vielfältige Einsatzmöglichkeiten für die MDS, die sich z. B. auf die Analyse subjektiver Produkt-, Marken- oder Verkaufsstättenwahrnehmungen, die Produkt- und Anbieterpositionierung (vgl. Kupke 2004; Gosh/Chakraborty 2004; Lilian/Rangaswamy 1997; Hodgkinson/Padmore/Tomes 1990; Wind/Robinson 1972; Bigne/Vila-Lopez/Küster-Bolude 2000), die Entwicklung von Neuprodukten (vgl. Green/Carmone 1968) oder die Visualisierung von Verbund- und Netzeffekten (vgl. Phillips/Phillips 1998; Decker 2001) beziehen.

Bezogen auf die Anwendungshäufigkeit rangiert die MDS jedoch, verglichen mit anderen Verfahren der multivariaten Datenanalyse, nur auf den hinteren Plätzen (vgl. Hussey/Hooley 1995, S. 24; Waheeduzzaman/Krampf 1992, S. 287ff.). Die zentralen Gründe hierfür liegen zum einen in den Schwierigkeiten, die mit der Interpretation von MDS-Konfigurationen und daraus ableitbaren Handlungsimplicationen (vgl. Carroll/Green 1997, S. 194) verbunden sind und zum anderen in dem relativ hohen Aufwand bei der Erhebung paarweiser Ähnlichkeitsdaten (vgl. MacCallum 1978, S. 69; Spence/Domoney 1974, S. 369).

Bei der Konzeption des Untersuchungsdesigns einer MDS sind vor allem drei Gestaltungsbereiche hervorzuheben, die vom Forscher auch entsprechende Auswahlentscheidungen erfordern:

(1) *Festlegung des Untersuchungsrahmens:*

Hier gilt es konkrete Gestaltungsoptionen zur Erhebung von Ähnlichkeiten zu wählen. Dabei ist insbesondere festzulegen, ob alle Beurteilungsobjekte den Befragten vorher bekannt gegeben werden sollen und in welcher Form die Präsentation der Objekte zu erfolgen hat oder wie probandenbezogene Beurteilungsunsi-

cherheiten bzgl. der Untersuchungsobjekte bei der Erhebung berücksichtigt werden sollten.

(2) *Auswahl einer Skalierungsmethodik zur Messung der Ähnlichkeitsurteile:*

Aus der Vielzahl an Varianten zur Quantifizierung von Ähnlichkeitsdaten, wie der Rangreihung, Dyaden- oder Triadenvergleiche (Taylor/Kinnear 1971) hat sich in der Praxis der Einsatz der metrischen MDS und hier die Verwendung von Ratingskalen etabliert. Unmittelbar mit der Anwendung dieser Skalierungsmethodik stellt sich die Frage nach dem Ausmaß der zu verwendenden Skalenabstufungen.

(3) *Bestimmung eines Abfragedesigns:*

Mit steigender Objektzahl wird der Einsatz eines vollständigen Designs, bei dem alle möglichen Objekt-Kombinationen zur Ähnlichkeitsbeurteilung den Probanden vorgelegt werden, immer schwieriger. So ist ab einer gewissen Objektzahl (10-15) eine Entscheidung zwischen der Verwendung eines vollständigen, eines reduzierten Designs (vgl. Spence/Domoney 1974; MacCullom 1978) oder einer interaktiven Erhebungsform (vgl. DeSarbo/Young/Rangaswamy 1997; Green/Bentler 1979) zu treffen.

Die im Rahmen dieser Gestaltungsbereiche getroffenen Entscheidungen besitzen einen unmittelbaren Einfluss auf die Qualität der mittels MDS gewonnenen Objekt-Konfigurationen, weshalb ihnen bei der Erstellung eines MDS-Untersuchungsdesigns eine herausragende Bedeutung beizumessen ist. Während die Konzeption und Evaluation reduzierter Abfragedesigns (3) im Forschungsbericht Nr. 9 analysiert wird (vgl. Mühlhaus/Hörstrup 2009), konzentriert sich der vorliegende Beitrag auf die Gestaltungsbereiche (1) „Untersuchungsrahmen“ und (2) „Messmethoden“ und geht hier der Frage nach, inwieweit unterschiedliche Gestaltungsentscheidungen bei der Erstellung des Untersuchungsdesigns einen unmittelbaren Einfluss auf die ermittelte MDS-Konfiguration und damit verbunden auch auf die Güte (i. S. v. Validität) der gesamten Untersuchung ausüben.

Um eine Validierung der individuellen sowie der aggregierten MDS-Konfiguration vornehmen zu können, ist nach der Erhebung der Ähnlichkeitsurteile eine zweite Befragung erforderlich (sog. probandenbezogene Ergebnisevaluation), bei der den Teilnehmern der ersten Befragung entsprechend ermittelte MDS-Konfigurationen zur Beurteilung vorgelegt werden. Da die Durchführung von probandenbezogenen Ergebnisevaluationen in der Literatur kontrovers diskutiert wird und nicht unstrittig ist (vgl. Summers/MacKay 1976; Wilkes/Wilcox 1977), wird sie im Folgenden zunächst hinsichtlich ihrer Eignung zur Validierung von MDS-Ergebnissen geprüft (Abschnitt 2.2). Aufbauend auf dieser Prüfung wird in Abschnitt 3 untersucht, ob bei der Erhebung von metrischen Ähnlichkeitsurteilen die Abstufung der verwendeten Ratingska-

la einen Einfluss auf das Ergebnisgüte einer MDS besitzt und ob der Einsatz der sog. Magnitudeskalierung, eine Alternative zur Verwendung von Ratingskalen darstellen könnte. Schließlich wird in Abschnitt 4 geprüft, ob sich durch Vorabnennung der Beurteilungsobjekte die Ergebnisse einer MDS signifikant verändern und inwieweit Positionseffekte sowie die Beurteilungssicherheit der Probanden einen Einfluss auf die Güte einer MDS-Konfiguration besitzen. Bei allen Analysen wird auf die, in praktischen Anwendungen meist gewählte, metrische MDS zurückgegriffen, da diese häufig vergleichbare Ergebnisse wie die nichtmetrische Variante liefert (vgl. Green/Rao 1972, S. 8; Hair et al. 1998, S. 534; Malhotra 1987, S. 164), gleichzeitig aber eine geringere kognitive Beanspruchung der Probanden aufweist (vgl. Backhaus et al. 2006, S. 630) und darüber hinaus deutlich weniger Zeitaufwand erfordert (vgl. Taylor/Kinnear 1971, S. 10). Weiterhin wird bei allen Analysen der in SPSS 16 implementierte und forschungstechnisch stark durchdrungene ALSCAL-Algorithmus (vgl. MacCallum 1978; Takane/Young/De Leeuw 1977) verwendet, der auch für metrische Daten auf individuellem Level mit fehlenden Werten geeignet ist (vgl. Schiffman/Reynolds/Young 1981, S. 251).

## **2 Struktur des verwendeten Untersuchungsdesigns**

### **2.1 Notwendigkeit und Aufbau des zweistufigen Untersuchungsdesigns**

Typischerweise werden zur Güteprüfung von MDS-Untersuchungen die ermittelten MDS-Konfigurationen anhand der klassischen Gütekriterien Stress und  $r$  bzw. dem Bestimmtheitsmaß  $R^2$  beurteilt, die das Ausmaß der Übereinstimmung zwischen den konfigurationsbezogenen Modelldistanzen und den erhobenen Distanzen widerspiegeln. Auf diese Weise ist jedoch nicht auszuschließen bzw. abzuschätzen, ob und in welchem Ausmaß Fehler bereits bei der Datenerhebung vorliegen, die dazu führen, dass die erzielten Konfigurationen von den subjektiven Wahrnehmungen der Probanden abweichen. Einsichtig ist, dass die Datenqualität direkt von der Motivation bzw. dem Kenntnisstand der Auskunftspersonen bzgl. des Untersuchungsgegenstandes beeinflusst wird (vgl. Green/Tull 1982, S. 106f.). Ist die Motivation bei der Beantwortung entsprechender Fragen nur sehr gering, so kann davon ausgegangen werden, dass auf den so erhobenen Daten basierende Ergebnisse nur von fraglichem Wert sind. Ähnlich verhält es sich mit dem Kenntnisstand der Probanden; denn Angaben zu unbekanntem Sachzusammenhängen, die bestenfalls vage Vermutungen darstellen, sollten analog zu den vorangegangenen Ausführungen, dem Prinzip

des „Garbage in – Garbage out“ folgend, auch nur zu zweifelhaften Resultaten führen. Aus diesem Grund erscheint der Rückgriff auf die Konfigurationsbewertung seitens der Probanden zweckmäßig, da nur so die Angemessenheit der Konfiguration umfassend abgeschätzt werden kann. Eine solche Güteprüfung erfordert jedoch eine zweite, der eigentlichen Erhebung nach gelagerte Befragung derselben Personen und somit ein zweistufiges Untersuchungsdesign. Im Januar 2007 wurden deshalb mit Studierenden der Universität Trier folgende zwei aufeinander aufbauende Befragungen durchgeführt (vgl. auch Abb. 1):

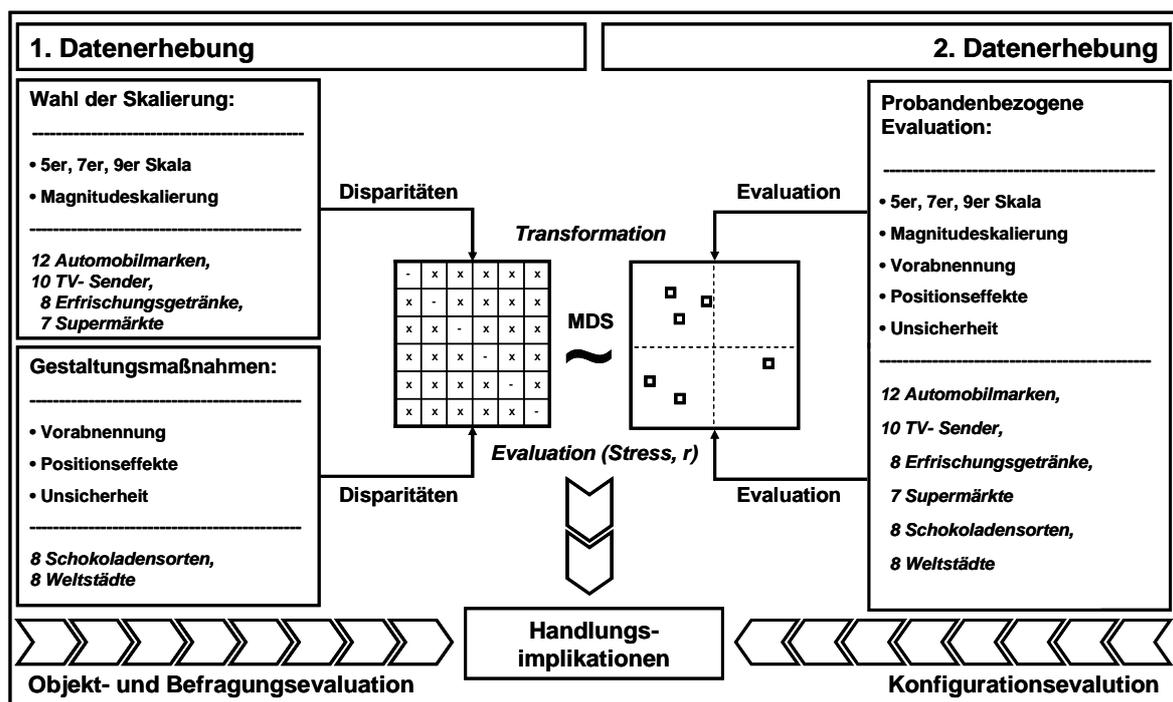


Abb. 1: Struktur der Untersuchung

(1) Erhebung der Ähnlichkeitsurteile (erste Datenerhebung)

Im Rahmen der *ersten Erhebung* wurden bei 132 Studierenden der Wirtschaftswissenschaften im Alter von 22 bis 31 Jahren Ähnlichkeitsurteile zu sechs verschiedenen Objektgruppen erhoben. Dabei wurde bei vier Objektgruppen (Automobilmarken, TV-Sender, Erfrischungsgetränke und Supermärkte als Einkaufsstätte) eine Variation des Erhebungsinstruments in Form einer Verwendung unterschiedlich abgestufter Ratingskalen (5-, 7- und 9-stufige Skalen) sowie der Magnitudeskaliierung zur Gewinnung der Ähnlichkeitsurteile vorgenommen ("Wahl der Skalierung"). Bei zwei Objektgruppen (Schokoladensorten, Weltstädte) wurde der Untersuchungsrahmen durch die Vorabnennung bzw. Nicht-Nennung aller Objekte sowie jeweils veränderte Objektpositionen variiert. Zusätzlich wurde mit der Kennzeichnungsoption unsicherer

Angaben eine weitere Gestaltungsvariante untersucht ("Gestaltungsmaßnahmen"). Für alle Objektgruppen wurden daraufhin MDS-Konfigurationen erstellt, wobei hierfür aufgrund der leichteren Interpretierbarkeit zweidimensionale Darstellungen gewählt wurden (vgl. ausführlich Abschnitt 2.2).

### *(2) Probandenbezogene Ergebnisevaluation (zweite Datenerhebung)*

Die im Rahmen der ersten Befragung gewonnenen MDS-Konfigurationen wurden eine Woche später den Probanden erneut zur Beurteilung vorgelegt ("Probandenbezogene Evaluation"). Für die zweite Befragung konnten noch 109 der ursprünglich 132 Studierenden gewonnen werden. Beide Erhebungen wurden dabei bewusst auf Studierende im Rahmen einer Vorlesung gestützt, da insbesondere für die probandenbezogene Ergebnisevaluation der MDS-Konfigurationen ein Rückgriff auf dieselben Personen (Panellerhebung) erforderlich ist und im Rahmen einer Lehrveranstaltung eine hohe befragungsübergreifende Teilnahmequote sichergestellt werden kann. Weiterhin weisen Studenten i. d. R. eine hohe Motivation, sowie ein hohes Engagement auf, weshalb sie für die nicht triviale visuelle Beurteilung von MDS-Konfigurationen, die zur Eignungsprüfung der einzelnen Gestaltungsmaßnahmen sowie Skalierungsvariationen erforderlich ist, besonders geeignet sind.

## **2.2 Eignung einer probandenbezogenen Ergebnisevaluation**

Die meisten Publikationen zur MDS beschränken ihre Analyse auf die aggregierte Ebene, und es wird keine Betrachtung der individuellen Probanden-Wahrnehmungen vorgenommen. Eine solche aggregierte Vorgehensweise erscheint jedoch aus Anwendersicht wenig zweckmäßig, da gerade bei stark voneinander abweichenden Wahrnehmungen der Probanden und einer damit einhergehenden inhomogenen Datenstruktur die Durchschnittswahrnehmungen nur wenig aussagekräftig sind. Die aggregierten Ergebnisse können damit vielfach in letzter Konsequenz keiner der individuellen Probandeneinschätzungen wirklich gerecht werden. Durch die Analysen auf disaggregierter Ebene sind deshalb meist qualitativ wesentlich höherwertige Erkenntnisse von MDS-Untersuchungen zu erwarten.

Um eine Validierung der individuell erzielten Ergebnisse vornehmen zu können, sollte ergänzend zu den klassischen Gütemaßen Stress und  $r$  bzw.  $R^2$  eine Evaluation der MDS-Ergebnisse in Form einer Beurteilung der visuell vorgelegten Konfigurationen durch die Probanden selbst vorgenommen werden. Mit Hilfe einer solchen probandenbezogenen Ergebnisevaluation kann die Eignung unterschiedlicher Messmetho-

den zur Anwendung der MDS abgeschätzt werden, deren Adäquanz anhand der statistischen Kennzahlen nur eingeschränkt zu prüfen ist. Letztgenannter Aspekt stand zur Erreichung der Zielsetzungen des vorliegenden Beitrags im Vordergrund, wobei zunächst zu prüfen ist, inwieweit diese Form der Konfigurationsevaluation grundsätzlich geeignet ist. Sind Probanden in der Lage ihre eigene individuelle MDS-Konfiguration zu erkennen, so sollten sie diese auch besser beurteilen als eine auf aggregierten Daten oder zufällig erstellte MDS-Konfiguration, da diese ihre eigenen Wahrnehmungen besser widerspiegeln.

Zur Prüfung dieses Zusammenhangs wurden deshalb die individuellen MDS-Konfigurationen für alle sechs Objektgruppen mit Hilfe der ersten Befragung erstellt und jedem Probanden im Rahmen der zweiten Erhebung jeweils drei Konfigurationen zur Beurteilung anhand einer sechsstufigen Ratingskala (Skalierung: 1=spiegelt meine Einschätzung sehr gut wider bis 6=spiegelt meine Einschätzung überhaupt nicht wider) in randomisierter Folge vorgelegt. Zusätzlich zu den individuellen Konfigurationen wurden zwei weitere Konfigurationskarten generiert und zur Beurteilung vorgelegt, wobei die eine der aggregierten MDS-Konfiguration (AGGR-Konfiguration) der jeweiligen Objektgruppe über alle Probandenangaben entsprach und eine zweite zufällig durch Manipulation (MAN) pro Objektgruppe erstellt wurde. Tab. 1 macht zunächst deutlich, dass die zweidimensionale Darstellung der individuellen Ähnlichkeitsurteile für einen überwiegenden Teil der Probanden anhand der Gütekriterien Stress und  $r$  gemäß den Kategorisierungen nach Kruskal/Carmone (1973) sowie Tschudi (1972) für den überwiegenden Teil der Probanden als angemessen einzustufen sind. Die insgesamt nur  $(7+2=)$  9 Konfigurationen, die mit Stresswerten über .10 oder  $r$ -Werten unter .70 als „nicht angemessen“ gelten, werden im Folgenden nicht mehr berücksichtigt.

<b>Stress</b>	Kruskal/Carmone (1973)	Anzahl	<b>r</b>	Tschudi (1972)	Anzahl
.200	"Poor"	<b>2</b>	< .70	"Unacceptable"	<b>7</b>
.100	"Fair"	40	.70-.92	"Fair"	131
.050	"Good"	110	.92-.98	"Good"	51
.025	"Excellent"	44	.98-.99	"Very Good"	5
.000	"Perfect"	-	> .99	"Excellent"	2

Tab. 1: *Statistische Güte der erzielten zweidimensionalen individuellen MDS-Konfigurationen*

Die Ergebnisse der obigen statistischen Güteprüfung sind nun mit den probandenbezogenen Ergebnisbewertungen zu vergleichen. Um die Eignung der individuellen Konfigurationen zu prüfen, sind in Tab. 2 den Beurteilungen der individuellen Konfigurationen (INDIV) die Beurteilungen der Probanden zu der aggregierten (AGGR) sowie einer zufällig manipulierten (MAN) Konfiguration gegenübergestellt. Es wird deutlich, dass die zufälligen Konfigurationen von den Probanden durchschnittlich die schlechtesten Bewertungen (4.72) erhielten. Demgegenüber wurden die aggregierten Konfigurationen mit durchschnittlich 2.82 zwar wesentlich besser beurteilt, liegen aber noch deutlich hinter der Durchschnittsbeurteilung der individuellen Konfigurationen (2.35) zurück.

	INDIV	AGGR	p Bewertung	p Auswahl	MAN
Gesamt	2.35 (70)	2.82 (35)	<.001	.001	4.72 (1)
hohe Sicherheit (S < 4)	2.03 (36)	2.67 (13)	<.001	.001	5.05 (0)
geringe Sicherheit (S > 3)	2.68 (34)	2.97 (22)	n.s.	n.s.	4.58 (1)

Tab. 2: *Probandenbezogene Ergebnisevaluation*

Die Differenzen zwischen den Beurteilungen der individuellen und den aggregierten Ergebnissen können auf Basis eines t-Test auf Mittelwertunterschiede als „hoch signifikant“ eingestuft werden ( $p < .001$ ). Diese deutlichen Unterschiede bestätigen sich auch anhand der Anzahl der Fälle, bei denen die individuellen gegenüber den aggregierten Karten bevorzugt wurden (70 zu 35) und sind mittels Binomialtest als hoch signifikant ( $p = .001$ ) einzustufen. Wird weiterhin der Grad der Beurteilungssicherheit der Probanden (S) berücksichtigt (Skalierung: 1 = keine Unsicherheit; 6 = hohe Unsicherheit), so zeigt sich allerdings, dass die Diskriminierfähigkeit bei hoher Unsicherheit erwartungsgemäß abnimmt (vgl. zur Beurteilungssicherheit Abschnitt 4.1 (c)).

Aus den hier erzielten Ergebnissen kann insgesamt gefolgert werden, dass eine probandenbezogene Ergebnisevaluation, entgegen den Erkenntnissen von Summers/MacKay (1976), nicht nur möglich ist, sondern bei hoher Beurteilungssicherheit auch ein geeignetes Mittel zur Validierung von MDS-Konfigurationen darstellt. Weiterhin kann der Schluss gezogen werden, dass die individuellen Wahrnehmungen bedeutsame Unterschiede zu den durchschnittlichen Einschätzungen über alle Probanden aufweisen. Dies spricht eindeutig für die Zweckmäßigkeit, die Datenanalyse

der MDS nicht ausschließlich auf die aggregierte Ebene zu beschränken, sondern eine Erweiterung der Analysen um individuelle, also probandenbezogene Auswertungen vorzunehmen.

### **3 Güte unterschiedlicher Messmethoden zur Erhebung von Ähnlichkeitsurteilen**

#### **3.1 Aufbau der Untersuchung und Auswahl der Messmethodik**

Bei der metrischen MDS wird zur Erhebung der paarweisen Ähnlichkeitsurteile typischerweise auf *Ratingskalen* zurückgegriffen. Dabei ist bekannt, dass die Güte einer metrischen MDS-Konfiguration, d. h. der Grad mit dem empirische Ähnlichkeitsdaten modelltheoretisch reproduziert werden können, mit zunehmender Abstufung (Kardinalität) der verwendeten Ratingskala ansteigt (vgl. Borg/Groenen 1997, S. 97ff.; Green 1975, S. 76f. oder Green/Wind 1973). Die Verwendung einer sehr feinen Skala mit bspw. 10 oder 20 Stufen erscheint jedoch unter Berücksichtigung von Erkenntnissen der psychologischen Forschung, nach denen Individuen kontextbezogen lediglich zwischen 5 bis 9 Abstufungen unterscheiden können, wenig sinnvoll (vgl. Miller 1956 oder Krantz/Tversky 1975). Eine derart stark untergliederte Skala könnte die Verlässlichkeit der erhobenen Daten reduzieren, da abhängig vom Differenzierungsvermögen der Probanden nicht alle Stufen gleichermaßen berücksichtigt werden und somit lediglich eine „Scheingenaugigkeit“ erzielt wird. Darüber hinaus erhöht die Verwendung einer sehr feinen Skala den von den Auskunftspersonen zu leistenden kognitiven Aufwand, was zusätzlich zu der aufwendigen und ermüdenden Beantwortung paarweiser Ähnlichkeitsurteile (vgl. Dong 1983; Malhotra/Jain/Pinson 1988, S. 95) einen negativen Effekt auf die Datenqualität zur Folge hat. Aufgrund dieses bestehenden „Trade-offs“ zwischen Reproduzierbarkeit der Ausgangsdaten und Güte der erzielbaren Ergebnisse wird in der einschlägigen Literatur die Verwendung von 7 bis 9-stufigen Skalen empfohlen (Backhaus et al. 2006, S. 630; Day/Deutscher/Ryans 1976, S. 94).

Neben dem Einsatz von Ratingskalen kann zur Erhebung der Ähnlichkeitsurteile aber auch auf die sog. *Magnitudeskalierung* (MS) zurückgegriffen werden. Dieses u. a. von Stevens (1975) entwickelte Messverfahren basiert auf Erkenntnissen der psychologischen Forschung, wonach die Stärke einer subjektiven Empfindung direkt auf einem vorgegebenen Antwortkontinuum, etwa durch eine numerische Ziffer oder die Größe (magnitude) eines physikalischen Reizes (z. B. Lautstärke, Helligkeit) ausge-

drückt werden kann. Durch die Psychophysik konnte nachgewiesen werden, dass die objektive Stärke eines Reizes (z. B. die Länge einer Linie) in einem gleich bleibenden und gesetzmäßigen Verhältnis zum subjektiven Empfinden eines Probanden steht (vgl. Lodge 1981, S. 27ff.). Im Vergleich zu Ratingskalen liefert die Magnitudeskalierung ein höheres, „echt metrisches“ Skalenniveau (vgl. Neibecker 1983a, S. 211; Hüttner 1997, S. 113), kann feiner zwischen ähnlichen Reizen diskriminieren (vgl. Grunert 1983, S. 111) und bietet in Kombination mit alternativen Skalen auch bei längeren und monotonen Befragungen eine größere Abwechslung (Neibecker 1983b, S. 185), was einen positiven Einfluss auf die Probandenmotivation ausübt. Vor diesem Hintergrund erscheint die Magnitudeskalierung auch für die Messung von Ähnlichkeiten bei MDS-Untersuchungen besonders geeignet.

Da bisher weder der „optimale“ Abstufungsgrad von Ratingskalen, noch der Einsatz der Magnitudeskalierung im Rahmen der MDS empirisch untersucht wurden, wird im Folgenden die Auswirkung des Einsatzes unterschiedlicher Ratingskalen sowie der Magnitudeskalierung auf die Güte der erzielbaren MDS-Ergebnisse untersucht. Zu diesem Zweck wurden vier Objektgruppen gebildet (12 Automobilmarken, 10 deutsche TV-Sender, 8 Erfrischungsgetränke und 7 Supermärkte), bei denen jeweils davon auszugehen war, dass diese allen Studenten auch bekannt sind.<sup>1</sup> Die Ähnlichkeiten innerhalb dieser Objektgruppen wurden bei verschiedenen Probanden mittels vier verschiedener Messmethoden erhoben und entsprechende MDS-Konfigurationen zur Evaluation vorgelegt. Auf diese Weise können folgende Forschungsfragen beantwortet werden:

- F1: Welche Messmethode ist aus Probandensicht (i. S. v. Angemessenheit bei der Abfrage der Ähnlichkeitsurteile) am besten geeignet?
- F2: Mit welcher Messmethode können MDS-Konfigurationen mit der höchsten Güte (i. S. v. Angemessenheit aus Probandensicht und Rekonstruktionsgüte) erzielt werden?

Als Messmethoden wurden die in der Literatur empfohlenen 7er und 9er-Ratingskalen, eine fünfstufige Ratingskala sowie die Magnitudeskalierung ausgewählt. Entsprechend den Empfehlungen der Literatur sollten bei der Magnitudeskalie-

---

<sup>1</sup> Folgende Einzelobjekte wurden in der Untersuchung berücksichtigt: (1) Automobilmarken: Alfa Romeo, Audi, BMW, Ferrari, Fiat, Honda, Mazda, Mercedes-Benz, Opel, Porsche, Renault und VW. (2) TV-Sender: ARD, Arte, N 24, Phoenix, Pro 7, RTL, Sat 1, Super RTL, Vox und ZDF. (3) Erfrischungsgetränke: Coca Cola, Fanta, Lift Apfelschorle, Limonade, Lipton Eistee, Mineralwasser, Pepsi und Sprite (4) Supermärkte: Accord, Aldi, Edeka, Kaufland, Lidl, Plus und Rewe.

rung im Rahmen von Paper-Pencil-Befragungen die Ähnlichkeitsangaben anhand vom Probanden zu zeichnenden Linien erhoben werden, wobei dies möglichst frei und unrestringiert erfolgen sollte (vgl. Behrens 1983, S. 133). Aus diesem Grund wurde bei den hier durchgeführten Untersuchungen weder ein Bezugsrahmen, d. h. eine maximale Begrenzung der Linienlänge vorgegeben, noch wurde festgelegt, ob hohe Ähnlichkeiten durch eine lange oder eine kurze Linie gekennzeichnet werden sollen. Da die Magnitudeskalierung und die damit verbundene Angabe von Urteilen den meisten Studenten unbekannt war, erfolgte vorab eine etwa 5-minütige Einführung und Trainingsphase (vgl. Lodge 1981, S. 9). Um systematische Effekte auszuschließen erfolgte die Aufteilung der Beurteilungsaufgaben (Kombination von Objektgruppe und Skala) auf die Probanden in randomisierter Form, so dass je Person jede Objektklasse und Skala genau einmal berücksichtigt wurde. Zusätzlich erfolgte eine fragebogeninterne Randomisierung der Beurteilungsaufgaben, um auf diese Weise Reihenfolgeeffekte auszuschließen.

Im Folgenden wird zunächst eine Eignungsprüfung für die vier verwendeten Messmethoden durchgeführt und anschließend die von den Probanden empfundene Angemessenheit der verschiedenen Varianten, sowie die Rekonstruktionsgüte der MDS-Konfigurationen untersucht.

### **3.2 Eignungsprüfung der verwendeten Messmethoden**

Zur Eignungsprüfung der verwendeten Messmethoden wurde zunächst deren Reliabilität untersucht (vgl. Peter 1979; Cronbach/Meehl 1955). Zur *Reliabilitätsprüfung* wurden zum einen die quadrierten multiplen Korrelationen (QMC) der standardisierten mittleren Ähnlichkeitsurteile innerhalb der vier Objektkonstellationen und zum anderen Cronbach's Alpha als Maß für die Gesamteignung einer Skala berechnet (vgl. Cronbach 1951; Nunnally 1967). Der Rückgriff auf die mittleren Angaben war erforderlich, da aufgrund des gewählten Experimentaldesigns jeder Proband jede Objektklasse nur anhand einer Skala beurteilt hat. Tabelle 3 zeigt die erzielten Ergebnisse. Die insgesamt sehr hohen Werte machen deutlich, dass die Wahl der Messmethodik keinen Einfluss auf die Struktur der Antworten ausübt, also unabhängig von der verwendeten Skala die durchschnittlichen Angaben vergleichbar sind. Die Reliabilität der Messung von Ähnlichkeiten anhand der vier Methoden ist somit im vorliegenden Fall gegeben.

	Objektgruppen				
	Auto	TV	Getränk	Markt	Gesamt
	<i>Quadierte Korrelationen (QMC)</i>				
5er Skala	.912	.932	.868	.698	.874
7er Skala	.895	.919	.878	.640	.804
9er Skala	.833	.760	.883	.648	.782
Magnitudeskalie- rung	.846	.913	.888	.672	.747
	<i>Cronbach's Aplpha</i>				
	.966	.973	.975	.918	.947

Tab. 3: Reliabilitätsprüfung der Messmethoden mittels QMC und Cronbach's Alpha der standardisierten Mittelwerte

### 3.3 Empfundene Angemessenheit verschiedener Messmethoden und Rekonstruktionsgüte der MDS-Konfigurationen

Zur Prüfung der Angemessenheit der vier verschiedenen Messmethoden zur Abfrage von Ähnlichkeitsurteilen wurden die Befragten gebeten, diese anhand einer sechsstufigen Ratingskala mit den Ausprägungen der Schulnotenskala (1=sehr gut bis 6=ungenügend) zu beurteilen (ANG). Die in Tabelle 4 dargestellten Durchschnittsbeurteilungen über alle Befragten machen deutlich, dass die 5er Skala im Mittel (2.35) am besten und signifikant besser als die anderen Skalen beurteilt wurde, was anhand der geringeren kognitiven Erfordernisse erklärt werden kann. Die mittlere Beurteilung der Magnitudeskalierung (3.26), die nur geringfügig schlechter als die 7er Skala (3.18), aber deutlich besser als die 9er (3.79) Skala ausfällt, ist insbesondere deshalb bemerkenswert, da die Magnitudeskalierung im Zusammenhang mit MDS-Untersuchungen eine neuartige Beurteilungsmethode darstellt und für die Probanden als kognitiv anspruchsvoller einzustufen ist. Die schlechte Beurteilung der 9er Skala, die im angloamerikanischen Raum häufig Anwendung findet, kann dadurch erklärt werden, dass diese Skala in Deutschland nur selten angewandt wird und die Probanden daher mit ihrem Umgang wenig geübt sind.

Skala	ANG	hohe Sicherheit (S < 4)	geringe Sicherheit (S > 3)	p
5 (n=31)	<b>2.35</b>	2.18	3.00	.023
7 (n=24)	<b>3.18</b>	2.81	3.72	.008
9 (n=26)	<b>3.79</b>	3.59	4.43	.003
MS (n=23)	<b>3.26</b>	3.16	3.50	.078

Tab. 4: Angemessenheit unterschiedlicher Messmethoden

Interessant ist weiterhin, dass die Unterschiede in der beurteilten Angemessenheit der Messmethoden abhängig von der Beurteilungssicherheit (S) bei der Magnitudeskalierung deutlich geringer als bei den Ratingskalen ausgeprägt ist und zu einem Konfidenzniveau von  $\alpha=.05$  nicht mehr signifikant sind. Das bedeutet, dass die Magnitudeskalierung sowohl für Personen, die sich bzgl. der Beurteilung der Paarvergleiche sicher als auch unsicher sind gleichermaßen geeignet ist, was darauf zurückgeführt werden kann, dass die „Feinheit der Skala“ und der vorgenommene Grad an Urteilsabstufungen vom Probanden hier selbst festgelegt wird.

Skala	statistische Gütekriterien		probandenbezogene Evaluation		
	Stress	r	INDIV	AGGR	p
<b>5 (n=31)</b>	.033	.912	1.90	2.32	n.s.
<b>7 (n=24)</b>	<b>.029</b>	<b>.916</b>	<b>1.88</b>	<b>2.44</b>	.014
<b>9 (n=26)</b>	.034	.888	<b>2.64</b>	2.21	n.s.
<b>MS (n=23)</b>	<b>.054</b>	<b>.893</b>	<b>1.76</b>	2.57	.007

Tab. 5: Rekonstruktionsgüte unterschiedlicher Messmethoden

Zur Beurteilung der *Rekonstruktionsgüte der zweidimensionalen (aggregierten sowie individuellen) MDS-Konfigurationen*, die auf unterschiedlichen Instrumenten zur Ähnlichkeitsabfrage basieren, wird auf folgende Gütekriterien zurückgegriffen. Zum einen werden die individuellen Konfigurationen anhand der klassischen Gütemaße Stress (vgl. Green 1975, S. 74) und Pearson's r (= Korrelationskoeffizient  $r(\delta, d)$  der Ähnlichkeitsurteile  $\delta$  und euklidischen Modelldistanzen  $d$  der ermittelten MDS-Konfigurationen) bewertet (vgl. Malhotra/Jain/Pinson 1988, S. 97). Dabei vermittelt der Korrelationskoeffizient bzw. dessen Quadrat (sog. Bestimmtheitsmaß  $R^2$ ), wie

Schiffman/Reynolds/Young (1981, S. 175) zeigen, oftmals einen besseren Eindruck der Konfigurationsgüte. Darüber hinaus werden zusätzlich die probandenbezogenen Konfigurationsbewertungen aus der zweiten Probandenbefragung berücksichtigt (Skalierung: 1=spiegelt meine Einschätzung sehr gut wider bis 6=spiegelt meine Einschätzung überhaupt nicht wider). Tabelle 5 fasst die Ergebnisse zu den verschiedenen Gütekriterien zusammen.

Bei Betrachtung der durchschnittlichen Gütekriterien Stress und  $r$  über alle Probanden je Messmethode zeigt sich, dass hiervon keine bedeutsamen Effekte ausgehen. Insgesamt liefert die 7er Skala im Mittel sowohl hinsichtlich der Stress-Maße (.029) als auch der  $r$ -Werte (.916) die besten Reproduktionsergebnisse. Hinsichtlich der probandenbezogenen Bewertung zeigen sich keine signifikanten Effekte, die auf die Messmethoden zurückzuführen sind. Lediglich die individuellen Konfigurationen (INDIV) der Proximitätsdaten, die mittels 9er Skala erhoben wurden, werden im Mittel (2.64) deutlich schlechter beurteilt. Allerdings fällt auf, dass die statistischen Gütekriterien *nicht* mit der Bewertung der Probanden einhergehen. So sind die mittels Magnitudeskalierung erhobenen Ähnlichkeiten gemäß der statistischen Gütekriterien schlechter rekonstruiert, werden jedoch von den Probanden am besten beurteilt. Der Grund hierfür kann darin gesehen werden, dass die mittels Magnitudeskalierung erzielten Konfigurationen aufgrund der besseren Abstufungsmöglichkeiten insbesondere zwischen sehr unähnlichen Objektpaaren stärker strukturiert und damit leichter zu interpretieren sind. Verglichen mit den wenig differenzierenden Konfigurationen, die etwa anhand einer 5er Skala erzielt werden, wird dies, wie Abb. 2 zu entnehmen ist, offenkundig. Diese höhere Trennschärfe bestätigt sich anhand des Vergleichs der Beurteilung der individuellen (INDIV) und der aggregierten (AGGR) Konfigurationen. Hier zeigt sich, dass lediglich bei der 7er Skala und der Magnitudeskalierung die Beurteilungen signifikant differieren.

Bezugnehmend auf die Forschungsfrage F2 kann festgehalten werden, dass die 9er Skala deutlich schlechtere Ergebnisse als die anderen Skalen liefert. Diese Erkenntnis führt zu der allgemeinen Empfehlung, den in der Fachliteratur häufig zu entnehmenden Ratschlag, bei der metrischen MDS eine Skala mit 7 bis 9 Abstufungen zu wählen, zumindest für Anwendungen in Deutschland zu ignorieren und die Skalenkardinalität auf den Bereich von 5 bis 7 Abstufungen einzuschränken.

Hinsichtlich der *Magnitudeskalierung* kann zusammenfassend festgestellt werden, dass diese eine adäquate Alternative bzw. Ergänzung zur Verwendung der klassi-

schen Ratingverfahren darstellt. Sie liefert vergleichbare Rekonstruktionsergebnisse und wird darüber hinaus aus Probandensicht ähnlich angemessen beurteilt. Ein wesentlicher Vorteil der Magnitudeskalierung ist vor allem darin zu sehen, dass die hiernach erzielten Konfigurationen stärker strukturiert sind, was eine leichtere Interpretation ermöglicht. Zur Verdeutlichung dieser Erkenntnis ist in Abb. 2 die aggregierte Konfiguration der auf Basis der Magnitudeskalierung gewonnenen Ähnlichkeitsurteile ausgewiesen und der Konfiguration die anhand derselben erhobenen Ähnlichkeitsdaten – bei Reskalierung in eine 5er-Skala resultierenden Konfiguration – gegenübergestellt.

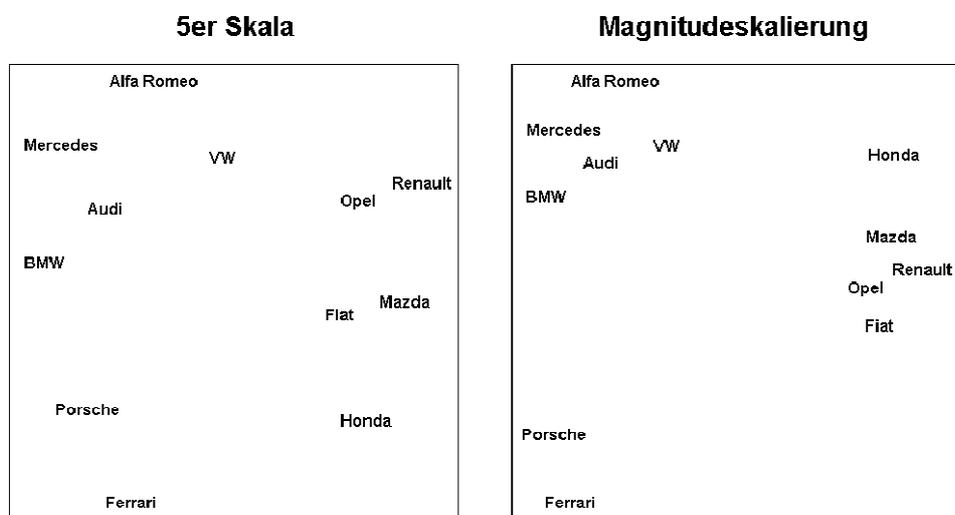


Abb. 2: Auswirkungen von 5er Ratingskalen und Magnitudeskalierung auf die MDS-Konfiguration

Um diesen Vergleich zu ermöglichen, wurden die magnitudeskalierten Daten in die entsprechenden Proximitäten der 5er- Skala transformiert. Da im vorliegenden Fall zwischen den anhand einer Ratingskala erhobenen Urteilen und denen bei Anwendung der Magnitudeskalierung ein linearer Zusammenhang besteht (vgl. hierzu Abb. 3 sowie die Ergebnisse von Grunert 1983, S. 111, der ebenfalls keine nennenswerten Abweichungen vom linearen Zusammenhang feststellt) wurden die magnitudeskalierten Angaben in linearer Weise gemäß der nachfolgenden Formel reskaliert:

$$(1) \delta_{i,j}^Z = \text{RUNDEN} \left( Z^{\min} + \frac{(\delta_{i,j}^{MS} - \min(\delta^{MS})) \cdot (Z^{\max} - Z^{\min})}{\max(\delta^{MS}) - \min(\delta^{MS})} \right) \text{ mit Skalenbereich: } (Z^{\min}; Z^{\max}) = (1; 5)$$

Wird der je Objektklasse von einem Probanden genutzte maximale Skalenwert durch den minimalen Skalenwert, der als Skaleneinheit zu interpretieren ist, dividiert, so erhält man die verwandte Skalenabstufung (vgl. Grunert 1983, S. 110).

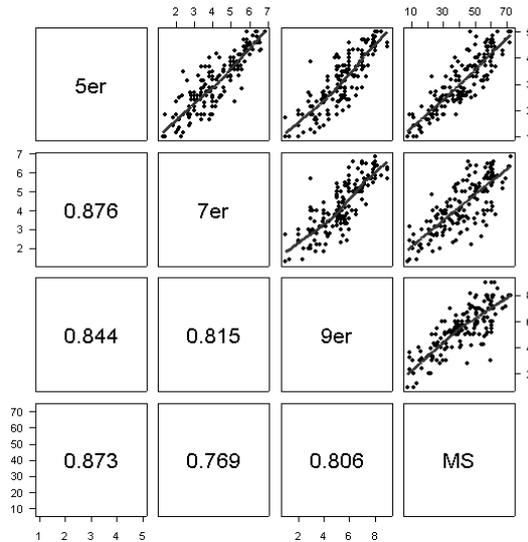


Abb. 3: Korrelationsbeziehungen zwischen den verschiedenen Skalen

Die in Tab. 6 ausgewiesenen mittleren Abstufungen je Skala zeigen, dass die mittels Magnitudeskalierung erhobenen Ähnlichkeitsdaten im Ergebnis deutlich feiner differenziert sind. Verglichen mit der 5er Skala liefert die Magnitudeskalierung eine um den Faktor 4 (16.34 zu 4.27) höhere Differenzierung der Ähnlichkeitsurteile, wodurch die wesentlich bessere Strukturierung der auf der Magnitudeskalierung erzielbaren MDS-Konfigurationen erklärt werden kann.

Skala	Mittelwert	Median	Max	Min
5 (n=31)	<b>4.27</b>	5	5	1.67
7 (n=24)	5.26	6	7	1.50
9 (n=26)	6.20	7	9	1.75
MS (n=23)	<b>16.34</b>	13.2	72	2.00

Tab. 6: Anzahl durchschnittlich genutzter Skalenabstufungen

Insbesondere bei Onlinebefragungen, die nach wie vor eine steigende Verbreitung erfahren (vgl. ADM 2005, S. 7ff.; Zerr 2003, S. 24; Weis/Steinmetz 2005, S. 108), sollten die oftmals im Rahmen der Magnitudeskalierung genannten Schwierigkeiten in der praktischen Umsetzung bei Erhebung und Erfassung der Daten (vgl. Neibecker 1984) umgangen werden können, weshalb die Magnitudeskalierung hierfür besonders geeignet erscheint.

## **4 Gestaltung des Untersuchungsrahmens und Ansatzpunkte zur Steigerung der Datenqualität**

### **4.1 Forschungsfragen und Hypothesen zum Untersuchungsrahmen**

Unter dem Aspekt der Steigerung der Datenqualität wurden folgende drei Problemfelder ausgewählt, um auf Basis der erzielten Erkenntnisse konkrete Hinweise zur Ausgestaltung des Untersuchungsrahmens zu erhalten:

- (a) Vorabnennung oder Nicht-Nennung der Beurteilungsobjekte
- (b) Prüfung der Transitivität der Ähnlichkeitsurteile
- (c) Einfluss der Beurteilungsunsicherheit der Befragten auf die MDS-Ergebnisse

Nachfolgend werden zunächst die Problemfelder im Detail diskutiert und die zu untersuchenden Hypothesen abgeleitet. Anschließend werden die entsprechend erzielten Untersuchungsergebnisse in Abschnitt 4.2 vorgestellt und konkrete Handlungsempfehlungen gegeben.

#### ***(a) Vorabnennung der Beurteilungsobjekte***

Die individuelle Wahrnehmung von Objekten erfolgt referenzbezogen, d. h. dass Urteile nicht absolut, sondern bezogen auf einen bestimmten Kontext bzw. einen Vergleichsstimulus erfolgen (vgl. Fechner 1877; Allais 1953; Kelly 1955; Rosch 1975). In Bezug auf die MDS bedeutet dies, dass die Probanden die verschiedenen Objekte einer Objektkategorie in einem Beurteilungsraum anhand von Referenzen positionieren und sie auf diese Art und Weise von einander abgrenzen. Sollen also Ähnlichkeiten zwischen verschiedenen Objekten untersucht werden, so ist ebenfalls davon auszugehen, dass eine diesbezügliche Wahrnehmung kontextbezogen von der Gesamtheit der betrachteten Objekte abhängt. Folgendes Beispiel möge dies veranschaulichen: Sollen die Erfrischungsgetränke Coca Cola, Fanta und Mineralwasser hinsichtlich deren Ähnlichkeit beurteilt werden, so erscheinen Coca Cola und Fanta im Vergleich zu Mineralwasser recht ähnlich. Wird hingegen das Mineralwasser durch Pepsi Cola ersetzt, so stellen sich nun Coca Cola und Fanta als unähnlich dar. Der Austausch von Mineralwasser durch Pepsi führt somit zu einer Neupositionierung der Objekte innerhalb des durch die drei Erfrischungsgetränke aufgespannten Beurteilungsraumes eines Probanden.

Weiterhin ist bei einer sukzessiven Ähnlichkeitsabfrage zwischen Objekten davon auszugehen, dass mit jeder durchgeführten Paarbeurteilung eine Adaption des Be-

wertungsrahmens der intern zugrundeliegenden Skala erfolgt (vgl. Borg/Groenen 1997, S. 94). Dies führt letztlich dazu, dass die Beurteilung der Objekte nicht unabhängig von der Reihenfolge der Paarvergleiche erfolgt, was einerseits die Vergleichbarkeit der Ähnlichkeitsdaten und darüber hinaus deren Gültigkeit beeinträchtigt. Da die Verwendung einer Ratingskala die Probanden bei Angabe der Ähnlichkeitsurteile auf einen vorgegebenen Bereich innerhalb der Skalenendpole einschränkt, sollte irgendwann der Fall eintreten, dass ein zu beurteilendes Paar deutlich ähnlicher oder unähnlicher erscheint als ein bereits mit den Extremkategorien der Skala beurteiltes Paar. In einem solchen Fall befindet sich die Auskunftsperson in einer Art Dilemma, da entweder bereits getätigte Angaben angepasst werden müssen oder eine empfindungstreue Abstufung zwischen den Urteilen nicht erfolgen kann. Dies kann zum einen die beurteilte Angemessenheit der Skalierung verringern und darüber hinaus bei ausbleibender Anpassung bereits abgegebener Urteile die Güte der erzielbaren MDS-Konfiguration reduzieren. Derartige „order effects“ (vgl. Jain/Pinson 1976, S. 435) sollten jedoch dadurch abgeschwächt werden können, indem den Probanden vor Bearbeitung der Ähnlichkeitsurteile alle Vergleichs- oder Beurteilungsobjekte zur Bildung eines Beurteilungsraumes (ex ante) präsentiert werden. Diese Einschätzung mündet in den drei folgenden Hypothesen:

H1a: *Die paarweisen Ähnlichkeitsurteile unterscheiden sich in Abhängigkeit davon, ob eine Vorabnennung aller Objekte erfolgt oder nicht.*

H1b: *Eine Vorabnennung aller Beurteilungsobjekte erhöht die wahrgenommene Angemessenheit der verwandten Skalierung und steigert die Beurteilungssicherheit.*

H1c: *Eine Vorabnennung aller Beurteilungsobjekte verbessert die Güte der erzielbaren MDS-Konfiguration.*

### **(b) Prüfung der Transitivität der Ähnlichkeitsurteile**

Zusätzlich zu den beschriebenen Effekten, die Nennungssequenz der Objekte betreffend, ist nicht auszuschließen, dass die Annahme der Symmetrie bzw. Transitivität der paarweisen Ähnlichkeitsurteile, die bei Anwendung der MDS üblicherweise unterstellt wird [also  $\delta(A,B)=\delta(B,A)$ ] stets gegeben ist (vgl. Coombs 1950, S. 146; Cox/Cox 1994, S. 89; Huber/Holbrook 1982). Untersuchungen zur Beurteilung von Ländern (vgl. Tversky 1977) oder der Ähnlichkeit der Codes des Morsealphabets (vgl. Rothkopf 1957) zeigen auf, dass abhängig von der Ordnung der zu beurteilenden Objekte starke Abweichungen auftreten können. Dieser Effekt, den

Borg/Groenen (1997, S. 94) als „position effect“ bezeichnen, wird im Folgenden ebenfalls untersucht und führt zu folgender Hypothese:

H2: *Der Einfluss von „position effects“ ist stark ausgeprägt und führt zu unterschiedlichen Ähnlichkeitsbeurteilungen.*

### **(c) Einfluss der Beurteilungsunsicherheit der Befragten auf die MDS-Ergebnisse**

Der Grad an Beurteilungssicherheit (S) mit dem Probanden Urteile über bestimmte Objekte treffen, hängt stark von deren Kenntnisstand hinsichtlich der betreffenden Beurteilungsobjekte ab. Wird bei der Erhebung von Ähnlichkeitsdaten klassisch vorgegangen, so bieten sich aus Probandensicht bei geringer Beurteilungssicherheit lediglich zwei Optionen: die Angabe einer vagen Vermutung oder die Auskunftsverweigerung. Da jedoch ohne eine explizite Abfrage des Grades an Sicherheit je Paarvergleich keinerlei Anhaltspunkte darüber bestehen, welcher probandenbezogene Kenntnisstand bei der Beantwortung vorlag bzw. welche Vergleiche nicht angemessen beurteilt werden konnten, ist nicht abzuschätzen, wie verlässlich die entsprechend erzielbaren Ergebnisse sind. Aufgrund des ohnehin sehr hohen Aufwandes zur Erhebung von Daten zur Durchführung einer MDS erscheint eine explizite Abfrage des Grades der Beurteilungssicherheit für jeden einzelnen Paarvergleich zusätzlich wenig zweckmäßig. Vielmehr sollte daher den Probanden die Möglichkeit geboten werden, sehr unsichere Einschätzungen zu kennzeichnen. Eine solche binäre Kennzeichnung (Vorliegen einer hohen Unsicherheit ja oder nein) könnte anschließend in der Auswertung berücksichtigt werden. Die einfachste Möglichkeit stellt dabei der Ausschluss derartiger Urteile dar und soll im Folgenden hinsichtlich der Auswirkung auf die Konfigurationsgüte untersucht werden. Wir gehen von folgenden Hypothesen aus:

H3a: *Die Möglichkeit zur Kennzeichnung unsicherer Angaben reduziert die Anzahl fehlender Werte.*

H3b: *Der Ausschluss von unsicheren Angaben verbessert die Güte der erzielbaren MDS-Konfiguration.*

## **4.2 Untersuchungsergebnisse zur Gestaltung des Untersuchungsrahmens**

Zur Überprüfung der Hypothese **H1** bis **H3** wurde ein vollständiges Experimentaldesign mit den vier Faktoren Vorabnennung (ja/nein), Unsicherheit (mit/ohne Kennzeichnung), Positionseffekt (A,B bzw. B,A) und Objekt (8 Städte bzw. 8 Schokoladensorten) gewählt und in randomisierter Form auf die Probanden angewandt. Die

Ähnlichkeitsurteile wurden anhand einer 7er Skala erhoben, die entsprechend den Ergebnissen in Abschnitt 3.3 zur Anwendung der MDS besonders geeignet ist. Da die einzelnen Faktoren inhaltlich einen nur geringen Bezug zueinander aufweisen und hierbei keine signifikanten Interaktionseffekte identifiziert werden konnten, erfolgt die Untersuchung der Hypothesen in den einzelnen drei Bereichen getrennt voneinander.

### **(a) Vorabnennung der Beurteilungsobjekte**

Entsprechend **Hypothese 1a** ist zu prüfen, ob die Vorabnennung der zu beurteilenden Objekte die Antwortstruktur beeinflusst. Hierzu wurden einerseits die Korrelationen zwischen den Ähnlichkeitsurteilen der Gruppen mit und ohne Vorabnennung berechnet sowie die mittleren absoluten Differenzen zwischen diesen Gruppen gemäß Formel (2) bestimmt:

$$(2) \quad ABW^{(A,B)} = \frac{1}{pv} \sum_{i < j} | \bar{\delta}_{ij}^A - \bar{\delta}_{ij}^B |$$

mit:  $\bar{\delta}_{ij}^A$  : durchschnittlich beurteilte Ähnlichkeit der Objekte i und j in Gruppe A

$$pv = \frac{1}{2} \cdot i \cdot (i - 1) : \text{Anzahl der Paarvergleiche}$$

Die in Tab. 7 zusammengefassten Ergebnisse machen deutlich, dass nicht nur die Ähnlichkeitsurteile zwischen den Gruppen mit und ohne Vorabnennung sowohl bei der Objektgruppe „Schokoladensorten“ ( $r=.849$ ) als auch bei der Objektgruppe „Städte“ ( $r=.962$ ) sehr hoch korrelieren, sondern auch die mittleren absoluten Abweichungen (ABW) sehr gering sind. Das bedeutet, dass die Vorabnennung aller Beurteilungsobjekte keinen nennenswerten Effekt auf das Urteilsverhalten der Probanden ausübt. Dies bestätigen auch die je Paarvergleich durchgeführten t-Tests, bei denen keiner bzw. nur einer der 28 je Objektgruppe durchgeführten Paarvergleiche mit und ohne Vorabnennung signifikant voneinander abweichend beurteilt wurde. Dies erscheint insbesondere für die Anwendung von reduzierten Abfragedesigns bedeutsam, bei denen nicht alle erforderlichen Paarvergleiche zu Beurteilung herangezogen, sondern lediglich eine Partition hiervon vorgelegt wird (vgl. Spence/Domoney 1974). Die hier gemessenen, nur geringen Unterschiede zwischen einer Vorgehensweise mit und ohne Vorabnennung sprechen für die grundsätzliche Eignung von reduzierten Abfragedesigns (siehe hierzu Mühlhaus/Hörstrup 2009).

Objekt- gruppe	Vorab- nennung	Prüfung Hypothese 1a		Prüfung Hypothese 1b			Prüfung Hypothese 1c		
		r	ABW	ANG	AUSG	S	Stress	r	INDIV
Schokolade (n=53)	mit (n = 23)	<b>.849</b>	.290	3.16	4.20	4.04	.040	.830	2.95
	ohne (n = 30)			3.48	4.32	4.24	.043	.840	2.47
Städte (n=49)	mit (n = 24)	<b>.962</b>	.365	3.25	3.77	4.08	.044	.827	2.72
	ohne (n = 25)			3.50	4.29	4.69	.039	.849	2.63
Gesamt (n=102)	mit (n = 47)	.892	.327	3.20	3.98	4.06	.042	.829	2.84
	ohne (n = 55)			3.49	4.31	4.47	.041	.845	2.55

Tab. 7: Effekte der Vorabnennung aller Objekte

Zur Prüfung der **Hypothese 1b** wurden im Rahmen der probandenbezogenen Befragung die wahrgenommene Angemessenheit der verwandten Skalierung (ANG), die vom Befragten empfohlene Ausgestaltung der Skala (AUSG) sowie die Beurteilungssicherheit der Probanden (S) mit Hilfe der in Tab. 8 zusammengefassten Fragestellungen bei den Probanden erfragt:

Abkürzung	Erläuterung	Abfrage und Skalierung
ANG	Im Durchschnitt beurteilte Angemessenheit der Skala	Für wie angemessen halten Sie die verwendete Skalierung zur Abbildung Ihrer Einschätzungen? (1=sehr angemessen bis 6=sehr unangemessen)
AUSG	Im Durchschnitt empfohlene Ausgestaltung einer "optimalen" Skala	Wie sollte die in Ihren Augen "optimale" Skala gestaltet sein? (1=sehr viel feiner (mehr Stufen) bis 6=sehr viel gröber (weniger Stufen))
S	Mittlere Sicherheit bei der Angabe der Ähnlichkeitsurteile	Wie sicher waren Sie insgesamt bei der Beurteilung der Paarvergleiche? (1=keine Unsicherheit bis 6=hohe Unsicherheit)
INDIV AGGR MAN	Durchschnittliche Beurteilung der vorgelegten individuellen (INDIV), aggregierten (AGGR) oder zufällig manipulierten (MAN) Karten	Wie gut kann diese Konfiguration Ihre Wahrnehmung über die abgebildeten Objekte wiedergeben? (1=spiegelt meine Einschätzung sehr gut wider bis 6=spiegelt meine Einschätzung überhaupt nicht wider)

Tab. 8: Fragestellungen im Rahmen der probandenbezogenen Befragung

Die in Tab. 7 aufgeführten Ergebnisse machen deutlich, dass sich hinsichtlich der Antwortsicherheit und der Beurteilung der verwendeten Skala ebenfalls keine signifikanten Effekte bezüglich der Vorabnennung der Objekte ergeben. Allerdings zeigt sich ein konsistentes Muster, das – wenngleich statistisch nicht signifikant – dennoch interessant erscheint: Unabhängig vom Untersuchungsgegenstand ist die durchschnittliche Beurteilungssicherheit (S) bei Vorabnennung der Objekte höher (4.06) als ohne Vorabnennung (4.47). Ebenso konsistent ist die Beurteilung der Skala in Bezug auf ihre Angemessenheit, die mit Vorabnennung besser ausfällt (3.20) als ohne (3.49). Weiterhin bevorzugen die Befragten im Durchschnitt eine feiner abgestufte Skala (3.98) gegenüber einer weniger abgestuften (4.39). Diese als Tendenzen zu bezeichnenden Effekte sind somit konform zur Hypothese 1b.

Wird schließlich der in **Hypothese 1c** postulierte Einfluss der Vorabnennung der Objekte auf die Güte der MDS-Konfigurationen betrachtet, so zeigt sich ein überraschender Effekt. Sowohl die statistischen Gütemaße (Stress- und r-Werte), als auch die Bewertung durch die Probanden (INDIV) deuten bei Vorabnennung auf schlechtere Ergebnisse hin. Obschon diese Effekte abermals statistisch nicht signifikant sind, widersprechen sie jedoch der Intuition, nach der eine tendenziell höhere Beurteilungssicherheit (S) auch zu besseren Ergebnissen führen sollte. Insgesamt ist hinsichtlich der Vorabnennung zu konstatieren, dass keine der Hypothesen statistisch bestätigt werden konnte und der hiermit erzielbare theoretische und intuitive Nutzen empirisch nicht gegeben ist.

### ***(b) Prüfung der Transitivität der Ähnlichkeitsurteile***

Die Ergebnisse der Prüfung von „position effects“ (**Hypothese 2**) zeigt Tab. 9. Sowohl bei der marketingspezifischen Beurteilungsaufgabe der Ähnlichkeit von Schokoladenmarken ( $r=.909$ ) als auch bei der Beurteilung der Städte ( $r=.892$ ) waren die mittleren Angaben bei beiden Positionen hoch korreliert. Darüber hinaus zeigten sich auch auf der Ebene der insgesamt  $2 \times 28 = 56$  Paarvergleiche die zur Beurteilung der jeweils 8 Objekte erforderlich sind bei keinem signifikante Abweichungen. Die Hypothese 2 muss somit anhand der vorliegenden Daten abgelehnt werden, und es ist davon auszugehen, dass „position effects“ keine Bedeutung besitzen.

	Position	r	ABW
Schokolade (n=53)	(A,B)	<b>.909</b>	.259
	(B,A)		
Städte (n=49)	(A,B)	<b>.982</b>	.341
	(B,A)		

Tab. 9: *Effekte der Objektpositionen*

**(c) Einfluss der Beurteilungsunsicherheit der Befragten auf die MDS-Ergebnisse**

Die Analyseergebnisse zum Einfluss der Beurteilungsunsicherheit der Befragten auf die Güte der MDS-Ergebnisse sind in Tab. 10 zusammengefasst:

Der in der **Hypothese 3a** vermutende Effekt, dass durch die Kennzeichnung unsicherer Angaben die Anzahl fehlender Werte reduziert werden kann, kann nicht bestätigt werden, da sich über alle Objektgruppen betrachtet die Anzahl der durchschnittlich fehlenden Werte mit Kennzeichnungsoption (1.07) von der ohne Kennzeichnungsoption (1.00) nicht signifikant unterscheidet. Weiterhin stellt sich heraus, dass die Kennzeichnungsoption nur in 1.8% der Fälle auch von den Befragten genutzt wurde, was einerseits auf die allgemeine Bekanntheit der zu bewertenden Objekte schließen lässt. Es ist andererseits aber auch denkbar, dass die Befragten, nachdem sie ihre Einschätzung hinsichtlich der Objektähnlichkeit getroffen haben, nicht mehr gewillt waren, noch einmal eine zusätzliche Angabe zu tätigen. Aus diesem Grunde wurden lediglich den 18 Personen mit einem substantiellen Anteil von mindestens fünf gekennzeichneten Paarvergleichen zwei Karten, einmal unter Verwendung (voll) und zusätzlich unter Ausschluss (reduziert) der als unsicher gekennzeichneten Werte, vorgelegt. Dabei zeigten sich jedoch entgegen der **Hypothese 3b** keine bedeutsamen Effekte hinsichtlich der beurteilten Angemessenheit der Konfiguration. Beide Konfigurationen, die innerhalb des Fragebogens randomisiert angeordnet wurden, erfuhren im Mittel ähnliche Bewertungen (voll=2.79, reduziert=2.84), wobei die anhand der vollständigen Daten erstellten Konfigurationen in 8 gegenüber 6 Fällen bevorzugt wurden, bei 4 indifferenten Fällen. Die geringen Unterschiede können dabei darauf zurückgeführt werden, dass als sehr unsicher eingestufte Paarvergleiche größtenteils erwartungstreu auch nur Ähnlichkeitswerte im Bereich des Skalenzent-

rum von durchschnittlich 3.42 erhielten, die auf die Struktur der erzielten MDS-Konfigurationen einen nur sehr geringen Einfluss ausüben.

Trotz der allgemein geringen Resonanz bzw. Verwendung der Kennzeichnungsoption stellt sich eine Besonderheit heraus: So sind die bei dieser Testgruppe erhobenen Daten durchweg besser reproduzierbar, was sich objektübergreifend einerseits anhand der statistischen Kennzahlen (Stress und r) zeigt und zusätzlich in den positiveren Probandenurteilen (INDIV) widerspiegelt. Auch wenn diese Ergebnisse wiederum keine hohe statistische Signifikanz aufweisen, so stellen sie ein konsistentes Muster dar und sprechen daher für eine Steigerung der Datenqualität, die über die Verwendung der Kennzeichnungsoption erzielt werden kann. Dieser Effekt ist dabei vermutlich auf die zweigeteilte Fragestellung, die die Beurteilung einerseits der Objektähnlichkeit und zusätzlich der Bewertungssicherheit betrifft, zurückzuführen. Sie bewirkt, dass die Probanden insgesamt anscheinend intensiver über die einzelnen Fragestellungen nachdenken und hierbei verlässlichere Angaben treffen.

	Berücksichtigung der Unsicherheit	durchschn. fehlende Werte	Stress	p	r	p	INDIV	p
<b>Schokolade (n=53)</b>	Ja	1.31	.040	n.s.	.842	n.s.	2.80	n.s.
	Nein	0.88	.043		.828		2.61	
<b>Städte (n=49)</b>	Ja	0.83	.039	n.s.	.860	.060	2.44	n.s.
	Nein	1.12	.044		.822		2.86	
<b>Gesamt (n=102)</b>	Ja	<b>1.07</b>	.039	n.s.	.850	.078	2.64	n.s.
	Nein	<b>1.00</b>	.044		.824		2.74	

Tab. 10: Effekte der Beurteilungsunsicherheit

## 5 Zusammenfassung und kritische Würdigung der erzielten Untersuchungsergebnisse

Die Ergebnisse zur probandenbezogenen Ergebnisevaluation zeigen, dass Probanden die anhand ihrer Angaben erstellten Wahrnehmungskarten identifizieren können. Dies zeigt einerseits, dass die Wahrnehmungen zumindest für die untersuchten Bereiche stark variieren und eine Betrachtung lediglich der aggregierten Angaben, wie in der Anwendung der MDS meist üblich, hier nicht angemessen erscheint. Weiterhin zeigt sich, dass die klassischen Gütemaße mit den Bewertungen der Probanden

größtenteils einhergehen (vgl. die Tabellen 5, 7 und 10), was als Nachweis für deren Eignung zur Beurteilung von MDS-Konfigurationen angesehen werden kann.

Die Erkenntnisse der Untersuchungen zur Eignung verschiedener Skalen widersprechen teilweise der traditionellen Literatur zur MDS. Den vorliegenden Ergebnissen folgend, sind sowohl die 7er Skala, als auch die Magnitudeskalierung am besten geeignet. Die 9er Skala liefert, vermutlich aufgrund der Ungeübtheit im Umgang mit dieser Skala, insgesamt die schlechtesten Ergebnisse und sollte zumindest für Anwendungen in Deutschland nicht verwendet werden. Die Vorabnennung aller Objekte vor Beginn der eigentlichen Beurteilung steigert die Sicherheit der Probanden und die empfundene Angemessenheit der verwendeten Skala, wenngleich diese keinen nennenswerten Effekt auf die Struktur der Ähnlichkeitsdaten ausübt. Entgegen der Intuition waren die mit Vorabnennung der Objekte erzielten Konfigurationen tendenziell jedoch schlechter als ohne, weshalb dieses Vorgehen nicht uneingeschränkt zu empfehlen ist. Weiterhin konnte gezeigt werden, dass die Position der Objekte im Paarvergleich weder einen Einfluss auf die durchschnittlichen Proximitätsdaten noch auf die Konfigurationsgüte ausübt. Die Anwendung einer Randomisierung der Positionen über die Probanden erscheint somit nicht zwingend erforderlich und kann bei der Planung der Erhebung vernachlässigt werden.

Bezüglich der Probandenunsicherheit zeigte sich, dass die Bereitstellung einer Kennzeichnungsoption für sehr unsichere Angaben, wenngleich diese in der vorgestellten Studie probandenseitig kaum genutzt wurde, einen positiven Einfluss auf die Güte der MDS-Konfiguration ausübt. Dieser Effekt ist konsistent und bestätigt sich sowohl anhand der Stress- und Korrelationswerte als auch nach der probandenbezogenen Evaluierung. Der Anteil fehlender Werte war mit unter 1% insgesamt sehr gering und unabhängig von der Option zur Kennzeichnung unsicherer Beurteilungen. Dieser Effekt ist in der vorliegenden Untersuchung sicherlich auch auf eine hohe Motivation der an der Befragung teilnehmenden Studierenden zurückzuführen, die bei Erhebungen unter nicht universitären Bedingungen in dieser starken Form wahrscheinlich nicht vorliegen dürften. Ist dies der Fall, so ist davon auszugehen, dass von der Kennzeichnungsoption zusätzlich ein positiver Effekt auf das Ausmaß fehlender Werte ausgeht. Weiterhin sind die Objektgruppen in dieser Untersuchung gezielt dergestalt gewählt worden, dass alle Objekte auch von einem überwiegenden Teil der Studenten verlässlich eingeschätzt werden konnten. Dies lag in der Zielsetzung der Befragung begründet, da hier vielmehr methodische Aspekte im Vorder-

grund standen und weniger inhaltliche Aussagen über die Untersuchungsobjekte abgeleitet werden sollten. Diese Besonderheit bei der Objektwahl zeichnete sich höchstwahrscheinlich für die nur geringen Effekte, die aus der Vorabnennung der Beurteilungsobjekte und der Unsicherheitskennzeichnung resultieren, verantwortlich. Bei einer Deduktion von Handlungsempfehlungen für die praktische Anwendung ist dies zu berücksichtigen, da zu vermuten ist, dass derartige Gestaltungsmaßnahmen hier einen stärkeren Effekt aufweisen.

	<b>Beurteilungssicherheit der Probanden</b>	
	<b>hoch</b>	<b>gering</b>
<b>Skalierung</b>	7-stufige Ratingskala, Magnitudeskalierung	5-stufige Ratingskala Magnitudeskalierung
<b>Vorabnennung der Objekte</b>	optional	
<b>Unsicherheits- kennzeichnung</b>	optional, eher sinnvoll	
<b>Berücksichtigung von Positionseffekte</b>	nicht relevant	
<b>Berücksichtigung von Reihenfolgeneffekte</b>	nicht relevant	

Tab. 11: *Empfohlene Gestaltungsmaßnahmen bei MDS-Datenerhebungen*

Da sich der überwiegende Teil der Effekte statistisch als nicht signifikant darstellte, basieren die vorangegangenen Empfehlungen größtenteils auf unterschiedlich stark ausgeprägten Tendenzaussagen. Insbesondere die Erkenntnisse zur Vorabnennung der Objekte, die der Intuition widersprechen, sollten anhand weiterer Studien geprüft werden. Darüber hinaus stellen die gewonnenen Erkenntnisse zur Skalierung ein Feld für weitere Forschungsbemühungen dar, wobei hier die im Rahmen der metrischen MDS bisher nicht untersuchte Magnitudeskalierung besonders interessant erscheint.

## Literaturverzeichnis

- ADM (2005): *Arbeitskreis Deutscher Marktforschungsinstitute, Jahresbericht 2005*, www.adm-ev.de, (Stand 13.06.2007).
- Allais, M. (1953): Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école Américaine. *Econometrica*, 21, S. 503-546.
- Backhaus, K./Erichson, B./Plinke, W./Weiber, R. (2006): *Multivariate Analysemethoden*. 11. Aufl., Berlin: Springer.
- Behrens, G. (1983): Magnitudeskalierung. In Forschungsgruppe Konsum und Verhalten (Hrsg.), *Innovative Marktforschung* (S. 125-137), Würzburg: Physica.
- Bigne, E./Vila-Lopez, N./Küster-Bolude, I. (2000): Competitive positioning and market orientation: two interrelated constructs. *European Journal of Innovation Management*, 3, S. 190-198.
- Borg, I./Groenen, P. (1997): *Modern Multidimensionale Scaling*. Berlin: Springer.
- Carroll, J.D./Green, P.E. (1997): Psychometric Methods in Marketing Research: Part II, Multidimensional Scaling. *Journal of Marketing Research*, 34, S. 193-204.
- Coombs, C.H. (1950): Psychological scaling without a unit of measure. *Psychological Review*, 57, S. 145-158.
- Coombs, C.H. (1964): *A Theory of Data*, New York: Wiley.
- Cox, T.F./Cox, M.A.A. (1994): *Multidimensional Scaling*. London: Chapman and Hall.
- Cronbach, L.J./Meehl, P.E. (1955): Construct Validity in Psychological Tests. *Psychological Bulletin*, 52, S. 281-302.
- Cronbach, L.J. (1951): Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, S. 297-335.
- Day, S.G./Deutscher, T./Ryans, A.B. (1976): Data Quality, Level of Aggregation, and Nonmetric Multidimensional Scaling Solutions. *Journal of Marketing Research*, 13, S. 92-97.
- Decker, R. (2001): Empirischer Vergleich alternativer Ansätze zur Verbundanalyse im Marketing, *Working Paper No. 1/01*, Chair of Marketing, University of Bielefeld.
- DeSarbo, W.S./Young, M.R./Rangaswamy, A. (1997): A Parametric Multidimensional Unfolding Procedure for Incomplete Nonmetric Preference & Choice Set Data in Marketing Research. *Journal of Marketing Research*, 34, S. 499-516.

- Dong, H.K. (1983): Method of Complete Triads: An Investigation of Unreliability in Multidimensional Perception of Nations. *Multivariate Behavioral Research*, 18, S. 85-96.
- Fechner, G.T. (1877): *In Sachen der Psychophysik*. Leipzig: Breitkopf & Härtel.
- Gosh, A.K./Chakraborty, G. (2004): Using positioning models to measure and manage brand uncertainty. *Journal of Product & Brand Management*, 13, S. 294-302.
- Green, P.E./Carmone, F.J. (1968): The Performance Structure of the Computer Market: A Multivariate Approach. *The Economic and Business Bulletin*. 20, S. 1-11.
- Green, P.E./Rao, V.R. (1972): *Applied Multidimensional Scaling*. Hinsdale: Dryden Press.
- Green, P.E./Tull, D.S. (1982): *Methoden und Techniken der Marketingforschung*. 4. Aufl., Stuttgart: C.E. Poeschel Verlag.
- Green, P.E./Wind, Y. (1973): *Multiattribute Decisions in Marketing: A Measurement Approach*. Hinsdale: Dryden Press.
- Green, P.E. (1975): On the Robustness of Multidimensional Scaling Techniques. *Journal of Marketing Research*, 12, S. 71-83.
- Green, R.S./Bentler, P.M. (1979): Improving the efficiency and effectiveness of interactively selected MDS data designs. *Psychometrika*, 44, S. 115-119.
- Grunert, K.G. (1983): Magnitude-Skalierung. *Marketing ZFP*, 3, S. 185-189.
- Gulliksen, H. (1946): Paired comparisons and the logic of measurement. *Psychological Review*, 53, S. 199-213.
- Hair, J.F./Anderson, R.E./Tatham, R.L./Black, W.C. (1998): *Multivariate Data Analysis*, 5. Aufl., New Jersey: Prentice-Hall.
- Hodgkinson, G.P./Padmore, J./Tomes, A.E. (1990): Mapping Consumers' Cognitive Structures: A Comparison of Similarity Trees with Multidimensional Scaling and Cluster Analysis. *European Journal of Marketing*, 25, S. 41-60.
- Huber, J./Holbrook, M.B. (1982): Estimating Temporal Trends in Preferences Measured by Graded Paired Comparisons. *Journal of Business Research*, 10, S. 459-473.
- Huber, S. (2000): Nonmetrische Multidimensionale Skalierung (NMDS) als Messmodell für subjektive Sichtweisen der Religiosität. In Henning, C./Nestler, E. (Hrsg.), *Religionspsychologie heute* (S. 361-377). Frankfurt am Main: Verlag Peter Lang.

- Hussey, M./Hooley, G. (1995): The diffusion of quantitative methods into marketing management. *Journal of Marketing Practice. Applied Marketing Science*, 1, S. 13-31.
- Hüttner, M. (1997): *Grundzüge der Marktforschung*. 5. Aufl., München: Oldenbourg.
- Jain, A.K./Pinson, C. (1976): The Effect of Order of Presentation of Similarity Judgements on Multidimensional Scaling Results: An Empirical Examination. *Journal of Marketing Research*, 13, S. 435-439.
- Kelly, G.A. (1955): *The psychology of personal constructs*. New York: Norton.
- Krantz, D.H./Tversky, A. (1975): Similarity of Rectangles: An Analysis of Subjective Dimensions. *Journal of Mathematical Psychology*, 12, S. 4-34.
- Kruskal, J.B./Carmone, F.J. (1973): *How to Use MDSCAL, A Program to Multidimensional Scaling and Multidimensional Unfolding* (Version 5M), Bell Laboratories, New York.
- Kruskal, J.B. (1964a): Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29, S. 1-27.
- Kruskal, J.B. (1964b): Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 29, S. 115-129.
- Kupke, V. (2004): Identifying the dimensions to retail centre image. *Journal of Property Investment & Finance*, 22, S. 298-306.
- Lilian, G./Rangaswamy, A. (1997): *Marketing engineering. Computer assisted market analysis and planning*, Reading: Addison-Wesley.
- Lodge, M. (1981): *Magnitude Scaling*. London: Sage.
- MacCallum, R.C. (1978): Recovery of structure in incomplete data by ALSCAL. *Psychometrika*, 42, S. 69-74.
- Malhotra, N.K. (1987): Validity and Structural Reliability of Multidimensional Scaling. *Journal of Marketing Research*, 24, S. 164-173.
- Malhotra, N.K./Jain A.K./Pinson, C. (1988): The Robustness of MDS Configurations in the Case of Incomplete Data. *Journal of Marketing Research*, 25, S. 95-102.
- Mathar, R. (1997): *Multidimensionale Skalierung*, Stuttgart: Teubner.
- Miller, G. A. (1956): The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, S. 81-97.
- Mühlhaus, D./Hörstrup, R. (2009): Konzeption und Evaluation reduzierter Abfragedesigns für MDS-Anwendungen, Forschungsbericht Nr. 9, hrsg. von Rolf Weiber, Trier.

- Neibecker, B. (1983a): Elektronische Datenerhebung: Computergestützte Reaktionsmessung. In Forschungsgruppe Konsum und Verhalten (Hrsg.), *Innovative Marktforschung* (S. 209-235). Würzburg: Physica.
- Neibecker, B. (1983b): Computerkontrollierte Magnitudeskalierung: Eine kausalanalytische Validierung. *Marketing ZFP*, 3, S. 185-189.
- Neibecker, B. (1984): The Validity of Computer-Controlled Magnitude Scaling to Measure Emotional Impact of Stimuli. *Journal of Marketing Research*, 21, S. 325-331.
- Nunnally J.C. (1967): *Psychometric theory*. New York: McGraw Hill.
- Peter, J.P. (1979): Reliability: A Review of Psychometric Basics and Recent Marketing Practices. *Journal of Marketing Research*, 16, S. 6-17.
- Phillips, D.M./Phillips, J.K. (1998): A social network analysis of business logistics and transportation. *International Journal of Physical Distribution & Logistics Management*, 28, S. 328-348.
- Richardson, M.W. (1938): Multidimensional psychophysics. *Psychological Bulletin*, 35, S. 659-660.
- Rosch, E. (1975): Cognitive Reference Points. *Cognitive Psychology*, 7, S. 532-547.
- Rothkopf, E.Z. (1957): A measure of stimulus similarity and errors in some paired-associate learning tasks. *Journal of Experimental Psychology*, 53, S. 94-101.
- Schiffman, S.S./Reynolds, M.L./Young, F.W. (1981): *Introduction to Multidimensional Scaling*, London: Academic Press.
- Shepard, R.N. (1963): Analysis of proximities as a technique for the study of information processing in man. *Human Factors*, 5, S. 33-48.
- Shepard, R.N. (1964): Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, 1, S. 54-87.
- Shepard, R.N. (1966): Metric structure in ordinal data. *Journal of Mathematical Psychology*, 3, S. 287-315.
- Shepard, R.N. (1974): Representation of structure in similarity data: Problems and prospects. *Psychometrika*, 39, S. 373-421.
- Spence, I./Domoney, D.W. (1974): Single subject incomplete designs for nonmetric multidimensional scaling. *Psychometrika*, 39, S. 469-490.
- Stevens, S.S. (1975): *Psychophysics. Introduction to Its Perceptual, Neural, and Social Prospects*. New York: Wiley.

- Summers, J.O./MacKay, D.B. (1976): On the Validity and Reliability of Direct Similarity Judgments. *Journal of Marketing Research*, 13, S. 289-295.
- Takane, Y./Young, F.W./De Leeuw, J. (1977): Nonmetric Individual Differences Multidimensional Scaling: An Alternating Least Squares Method with Optimal Scaling Features. *Psychometrika*, 28, S. 7-67.
- Taylor, J.R./Kinnear, T.C. (1971): Empirical Comparison of Alternative Methods for Collecting Proximity Judgments, *Working Paper No. 46*, University of Michigan.
- Torgerson, W.S. (1952): Multidimensional scaling: I. Theory and method. *Psychometrika*, 17, S. 401-419.
- Torgerson, W.S. (1965): Multidimensional scaling of similarity. *Psychometrika*, 30, S. 333-367.
- Tschudi, F. (1972): The latent, the manifest, and the reconstructed in multivariate data reduction models, *unpublished Ph.D. thesis*, University of Oslo.
- Tversky, A. (1977): Features of Similarity. *Psychological Review*, 84, S. 327-352.
- Waheeduzzaman, A./Krampf, R. (1992): Use of Quantitative Techniques in Marketing Research: the Past Twenty Five Years. In Allen, C./Madden, T./Shimp, T. (Hrsg.), *Marketing Theory and Applications* (S. 285-294). Chicago: 1992 AMA Winter Educator's Conference.
- Weis, H.C./Steinmetz, P. (2005): *Marktforschung*. 6. Aufl., Ludwigshafen: Kiehl.
- Wilkes, R.E./Wilcox, J.B. (1977): On the Validity and Reliability of Direct Similarity Judgments: A Comment. *Journal of Marketing Research*, 14, S. 261-262.
- Wind, Y./Robinson, P.J. (1972): Product Positioning: An Application of Multidimensional Scaling. In Haley, R.I. (Hrsg.), *Attitude Research in Transition* (S. 155-175). Chicago: American Marketing Association.
- Wunder, E. (2001): Die Wahrnehmung der Struktur der deutschsprachigen UFO-Szene: eine multidimensionale Skalierung von Expertenurteilen. *Zeitschrift für Anomalistik*, 1, S. 75.
- Young G.W./Householder, A.S. (1938): Discussion of a set of points in terms of their mutual distances. *Psychometrika*, 1, S. 19-22.
- Young, F.W./Hamer, R.M. (1987): *Multidimensional Scaling*, New Jersey: Erlbaum Associates.

Zerr, K. (2003): Online-Marktforschung - Erscheinungsformen und Nutzenpotentiale.  
In Theobald, A./Dreyer, M./Starsetzki, T. (Hrsg.), *Online- Marktforschung*  
(S. 8-26). 2. Aufl., Wiesbaden: Gabler.

## Forschungsberichte zum Marketing

- Nr. 1: Weiber, Rolf/Jacob, Frank:  
Informationsgewinnung im Business-to-Business-Marketing: Grundlagen der Marktforschung, 2. Aufl., Berlin/Trier 2000.  
89 Seiten.
- Nr. 2: Weiber, Rolf/Kollmann, Tobias/Pohl, Alexander:  
Das Management technologischer Innovationen, 2. Aufl., Berlin/Trier 1999.  
128 Seiten.
- Nr. 3: Weiber, Rolf/Kollmann, Tobias:  
Die Vermarktung von Multimedia-Diensten - Akzeptanzprobleme bei interaktivem Fernsehen, Trier 1995. 67 Seiten. (ISBN 3-930230-08-9)
- Nr. 4: Weiber, Rolf/Kollmann, Tobias:  
Akzeptanz von interaktiven Multimedia-Programmen im universitären Einsatz, Trier 1997. 67 Seiten. (ISBN 3-930230-14-3)
- Nr. 5: Weiber, Rolf/Adler, Jost:  
Ausgewählte Probleme des internationalen Marketing, 2. Aufl., Berlin/Trier 1999. 109 Seiten.
- Nr. 6: Kirschner, Uwe C. M.:  
Zur Notwendigkeit der Neupositionierung schienenengebundener Mobilitätsanbieter: Eine strategische Analyse am Beispiel der Deutschen Bahn Reise&Touristik AG, Trier 2000. 180 Seiten. (ISBN 3-930230-17-8)
- Nr. 7: Pohl, Alexander/Mühlhaus, Daniel:  
Modernes Innovationsmarketing im Kontext von Open Innovation, Trier 2008.  
167 Seiten. (ISBN 3-930230-26-7)
- Nr. 8: Hörstrup, R./Mühlhaus, D. (2009):  
Der Einfluss von Untersuchungsrahmen und Messmethodik auf die Güte von MDS-Konfigurationen, Forschungsbericht Nr. 8, hrsg. von Rolf Weiber, Trier.  
40 Seiten. (ISBN 3-930230-27-5)
- Nr. 9: Mühlhaus, D./Hörstrup, R. (2009):  
Konzeption und Evaluation reduzierter Abfragedesigns für MDS-Anwendungen, Forschungsbericht Nr. 9, hrsg. von Rolf Weiber, Trier.  
43 Seiten. (ISBN 3-930230-28-3)

## **Arbeitspapiere zur Marketingtheorie**

- Nr. 1: Weiber, Rolf:  
Was ist Marketing - Ein informationsökonomischer Erklärungsansatz, 2. Aufl., Trier 1996. 120 Seiten. (ISBN 3-930230-06-2)
- Nr. 2: Weiber, Rolf/Pohl, Alexander:  
Leapfrogging bei der Adoption neuer Technologien - Theoretische Fundierung und empirische Prüfung, 2. Aufl., Trier 1994. 60 Seiten. (ISBN 3-930230-02-X)
- Nr. 3: Adler, Jost:  
Informationsökonomische Fundierung von Austauschprozessen im Marketing, Trier 1994. 110 Seiten. (ISBN 3-930230-03-8)
- Nr. 4: Pohl, Alexander:  
Ausgewählte Theorieansätze zur Erklärung des Nachfragerverhaltens bei technologischen Innovationen, Trier 1994. 163 Seiten. (ISBN 3-930230-04-6)
- Nr. 5: Beinlich, Georg:  
Geschäftsbeziehungen - Ein integrativer Überblick auf Basis der politischen Ökonomie, Trier 1995. 69 Seiten. (ISBN 3-930230-05-4)
- Nr. 7: Kollmann, Tobias:  
Die Akzeptanz technologischer Innovationen - eine absatztheoretische Fundierung am Beispiel von Multimedia-Systemen, Trier 1996. 146 Seiten. (ISBN 3-930230-12-7)
- Nr. 8: Raff, Tilmann:  
Informationsökonomische Fundierung nachfragerseitiger Unsicherheitspositionen im Systemgeschäft, Trier 1998. 105 Seiten. (ISBN 3-930230-15-1)
- Nr. 9: McLachlan, Christopher:  
Die Gestaltung von Informationsasymmetrien durch das Marketing, Trier 1999. 114 Seiten. (ISBN 3-930230-16-X)