

Westfälische Wilhelms-Universität Münster

Philosophische Fakultät

**Multiattributive Präferenzstrukturmodellierung durch die
Conjoint Analyse.**

**Diskussion der Verfahrensmöglichkeiten und Optimierung von
Paarvergleichsaufgaben bei der adaptiven Conjoint Analyse.**

**Dissertationsschrift
zur Erlangung des Doktorgrades**

Wolfram Reiners

Westfälische Wilhelms-Universität Münster

Psychologisches Institut IV

Fliednerstraße 21

48149 Münster

Tel. (0251) 83 9140

✉ REINERS@PSY.UNI-MUENSTER.DE

Datum der Abgabe: 09. Mai 1996

Begutachtung: Prof. Dr. Heinz Holling
Westfälische Wilhelms-Universität Münster

Zweitbegutachtung: Prof. Dr. Wolfgang Hell
Westfälische Wilhelms-Universität Münster

Dekan: Prof. Dr. Hortschansky
Westfälische Wilhelms-Universität Münster

Datum der mündlichen Prüfungen: 12. Juli 1996

S t u f e n

Wie jede Blüte welkt und jede Jugend
Dem Alter weicht, blüht jede Lebensstufe,
Blüht jede Weisheit auch und jede Tugend
Zu ihrer Zeit und darf nicht ewig dauern.
Es muß das Herz bei jedem Lebensrufe
Bereit zum Abschied sein und Neubeginne,
Um sich in Tapferkeit und ohne Trauern
In andre, neue Bindungen zu geben.
Und jedem Anfang wohnt ein Zauber inne,
Der uns beschützt und der uns hilft, zu leben.

Wir sollen heiter Raum um Raum durchschreiten,
An keinem wie an einer Heimat hängen,
Der Weltgeist will nicht fesseln uns und engen,
Er will uns Stuf' um Stufe heben, weiten.
Kaum sind wir heimisch einem Lebenskreise
Und traulich eingewohnt, so droht Erschlaffen,
Nur wer bereit zu Aufbruch ist und Reise,
Mag lähmender Gewöhnung sich entrafen.

Es wird vielleicht auch noch die Todesstunde
Uns neuen Räumen jung entgegenschicken,
Des Lebens Ruf an uns wird niemals enden ...
Wohlan denn, Herz, nimm Abschied und gesunde!

Vorwort

Zu den Erkenntnissen, die in dieser Arbeit festgehalten sind, haben mehrere Kollegen entscheidend beigetragen. Dies liegt in der Natur der Sache, da effiziente Forschung nur im Team erfolgen kann. Bei der Planung der Forschungsfragen und empirischen Untersuchungen waren insbesondere mein Doktorvater Prof. Dr. Heinz Holling und Dipl.-Psych. Bernd Runde beteiligt. Ohne die Beiträge von Dipl.-Math. Andreas Jütting bei der Softwareentwicklung hätten die Untersuchungen in der vorliegenden Form nicht stattfinden können.

Wichtige Hilfen im Rahmen der Durchführung der empirischen Untersuchungen gaben mir die Diplomanden cand. psych. Tobias Arens, cand. psych. Georgios Melikidis und cand. psych. Torsten Melles.

Einige meiner Kollegen haben diese Arbeit während ihrer Entstehung mindestens teilweise gelesen. Für die wertvolle Kritik und die unzähligen konstruktiven Vorschläge danke ich meinem Doktorvater Prof. Dr. Heinz Holling, Petra Kleining, Dipl.-Psych. Ira Kokavec, Dipl.-Psych. Christoph Nienaber, Dipl.-Psych. Bernd Runde und meinem Vater Gerd Reiners.

Inhaltsverzeichnis

VORWORT	V
INHALTSVERZEICHNIS	VI
1 EINLEITUNG	1
2 KONZEPT DER CONJOINT ANALYSE	5
2.1 Prinzip und Ergebnisse einer typischen CA-Untersuchung	5
2.2 Multiattributive Präferenzstrukturmodellierung: Grundbegriffe	8
2.2.1 Zentrale Begriffe	8
2.2.2 Notation	11
2.2.2.1 Grundbausteine der Präferenzstrukturmodelle	11
2.2.2.2 Objektdefinition	11
2.2.2.3 Das Präferenzintegrationsmodell	11
2.2.2.4 Das Präferenzmodell.....	12
2.2.2.5 Das Präferenzstrukturmodell	13
2.3 Nutzenschätzung bei additiven Modellen	13
2.3.1 Grundüberlegung.....	14
2.3.2 SMART als Beispiel für ein direktes Nutzenschätzverfahren.....	14
2.3.3 Ein indirektes Nutzenschätzverfahren: Das Conjoint Measurement.....	16
2.4 Wo ist die Conjoint Analyse anzusiedeln?	18
2.4.1 Conjoint Analyse und Conjoint Measurement	18
2.4.2 Conjoint Analyse und kompositorische Modellierung	19
2.4.3 Conjoint Analyse und Functional Measurement.....	20
3 PROZEß DER CONJOINT ANALYSE	22
3.1 Ziel und Setting einer CA-Untersuchung	23
3.1.1 Einstellung, Präferenz, Verhaltensabsicht und Verhalten	24
3.1.2 Präferenz und Verhalten	27
3.1.3 Auskunftspersonen.....	29
3.2 Attribute	30
3.2.1 Grundlegende Anforderungen an das Attributset.....	31

3.2.2 Art der Attribute	32
3.2.2.1 Potentiell relevante Attribute	33
3.2.2.2 Determinante Attribute	33
3.2.2.3 Empirisch unabhängige Attribute	34
3.2.2.4 Präferenzunabhängige Attribute	34
3.2.2.5 Relevante Attribute	35
3.2.2.6 Maßgebliche Attribute	35
3.2.2.7 Fazit	36
3.2.3 Verfahren zur Gewinnung potentiell relevanter Attribute	36
3.2.3.1 Direkte Befragung	37
3.2.3.2 Elicitationstechnik	38
3.2.3.3 Gruppeninterview und Gruppendiskussion	38
3.2.3.4 Kreativitätstechniken	38
3.2.3.5 Repertory Grid Technik	39
3.2.3.6 Lautes Denken	40
3.2.3.7 Tiefeninterview	41
3.2.3.8 Projektive Verfahren	41
3.2.3.9 Multidimensionale Skalierung	41
3.2.4 Verfahren zur Selektion der Attribute	42
3.2.4.1 Grundüberlegungen	42
3.2.4.2 Stufenverfahren	43
3.2.5 Repräsentation der Attribute	47
3.3 Attributstufen	50
3.3.1 Endpunkteproblem	51
3.3.2 Abstufungsproblem	52
3.4 Präferenzstrukturmodell	54
3.4.1 Präferenzmodell	54
3.4.1.1 Lineares Vektormodell	55
3.4.1.2 Idealpunktmodell	55
3.4.1.3 Teilnutzenwertmodell	56
3.4.1.4 Anwendungen und Fazit	57
3.4.2 Präferenzintegrationsmodell	58
3.4.2.1 Kompensatorische Präferenzintegrationsmodelle	58
3.4.2.2 Paramorphe Repräsentation	60
3.4.2.3 Strukturangemessene Repräsentation	62
3.4.2.4 Fazit	64
3.4.3 Prozeßmodell	64
3.5 Beurteilungsobjekte	66
3.5.1 Aufwärmphase	67
3.5.2 Definition der Beurteilungsobjekte	69
3.5.2.1 Vollprofilmethode	69

3.5.2.2 Teilprofilmethoden.....	70
3.5.2.3 Vergleich Vollprofilmethode und Trade-off-Ansatz	71
3.5.2.4 Fazit.....	73
3.5.3 Konstruktion des Objektsets.....	74
3.5.3.1 Systematisch reduzierte Designs.....	74
3.5.3.2 Zufallsdesigns	75
3.5.3.3 Paarvergleichsdifferenzendesigns	75
3.5.3.4 Differenzendesigns: Fazit.....	81
3.5.4 Objektpräsentation	81
3.5.4.1 Objektpräsentationsmodus	82
3.5.4.2 Unterstützung der Objektpräsentation	86
3.5.4.3 Fazit.....	88
3.6 Beurteilungsaufgabe	88
3.6.1 Art des Urteils	88
3.6.2 Antwortmodus: Erhebung der abhängigen Variablen	90
3.6.2.1 Kategorienzuordnung.....	90
3.6.2.2 Rangreihung.....	91
3.6.2.3 Rating	92
3.6.2.4 Paarvergleichsverfahren	93
3.6.2.5 Validität und Anwendungshäufigkeit der Antwortmodi.....	96
3.6.2.6 Fazit.....	98
3.7 Verfahren zur Nutzenschätzung.....	98
3.7.1 Kompositorischer Ansatz - explizite Modellbildung	99
3.7.1.1 Kompensatorische, explizite Modelle.....	99
3.7.1.2 Konjunktiv-kompensatorische explizite Modellbildung.....	101
3.7.1.3 Probleme der kompositorischen Modellbildung	102
3.7.1.4 Übereinstimmung von kompositorischen und dekompositorischen Ansätzen und deren kriterienorientierte Validität.....	103
3.7.1.5 Fazit.....	106
3.7.2 Dekompositorischer Ansatz - Conjoint Analyse	106
3.7.2.1 Klassische Verfahren der CA.....	107
3.7.2.2 Hierarchische Informationsintegrations-Conjoint Analyse (HICA).....	107
3.7.2.3 Hybrid-Conjoint Analyse (HCA)	111
3.7.2.4 Adaptive Conjoint Analyse (ACA)	114

3.7.3 Vergleichende Diskussion der Verfahren hinsichtlich ihrer Validität.....	126
3.7.3.1 HCA gegenüber der expliziten Modellbildung und klassischen CA.....	126
3.7.3.2 ACA gegenüber der klassischen CA.....	128
3.7.4 Fazit.....	130
3.8 Parameterschätzverfahren.....	132
3.8.1 Überblick über die Verfahren.....	133
3.8.1 Das Parameterschätzverfahren bei ACA	135
4 WEITERVERARBEITUNG DER PRÄFERENZSTRUKTURMODELLE.....	137
4.1 Normierung.....	138
4.2. Aggregation	139
4.3 Segmentierung.....	141
4.4 Wahlanteilsimulation	143
4.4.1 Maximum-utility-Modell.....	144
4.4.2 BTL-Modell	145
4.4.3 Logit-Modell und ACA-Logit-Modell.....	146
4.4.4 Adjusted ACA-Logit-Modell	148
4.4.5 Logistisches Modell	149
4.4.6 Wahlanteilsimulation - Marktanteilsimulation	150
5 MEßGÜTEKRITERIEN DER CONJOINT ANALYSE.....	152
5.1 Subjektive und objektive Reaktion	152
5.2 Reliabilität.....	153
5.3 Modellanpassung.....	155
5.4 Kriterienorientierte Validität	156
5.4.1 Kriterienorientierte Validität auf individueller Ebene.....	158
5.4.2 Kriterienorientierte Validität auf aggregierter Ebene	162
5.4.3 Welche Art der Validierung und welche Maßzahl?	164
5.4.4 Prognostische Validität.....	166
5.5 Strukturwiedererkennung.....	167
5.5.1 Beurteilung der Strukturwiedererkennungsleistung	168
5.6 Fazit.....	169
6 EMPIRISCHER TEIL: OPTIMIERUNG DER ADAPTIVEN CONJOINT ANALYSE.....	171

6.1 Forschungssoftware ALASCA	172
6.1.1 Computersimulationen: Vergleich von ALASCA und ACA 4.0.....	174
6.2 Welche Effekte sind theoretisch zu erwarten?	175
6.3 Begründung und Darstellung des Forschungsdesigns	179
6.3.1 Anforderungen durch das Problem der Kriterienreliabilität	179
6.3.2 Empirische Untersuchung I	180
6.3.2.1 Methode	180
6.3.2.2 Ergebnisse	181
6.3.2.3 Diskussion	182
6.3.3 Strukturwiedererkennung bei empirischen Untersuchungen	184
6.3.4 Das Forschungsdesign.....	185
6.4 Empirische Untersuchung II: Meßgüte und Antwortfehler	
in Abhängigkeit der Anzahl der Paare und der Profilstärke	189
6.4.1 Methode.....	189
6.4.1.1 Die wahre Nutzenstruktur	189
6.4.1.2 Objektklasse.....	190
6.4.1.3 Attribute und Attributstufen.....	191
6.4.1.4 Auskunftspersonen und Untersuchungsrahmenbedingungen	192
6.4.1.5 Lernphase	193
6.4.1.6 Lernerfolgsüberprüfung und -rückmeldung.....	195
6.4.1.7 Untersuchungsaufbau und Versuchsplan.....	197
6.4.1.8 Parameterschätzung	204
6.4.2 Ergebnisse.....	204
6.4.2.1 Lernerfolg und Abhängigkeit der Ergebnisse vom Objektset	204
6.4.2.2 Stabilität der induzierten Nutzenstruktur	207
6.4.2.3 Retest-Reliabilität der Holdout-Aufgabe.....	209
6.4.2.4 Beantwortungszeiten der ALASCA-Paarvergleiche.....	210
6.4.2.5 Validität und Strukturwiedererkennung der Vollprofil-CA	212
6.4.2.6 Verlauf der Validität der ALASCA-Paarvergleichs-CA.....	212
6.4.2.7 Verlauf der Strukturwiedererkennung der ALASCA-Paarvergleichs-CA	219
6.4.2.8 Bestimmung des Antwortfehlers	222
6.4.2.9 Validität und Strukturwiedererkennung in Relation zur Antwortzeit.....	224
6.4.2.10 Vergleich der Meßgüte von Vollprofil-CA und ALASCA-Paarvergleichen	227
6.4.2.11 Wirkung der Profilstärke auf die kognitiven Ressourcen.....	229
6.4.2.12 Auswirkungen der kognitiven Leistungsfähigkeit	
und der Leistungsmotivation.....	231
6.5 Implikationen der Befunde und Diskussion.....	233
6.5.1 Die zentralen Ergebnisse	233
6.5.2 Aussagekraft und Reichweite der Befunde	236
6.5.3 Ausblick	238

6.6 Zusammenfassung	240
ANHANG	243
Anhang A Der schrittweise Ablauf einer Conjoint Analyse im Überblick	243
Anhang B Beispiel einer konjunktiv-kompositorischen expliziten Modellbildung nach Srinivasan (1988)	248
Anhang C Computersimulationen: Vergleich von ALASCA und ACA 4.0	250
Methode	250
Ergebnisse	252
Diskussion	256
Anhang D Die möglichen Objekte und deren Einsatz	257
Anhang E Schriftliche Instruktionen der Lernphase	260
Anhang F Die Aufgaben der dritten Stufe der Lernphase	263
Anhang G Mündliche Instruktionen der Lernphase	264
Anhang H Mündliche Instruktionen der Untersuchungsphase	265
Anhang I Items der Zwischenfragebögen und des Abschlußfragebogens	267
Anhang J LPS-Skala 12 zur Erfassung der allgemeinen Intelligenz	269
Anhang K MARPS-Skala zur Erfassung der Leistungsmotivation	269
Anhang L Ausführliche Ergebnisdarstellung der zentralen Varianzanalysen und Kovarianzanalysen	271
LITERATUR	275

Verzeichnis der Tabellen

Tabelle 2.1:	Die Attribute, deren Ausprägung und die Teilnutzenwerte eines fiktiven Kunden.	6
Tabelle 3.1:	Der Prozeß der Conjoint Analyse, typische Fragen und Probleme und die Gliederung des Abschnitts 3.	23
Tabelle 3.2:	Modellierungsziele einer Untersuchung und die jeweiligen kognitiven Repräsentationen der Objekte und der Situation als Beurteilungsgrundlage.	26
Tabelle 3.3:	Ein fiktives Beispiel des Objekte-Beurteilungsdimensionen-Gitters... 40	
Tabelle 3.4:	Relative Häufigkeit der Verwendung verschiedener Objektpräsentationsmodi bei der CA in kommerziellen Studien [K] und Studien zu Forschungszwecken [F]. Alle Angaben in %.....	86
Tabelle 3.5:	Relative Häufigkeit der Verwendung verschiedener Antwortmodi bei der CA in kommerziellen Studien [K] und Studien zu Forschungszwecken [F]. Alle Angaben in %.	97
Tabelle 3.6:	Die kriterienorientierte Validität der Studien von Green, Krieger und Agarwal (1991) und Green, Schaffer und Patterson (1991) im Überblick. Angegeben ist die Rangkorrelation, der Prozentsatz von Erstewahl-Hits und der Prozentsatz richtig vorhergesagter Rangplätze zu einem Kreuzvalidierungsobjektset (jeweils Mittelwert und Standardfehler).....	124
Tabelle 3.7:	Die wichtigsten Parameterschätzverfahren.....	134
Tabelle 3.8:	Relative Häufigkeit der Verwendung verschiedener Parameterschätzprozeduren bei der CA in kommerziellen Studien [K] und Studien zu Forschungszwecken [F]. Alle Angaben in %.....	134
Tabelle 4.1:	Überblick über die Weiterverarbeitung von Präferenzstrukturmodellen auf aggregierter Ebene.	137
Tabelle 5.1:	Systematisierung der Validierungsebene und der Art der Kriterien-gewinnung bei der Beurteilung der kriterienorientierten Validität von PSM. Die Validitätsmaßzahlen beziehen sich auf den Vergleich zwischen vorhergesagten und tatsächlichen Kriteriumsbeurteilungen. Weitere Erläuterungen im Text.	165
Tabelle 6.1:	Die TNW des wahren PSM in verschiedenen Darstellungsformen (Erläuterungen im Text).	190
Tabelle 6.2:	Die fünf Attribute mit jeweils drei Attributstufen und die wahre Nutzenstruktur.	191

Tabelle 6.3:	Zeitbedarfsäquivalenzen von Paarvergleichsaufgaben der Profilstärke 2, 3 und 4.	211
Tabelle 6.4:	Korrelation zwischen Indikatoren der kriterienorientierten Validität hinsichtlich der Holdout-Sets 1 und 2 und der Reliabilität des Kriteriums (Erläuterung der Reliabilitätsindikatoren im Text).	218
Tabelle 6.5:	Die Produkt-Moment-Korrelation r und r^2 zwischen der wahren und der empirischen Antwort auf die abgestuften Paarvergleiche. Die Maßzahlen sind getrennt nach Objektprofilstärke (2, 3, 4) und nach der Anzahl der Paarvergleiche (bis 10, 11-20, 21-30, 31-40) dargestellt.	223
Tabelle 6.6:	Die erwartete Anzahl der Paarvergleiche, die in einer bestimmten Zeiteinheit beantwortet werden. Die Maßzahlen der Meßgüte werden innerhalb der Apn über jeweils 5 Paarvergleiche gemittelt.	226
Tabelle 6.7:	Mittelwerte der Items 1, 6, 10 und 2.	230
Tabelle 6.8:	Mittelwerte der Items 18 und 19.	231
Tabelle A:	Der Prozeß der Conjoint Analyse und Verfahrensmöglichkeiten im Überblick.	243
Tabelle C-1:	Theoretische Auftretenswahrscheinlichkeit [in %] des zur Antwort addierten Fehlerwertes für verschiedene Antwortfehlerstufen beim Antwortsimulationsprogramm ACASIM.	250
Tabelle C-2:	Die TNW der simulierten wahren Modelle vom Typ I, II und III.	251
Tabelle D:	Die möglichen Objekte und deren Verwendung in verschiedenen Phasen der empirischen Untersuchung II.	257
Tabelle F:	Die dritte Stufe der Lernphase 'vergleichende Beurteilung von Wohnungskonzepten': Die Aufgaben sind in die Gruppen a bis e eingeteilt, je nachdem, auf wie vielen Attributen kompensatorische Urteile gefordert sind. Unterschiedliche Attributausprägungen einer Aufgabe sind kursiv gedruckt. Die Lösung der Aufgaben ist kurz erklärt.	263
Tabelle I-1:	Die Items der (identischen) Zwischenfragebögen.	267
Tabelle I-2:	Die Items des Abschlußfragebogens.	268
Tabelle L-1:	Die Ergebnisse der (2X4X3)-ANOVA aus Abschnitt 6.5.2.6.	271
Tabelle L-2:	Die Ergebnisse der (4X3)- ANOVA aus Abschnitt 6.5.2.6 (kriterienorient. Validität zu Holdout-Set 1).	271
Tabelle L-3:	Die Ergebnisse der (4X3)-ANOVA aus Abschnitt 6.5.2.6 (kriterienorient. Validität zu Holdout-Set 2).	272
Tabelle L-4:	Die Ergebnisse der (4X3)-ANOVA aus Abschnitt 6.5.2.7 (Strukturwiedererkennung).	272
Tabelle L-5:	Die Ergebnisse der (3X7)-ANOVA aus Abschnitt 6.5.2.9 (Strukturwiedererkennung/ Antwortzeit).	272

Tabelle L-6:	Die Ergebnisse der (4X3)-ANCOVA aus Abschnitt 6.5.2.12 mit der Kovariate 'LPS-Wert'	273
Tabelle L-7:	Die Ergebnisse der (4X3)-ANCOVA aus Abschnitt 6.5.2.12 mit der Kovariate 'Abiturnote'	273
Tabelle L-8:	Die Ergebnisse der (4X3)-ANCOVA aus Abschnitt 6.5.2.12 mit der Kovariate 'MARPS'	274

Verzeichnis der Abbildungen

Abbildung 2.1: ..Eine fiktive Vorhersage des Marktanteils.....	6
Abbildung 3.1: ..Beispiel einer schriftlichen, hoch strukturierten Objektbeschreibung.	83
Abbildung 3.2: ..Beispiel einer schriftlichen, weniger strukturierten Objektbeschreibung.....	83
Abbildung 3.3: ..Beispiel für die Beantwortung eines abgestuften Paarvergleichs auf zwei Antwortskalen.....	95
Abbildung 3.4: ..Beispiel für die Beantwortung eines abgestuften Paarvergleichs auf einer Antwortskala.....	95
Abbildung 3.5: ..Der Präferenzbildungsprozeß (a) und die Vorgehensweise bei der Präferenzstrukturmodellierung (b) mit der 'Brücken-CA'. Die gestrichelt gezeichneten Kästen sind conjointanalytische Subuntersuchungen.	108
Abbildung 3.6: ..Der Präferenzbildungsprozeß (a) und die Vorgehensweise bei der Präferenzstrukturmodellierung (b) mit der HICA. Die gestrichelt ge- zeichneten Kästen sind conjointanalytische Subuntersuchungen. ...	109
Abbildung 3.7: ..Der Präferenzbildungsprozeß (a) und die Vorgehensweise bei der Präferenzstrukturmodellierung (b) mit der erweiterten HICA. Die gestrichelt gezeichneten Kästen ist eine der drei conjointanalytischen Subuntersuchungen.....	111
Abbildung 5.1: ..Die Konzepte Reliabilität, Modellanpassung und Kreuzvalidität bei ähnlichen Conjoint Analysen I und II (Pfeil: Vorhersage, Doppelpfeil: Zusammenhang).....	155
Abbildung 5.2: ..Schematische Darstellung des Reliabilitätsproblems der Kriteriums- aufgabe. Individuell verschiedene Nutzenstrukturen haben zur Folge, daß die Schwierigkeit der Kriteriumsaufgabe - trotz objektiv gleicher Beurteilungsobjekte - individuell unterschiedlich ist.....	161
Abbildung 5.3: ..Zentrierung und Standardabweichungsnormierung eines wahren und eines geschätzten TNW-PSM. Anschließend kann eine Maßzahl der Strukturwiedererkennungslleistung (z.B. SSE) berechnet werden. ...	170
Abbildung 6.1: ..Die Bildschirmoberfläche des Programms ALASCA beim abgestuften Paarvergleich.	173
Abbildung 6.2: ..Der Zusammenhang zwischen der Anzahl der Paarvergleiche und der Meßgüte.....	176
Abbildung 6.3: ..Der Zusammenhang zwischen der Profilstärke und der Meßgüte. ...	176
Abbildung 6.4: ..Der Zusammenhang zwischen dem Antwortfehler und der Meßgüte.	176
Abbildung 6.5: ..Der Zusammenhang zwischen der Anzahl der Paarvergleiche und dem Antwortfehler.....	177

Abbildung 6.6: ..Der Zusammenhang zwischen der Objektprofilstärke und dem Antwortfehler.....	177
Abbildung 6.7: ..Histogramm der kriterienorientierten Validität gegenüber einem Holdout-Set. Dargestellt sind Rangkorrelationen (n=66).	182
Abbildung 6.8: ..Schematische Darstellung des Forschungsdesigns. Der gestrichelt dargestellte Kasten ist nur bei einem 'within subject design' vorhanden.	186
Abbildung 6.9: ..Schematische Darstellung der dritten Stufe der Lernphase (vergleichende Beurteilung von Wohnungskonzepten) mit seinen vier Lernblöcken und den vier Lernerfolgsmessungen A, B, C und D. ...	197
Abbildung 6.10: Schematischer Aufbau der empirischen Untersuchung. Erläuterungen im Text.	198
Abbildung 6.11: Mittlere Rangkorrelation der empirischen mit den wahren Rangreihen getrennt für die Versuchsbedingungen 1 und 2.	205
Abbildung 6.12: Histogramm der Rangkorrelation zwischen der wahren und der empirischen Objektrangreihe beim Holdout-Set 1.	207
Abbildung 6.13: Histogramm der Rangkorrelation zwischen der wahren und der empirischen Objektrangreihe beim Holdout-Set 2.	207
Abbildung 6.14: Die Beantwortungszeit [in s] der Paarvergleichsfragen gemittelt über alle Apn. Die Antwortzeiten sind getrennt für die Profilstärkebedingungen dargestellt.....	211
Abbildung 6.15: Kriterienorientierte Validität der ALASCA-PSM hinsichtlich des Holdout-Sets 1.	214
Abbildung 6.16: Kriterienorientierte Validität der ALASCA-PSM hinsichtlich des Holdout-Sets 2.	214
Abbildung 6.17: Kriterienorientierte Validität der ALASCA-PSM nach 10, 20, 30 und 40 Paarvergleichen hinsichtlich des Holdout-Sets 1 (über je 5 Paarvergleiche geglättet).....	216
Abbildung 6.18: Kriterienorientierte Validität der ALASCA-PSM nach 10, 20, 30 und 40 Paarvergleichen hinsichtlich des Holdout-Sets 2 (über je 5 Paarvergleiche geglättet).....	216
Abbildung 6.19: Die durchschnittliche Strukturwiedererkennungsleistung der geschätzten PSM für die Profilstärken 2, 3 und 4 in Abhängigkeit der Anzahl der Paarvergleiche.....	220
Abbildung 6.20: Strukturwiedererkennung der ALASCA-PSM nach 10, 20, 30 und 40 Paarvergleichen (über je 5 Paarvergleiche geglättet).....	220
Abbildung 6.21: Strukturwiedererkennung der ALASCA-PSM abgetragen über der durchschnittlichen Antwortzeit (über je 5 Paarvergleiche geglättet).226	

Abbildung C-1: . Typ I: Profilstärke 2, Antwortfehler $s=0$	253
Abbildung C-2: . Typ I: Profilstärke 4, Antwortfehler $s=0$	253
Abbildung C-3: . Typ I: Profilstärke 2, Antwortfehler $s=4$	253
Abbildung C-4: . Typ I: Profilstärke 4, Antwortfehler $s=4$	253
Abbildung C-5: . Typ II: Profilstärke 2, Antwortfehler $s=0$	253
Abbildung C-6: . Typ II: Profilstärke 4, Antwortfehler $s=0$	253
Abbildung C-7: . Typ II: Profilstärke 2, Antwortfehler $s=4$	254
Abbildung C-8: . Typ II: Profilstärke 4, Antwortfehler $s=4$	254
Abbildung C-9: . Typ III: Profilstärke 2, Antwortfehler $s=0$	254
Abbildung C-10: Typ III: Profilstärke 4, Antwortfehler $s=0$	254
Abbildung C-11: Typ III: Profilstärke 2, Antwortfehler $s=4$	254
Abbildung C-12: Typ III: Profilstärke 4, Antwortfehler $s=4$	254

Verzeichnis der verwendeten Abkürzungen

ANOVA	Varianzanalyse (<i>ANalysis Of VAriance</i>)
ANCOVA	Kovarianzanalyse (<i>ANalysis Of CovAriance</i>)
Ap(n)	Auskunftsperson(en)
CA	Conjoint Analyse
CM	Conjoint Measurement
FM	Functional Measurement
HII	Hierarchische Informationsintegrationstheorie
IIA	<i>Independence of Irrelevant Alternatives</i>
LINMAP	<i>Linear Programming Techniques for Multidimensional Analysis of Preference</i>
MAE	mittlerer absoluter Fehler (<i>Mean Absolute Error</i>)
MDS	multidimensionale Skalierung
MONANOVA	monotone Varianzanalyse (<i>Monotonic Analysis of Variance</i>)
MSAE	Minimierung der Summe der absoluten Fehler
MSE	mittlerer quadrierter Fehler (<i>Mean Square Error</i>)
OHD	orthogonales Haupteffektdesign
OLS	Minimierung der Summe der quadrierten Fehler (<i>Ordinary Least Square</i>)
PRMSEP	<i>Proportional Reduction in Mean Square Error of Prediction</i>
PSM	Präferenzstrukturmodell
SAE	Summe der absoluten Fehler (<i>Sum of Absolute Error</i>)
SSE	Summe der quadrierten Fehler (<i>Sum of Square Error</i>)
StaNorm	<i>Standardabweichungs-Normierung</i>
TMT	Telefon-Mail-Telefon Technik
TNW	Teilnutzenwert(e)

Einleitung

Die scheinbar einfachste Möglichkeit, Informationen über die Präferenzen eines Menschen zu gewinnen, besteht darin, die Person zu befragen. Möchte man beispielsweise ihre Präferenzen hinsichtlich verschiedener Wohnungen wissen, kann man sie bitten zu erzählen, in welcher Art und Weise sie Wohnungen beurteilt und in diesem Realitätsbereich Entscheidungen fällt. In aller Regel wird die so gewonnene Information recht unstrukturiert sein. Dies wird spätestens dann negativ auffallen, wenn man sich ein Bild der Präferenzen einer Vielzahl von Personen machen muß.

Es liegt nahe, die Vorgehensweise etwas zu systematisieren. Präferenzen bestehen nie losgelöst für sich, sondern immer gegenüber Objekten. Bei diesen Objekten kann es sich um Produkte, Dienstleistungen, Unternehmen, Personen, etc. handeln. Allen diesen Objekten ist gemeinsam, daß sie durch ihre Merkmale - oder ihre Ausprägung auf sogenannten Attributen - beschreibbar sind. Eine Wohnung kann man z.B. auf den Attributen Wohnfläche, Anzahl der Zimmer, Mietpreis, Alter, Ausstattung, Baustil des Hauses, Lage, ... charakterisieren. Das Objekt wird sozusagen mit einem Attributbündel gleichgesetzt. Menschliche Präferenzen setzen in unserem Denkmodell an einem solchen multiattributiven Objekt an. Macht man sich zur Aufgabe, die Präferenzen eines Menschen abzubilden, betreibt man "multiattributive Präferenzstrukturmodellierung".

Der Titel dieser Arbeit ist damit zur Hälfte erklärt. Die andere Hälfte heißt "... durch die Conjoint Analyse". Es ist damit angedeutet, daß es sich bei der Conjoint Analyse um eine *Verfahrensmöglichkeit* handelt, diese multiattributiven Präferenzstrukturen zu modellieren.

In der Marktforschung spielt die realistische Abbildung solcher Präferenzen z.B. im Rahmen von Produktkonzepttests eine wichtige Rolle. Vor der Einführung der Conjoint Analyse in diesen Bereich erhob man Präferenzen meist durch direkte Fragen. Beispielsweise sollten Kunden auf Ratingskalen einstufen, wie wichtig ihnen eine positive Ausprägung eines Objekts auf einem Attribut ist. Man fragt also: Wie wichtig ist ihnen eine große Wohnfläche? Eine hohe Zimmeranzahl? Ein niedriger Mietpreis? ... - Durch diese Systematisierung liegt die gewonnene Information gegenüber der einfachen Befragung strukturiert vor.

Ein großer Nachteil bleibt jedoch bestehen. Typischerweise erhält man bei diesem Vorgehen das Ergebnis, daß alle Objektattribute sehr wichtig sind. Schlimmer noch - es gibt in aller Regel gar keine Objekte, die den Anforderungen der Befragten genügen. Jeder möchte verständlicherweise gerne billige, große Wohnungen in guter Lage haben.

Die Conjoint Analyse überwindet die Nachteile solcher einfacher Befragungsverfahren. Die Conjoint Analyse ist eine Verfahrensklasse, mit der die Teilnutzen der verschiedenen Attributausprägungen eines Objekts effektiv geschätzt werden können. Diese Verfahren umfassen dabei 1) die Sammlung von Informationen über Präferenzurteile durch die effiziente Gestaltung eines Interviews und 2) die Schätzung der Parameter der algebraisch formulierten Präferenzstrukturmodelle.

Das Ziel dieser Arbeit besteht darin, einen Überblick über die verschiedenen Spielarten der Conjoint Analyse zu geben. Verschiedene Vorgehensmöglichkeiten sollen vor dem Hintergrund aktueller Forschungsergebnisse diskutiert werden. Dabei werden wir auf einige Probleme und Forschungsfragen stoßen, die bisher noch ungelöst sind. Eine kleine Teilmenge dieser Probleme wird im zweiten Teil der Arbeit mit empirischen Methoden bearbeitet und einer Lösung nähergebracht.

Diese Arbeit nimmt eine marketingorientierte Perspektive ein. Dies zeigt sich offen an den Beispielen, die zur Illustration eingesetzt werden. Weniger offen, und dies wird nur der Fachmann erkennen, ist dies daran ablesbar, daß wir die Conjoint Analyse weitgehend so verstehen, wie sie sich in der Marketingliteratur präsentiert.

Diese Arbeit ist in fünf inhaltliche Abschnitte gegliedert.

Abschnitt 2: Er dient dazu, das grundlegende Konzept der Conjoint Analyse anhand eines Beispiels vorzustellen. Dabei werden die wichtigsten Begriffe eingeführt und die algebraische Notation festgelegt, die in dieser Arbeit Verwendung finden.

Das Kernproblem der Präferenzmodellierung besteht in der Bestimmung des *Nutzens* von Handlungskonsequenzen. Das Prinzip verschiedener Verfahren zur Bestimmung dieser theoretischen Größe wird vorgestellt.

Abschnitt 3: Er bespricht ausführlich die Verfahrensmöglichkeiten einer conjointanalytischen Untersuchung aus Sicht des Untersuchers. Da dieser Vorgang

komplex ist, wird er zur besseren Übersichtlichkeit in Schritte eingeteilt. Zu jedem dieser Schritte stellen wir mögliche Vorgehensweisen vor und besprechen diese vor dem Hintergrund verfügbarer Forschungsbefunde. Wir haben uns dazu entschlossen, Abschnitt 3 recht umfangreich zu gestalten. Wir sind der Meinung, daß sich dieser Aufwand gelohnt hat, da derzeit keine Monographie verfügbar ist, welche die Conjoint Analyse in einer vergleichbaren Form bespricht.

Abschnitt 4: In aller Regel ist eine conjointanalytische Untersuchung mit der Erstellung individueller Präferenzstrukturmodelle nicht zu Ende. Diese primären Ergebnisse verarbeitet man meist weiter, indem Personengruppen segmentiert werden (z.B. Bildung von Käufersegmenten) oder indem man Präferenzstrukturen über mehrere Personen aggregiert. Eine besonders attraktive Form der Aggregation sind Wahlanteil- bzw. Marktanteilsimulationen. Diese und weitere Möglichkeiten der Weiterverarbeitung conjointanalytischer Ergebnisse enthält Abschnitt 4.

Abschnitt 5: Wie bei allen Meßverfahren, ist auch bei der Conjoint Analyse von entscheidendem Interesse, welche Meßgüte die Ergebnisse des Verfahrens aufweisen. Abschnitt 5 beschäftigt sich ausführlich mit der Beurteilung der Meßgüte von conjointanalytischen Ergebnissen. Dieser Abschnitt verfolgt zwei Ziele. Er dient einerseits dazu, die Forschungsbefunde, die in Abschnitt 3 dargestellt wurden, besser und kritischer zu verstehen. Gewissermaßen ist die Kenntnis der grundlegenden Inhalte des Abschnitts 5 (und auch Teile des Abschnitts 4) die Voraussetzung dafür, um die kritische Würdigung der Verfahrensmöglichkeiten in Abschnitt 3 verstehen zu können. Zweitens wird in Abschnitt 5 das Ziel verfolgt, einige Probleme üblich gewordener Meßgütekriterien aufzuzeigen. Die Kenntnis problematischer Meßgütekriterien und die konstruktive Lösung zu ihrer Überwindung fließen direkt in ein spezielles Forschungsdesign ein, nach dem die eigenen Untersuchungen aufgebaut sind.

Abschnitt 6: Diesem Abschnitt ist die Darstellung und Diskussion unserer eigenen Untersuchungen vorbehalten. Wir versuchen zunächst, Charakteristika eines conjointanalytischen Verfahrens herauszuarbeiten, damit dieses als derzeit optimales Verfahren gelten kann. Die Grundlage hierfür bildet die detaillierte Besprechung der Verfahrensmöglichkeiten in Abschnitt 3. Ziel unserer Bemühungen ist die weitere Optimierung eines solchen Verfahrens, indem wir

eine offene Forschungsfrage bearbeiten, die ein derzeit optimales Verfahren aufwirft.

Durch die Bearbeitung einer aktuellen Forschungsfrage hoffen wir einen kleinen Teil zur wissenschaftlichen Erkenntnis beizutragen. Dies ist jedoch nur ein Ziel des Abschnitts 6. Damit gleichberechtigt ist die exemplarische Vorstellung eines speziellen Forschungsdesigns. Dabei handelt es sich um eine besondere Art und Weise, eine beliebige Fragestellung zur Optimierung der Conjoint Analyse effektiv zu bearbeiten.

2 Konzept der Conjoint Analyse

In diesem Abschnitt soll das Grundkonzept der *Conjoint Analyse* (CA) und die typischen Ergebnisse einer CA-Untersuchung anhand eines fiktiven Beispiels dargestellt werden. Danach werden die Grundbegriffe der multiattributiven Präferenzstrukturmodellierung definiert und die verwendete algebraische Notation dargestellt.

2.1 Prinzip und Ergebnisse einer typischen CA-Untersuchung

Das Ziel einer Untersuchung nach dem Prinzip der CA besteht in der Vorhersage menschlichen Beurteilungs- und Entscheidungsverhaltens. Im Marketing handelt es sich dabei um das Verhalten von z.B. Konsumenten, Absatzmittlern oder Zulieferern. Die häufigste Anwendung der CA ist der Produkt-(Dienstleistungs-)konzepttest, die Preisfindung und die Marktsegmentierung. Diese Aufgaben sind grundlegend bei der Innovation und Modifikation von Produkten und Dienstleistungen und bei der Planung von zielgerichteten Kommunikationskampagnen.

Die Grundannahme der CA besteht darin, daß Produkte und Dienstleistungen (allgemein: Objekte) sich als Bündel von nutzenstiftenden Attributen auffassen lassen. Eine Flugreise hat z.B. die Attribute 'Preis', 'Verkehrsverbindung zum Abflughafen', 'Service an Bord' und 'Umbuchungsflexibilität'. Jede konkrete Flugreise läßt sich auf diesen Attributen charakterisieren. Die CA liefert als grundlegendes Ergebnis *Teilnutzenwerte* - numerische Werte, die den subjektiv wahrgenommenen Teilnutzen einer bestimmten Attributausprägung abbilden. Beispielsweise könnten die Teilnutzenwerte eines bestimmten Kunden wie in Tabelle 1 dargestellt aussehen.

Welche Information läßt sich aus den Ergebnissen entnehmen? Der Gesamtnutzen einer jeden beliebigen Flugreise läßt sich durch einfache Addition errechnen. Dadurch läßt sich bei verschiedenen Flugangeboten vorhersagen, für welche Alternative sich die Person entscheiden würde.

Die Teilnutzenwerte verschiedener Attribute sind direkt vergleichbar. Ein kostenloser Zubringerservice bedeutet für die Person ein Gewinn von 10 Teilnutzenwerten.

Tabelle 2.1: Die Attribute, deren Ausprägung und die Teilnutzenwerte eines fiktiven Kunden.

Attribut und Attributausprägung	Teilnutzen
Preis	
20% unter Durchschnitt	70
10% unter Durchschnitt	60
im Durchschnitt	50
10% unter Durchschnitt	30
20% über Durchschnitt	0
Verkehrsverbindung	
kostenloser Zubringerservice	40
Bahnanschluß	30
keine spezielle Verbindung	0
Service an Bord	
Essen und Getränke kostenfrei	10
nur gegen Aufpreis	0
Umbuchungsflexibilität	
Umbuchung jederzeit möglich	50
Umbuchung 14 Tage vorher	10
keine Umbuchungsmöglichkeit	0

Diese kompensieren einen Flug mit kostenpflichtigem Bordservice oder einen bestimmten Mehrpreis. Die Höhe dieses zusätzlichen Flugpreises, die der Kunde für den Zubringerservice zu bezahlen bereit ist, hängt vom tatsächlichen Preisniveau ab, weil die Preiswahrnehmung nicht linear verläuft.

Bei Marktforschungsuntersuchungen nach dem Prinzip der CA erhält man die Wertvorstellungen einer jeden befragten Person individuell. Personen mit einer ähnlichen Nutzenstruktur lassen sich so einfach in homogene Marktsegmente einteilen.

Eine der faszinierendsten Möglichkeiten dieses Untersuchungstyps ist die Marktanteilsimulation. Verfügt man über die Teilnutzenwerte einer repräsentativen Stichprobe, dann können verschiedene Flugangebote dem Simulator vorgegeben werden. Dieser bestimmt die Wahrscheinlichkeit, mit der sich jede einzelne Person für die Alternativen entscheiden würde. Aggregiert man diese Vorhersagen über alle Personen, erhält man Marktanteilsvorhersagen. Auf diese Weise lassen sich einfach „Was wäre wenn ...?“ Szenarien durchspielen. Typische Fragen an den Simulator sind z.B.: „Wieviel Marktanteil kann ein geplantes

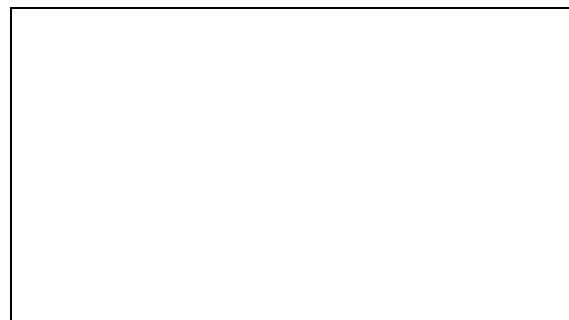


Abbildung 2.1: Eine fiktive Vorhersage des Marktanteils.

Flugangebot auf sich vereinen?“ „Mit wieviel Marktanteilszuwachs ist zu rechnen, wenn der Preis eines Fluges um X gesenkt wird?“ „Was geschieht, wenn ein Konkurrent Flüge mit diesen und jenen Attributen anbietet?“

Damit ist die CA nicht nur eines der flexibelsten Forschungsinstrumente, es kann auch Managemententscheidungen wie kaum ein anderes Forschungsinstrument sinnvoll unterstützen.

Wie bei jedem anderen Meßinstrument stellt sich auch bei der CA die Frage nach der Validität der Ergebnisse. Eine Möglichkeit der Validitätsüberprüfung besteht darin, die vorhergesagten Wahlanteile mit tatsächlichen Marktanteilen zu vergleichen. In diesem Falle spricht man auch von der *kriterienorientierten Validität auf aggregierter Ebene*. Die genaue Vorgehensweise ist in Abschnitt 5.4.2 beschrieben. Da die CA versucht, die Präferenzstruktur einer jeden Person individuell abzubilden, kann die *kriterienorientierte Validität* auch *auf der individuellen Ebene* bestimmt werden. Hier stellt sich die Frage, inwieweit ein individuelles Modell der Präferenzstruktur in der Lage ist, tatsächliche Beurteilungen und Entscheidungen einer Person vorherzusagen. Häufig werden hierzu den Personen sogenannte 'Holdout-Objekte' vorgelegt. Hierbei würde es sich in diesem Beispiel um mehrere Flugangebote handeln. Die Person erhält die Aufgabe, diese Angebote gemäß ihren Präferenzen zu ordnen. Die so entstandene Rangreihe wird mit der Rangreihe verglichen, die aus den Teilnutzenwerten der jeweiligen Person vorhergesagt würden. Der Grad der Übereinstimmung dient als Hinweis über die Validität der CA-Ergebnisse. Die Verfahrensmöglichkeiten der individuellen Validitätsbestimmung sind ausführlich in Abschnitt 5.4.1 dargestellt.

Wie aus dem Beispiel deutlich wird, dient die CA meist zur Lösung konkreter Probleme im Anwendungsbereich. Insbesondere wird sie zur Vorhersage von Konsumentenbewertungen im Rahmen produkt-, preis- und kommunikationspolitischer Entscheidungen eingesetzt. Dahinter steht die Annahme, daß die Bewertung von Objekten (z.B. Produkte, Dienstleistungen, Unternehmen, ...) eine Funktion der Bewertung der Attribute dieser Objekte ist. Dieses Attributbewertung wird als *Nutzen* bezeichnet.

2.2 Multiattributive Präferenzstrukturmodellierung: Grundbegriffe

Entsprechend üblicher Konventionen, die sich im Zusammenhang mit der Modellierung multiattributiver Präferenzstrukturen und der Literatur zur CA gebildet haben, werden im folgenden einige grundlegende Begriffe und die verwendete Schreibweise der algebraischen Ausdrücke definiert.

2.2.1 Zentrale Begriffe

Das Objekt. Ein Objekt ist ein komplexer Beurteilungs- und Entscheidungsgegenstand. Es kann sich dabei um eine Handlungsalternative, ein Handlungskonsequenzenbündel, ein Produkt, eine Dienstleistung, eine Person, eine Personengruppe, ein Unternehmen, eine Werbeanzeige, eine Verordnung etc. handeln.

Das Attribut. Ein Attribut ist eine entscheidungsrelevante Beurteilungsdimension eines Objekts. Diese Arbeit beschäftigt sich mit Objekten, die auf mehreren entscheidungsrelevanten Attributen spezifiziert sind. Entscheidungsrelevante Attribute sind solche, die bei Vorliegen mehrerer Objekte aus einer Objektklasse tatsächlich zu Präferenzunterschieden zwischen den Objekten führen. Nehmen wir als Beispiel eine Person, die in einem Schreibwarengeschäft einen Kugelschreiber (Objekt) kaufen will und diesen unter mehreren ähnlichen Objekten aussuchen soll. Entscheidungsrelevante Attribute der Kugelschreiber könnten z.B. der Preis, die Farbe, die Form, die Schreibqualität etc. sein. Diese Attribute gehen in die Präferenzbildung der Person ein. Ein anderes potentiell Attribut, z.B. die Haltbarkeit des Kugelschreibers oder die chemische Zusammensetzung seiner Hülle sind zwar Eigenschaften des Objekts, sie gehen jedoch nicht in die Präferenzbildung des Entscheiders ein, sind also keine entscheidungsrelevanten Bewertungsdimensionen, da entweder keine Unterschiede zwischen den Objekten bekannt sind (Haltbarkeit) oder nicht wahrgenommen bzw. beachtet werden (chemische Zusammensetzung). Wenn von Attributen die Rede ist, sind normalerweise nur entscheidungsrelevante Attribute gemeint.

Die Attribute sind einerseits objektive entscheidungsrelevante Objekteigenschaften. Andererseits können sie auch als mentale Repräsentationen eines Beurteilers oder Entscheiders aufgefaßt werden. Bei der Präferenzbildung werden diese Repräsentationen der objektiven Eigenschaften evoziert. Diese Auf-

fassung berücksichtigt den Umstand, daß Attribute nicht in sich bedeutungshaltig sind, sondern Interpretationen unterliegen. Das Attribut 'Farbe' des Kugelschreibers kann verschieden in Beurteilungsprozesse einfließen. Soll der Kugelschreiber als Geschenk Verwendung finden, so wird die Farbe möglicherweise unter ästhetischen Gesichtspunkten betrachtet. Die Auffälligkeit und Signalwirkung kann im Vordergrund stehen, wenn er als Werbemedium dienen soll. Die Attributdefinition kann sich an den objektiven Eigenschaften orientieren oder an der Eigenschaftsrepräsentation. In der Literatur finden sich eine Reihe von Klassifikationssystemen, die versuchen, Ordnung in verschiedene Möglichkeiten der Repräsentation von Attributen zu bringen (s. Abschnitt 3.2.5).

Die Attributstufen (levels). Dies sind die möglichen Ausprägungen eines Objekts auf einem Attribut. Ist z.B. das Attribut der Preis eines Objekts, dann ist die Attributstufe ein bestimmter Preis, also z.B. 10 DM oder 15 DM.

Kann man die Stufen auf mindestens Intervallskalenniveau angeben, dann spricht man auch von einem quantitativen Attribut oder metrischen Attribut. Ein quantitatives Attribut, wie z.B. der Preis, hat die Eigenschaft, daß die Attributstufen prinzipiell an beliebiger Stelle der Attributdimension liegen können. Es ist demnach natürlich auch vorstellbar, daß der Preis eines Objekts nicht nur 10 DM oder 15 DM beträgt, sondern auch 10,46 DM.

Anders verhält sich dies mit sogenannten qualitativen Attributen. Die Attributstufen bei qualitativen Attributen sind nominalskaliert. Das Attribut 'Farbdruckfähigkeit' ist ein Beispiel dafür. Entweder besitzt ein Drucker diese (Farbdruckfähigkeit vorhanden) oder nicht (Farbdruckfähigkeit nicht vorhanden). Unter Umständen muß zwischen verschiedenen Ausprägungen der Farbdruckfähigkeit unterschieden werden. Dies bedeutet jedoch lediglich, daß das Attribut auf mehreren Stufen definiert ist (kein Farbdruck, 16 Farben, 256 Farben).

Die Nutzenfunktionen quantitativer Attribute sind in einem bestimmten Definitionsbereich stetig. Qualitative Attribute dagegen sind abschnittsweise definiert und besitzen demnach nur diskrete Nutzenwerte.

Die Objektklasse. Der Begriff der Objektklasse ist mehr oder weniger hilfreich. Wir meinen damit eine Menge von Objekten, die auf den gleichen Attributen beschreibbar sind. Beispielsweise wäre die Objektklasse der Autos vorstellbar. Alle Objekte dieser Menge sind alle auf den Attributen Verarbeitungsqualität, Kaufpreis, Höchstgeschwindigkeit, Motorleistung, Benzinverbrauch, Unterhaltskosten, Anzahl der Sitze usw. beschreibbar. Ein Drucker dagegen ist

nicht auf allen diesen Attributen beschreibbar. Er gehört nicht zu dieser Objektklasse.

Die Präferenz und der Nutzen. Die Präferenz ist das Ergebnis eines kognitiven Beurteilungsprozesses bei der Beurteilung von Objekten auf einem oder mehreren Attributen. Präferenzen sind Entitäten des empirischen Relativs und damit prinzipiell beobachtbar. Werden Präferenzurteile über mehrere Objekte gebildet, so sollten sich diese mindestens auf einer Ordinalskala abbilden lassen.

Bei der multiattributiven Präferenzstrukturmodellierung wird angenommen, daß die Präferenz das Ergebnis eines Informationsintegrationsprozesses ist. Bei diesem Prozeß spielen die wahrgenommenen Ausprägungen der Objekte auf den Attributen eine entscheidende Rolle. Die Präferenz eines Objekts wird nach bestimmten Gesetzmäßigkeiten gebildet. Um die Art und Weise der Präferenzbildung modellieren zu können, bedient man sich des Konstrukts des Nutzens. Man versteht den Nutzen als Funktion der Ausprägungen eines Objekts auf den Attributen. Letztendlich kommt bei dieser theoretischen Sichtweise die Präferenz für ein Objekt durch den Nutzen der Objektattribute zustande. Die Präferenzstrukturmodellierung macht sich zur Aufgabe, die Beziehung zwischen den Ausprägungen eines Objekts auf den Attributen und seiner Präferenz herzustellen. Hierbei geht man über den Umweg des Nutzens als erklärendes Konzept.

Die Nutzenfunktion bzw. das Präferenzmodell. Die Nutzenfunktion ist ein algebraisches Modell, das den Zusammenhang zwischen objektiven Objekteigenschaften und subjektivem Nutzen auf einem Attribut beschreibt. Die Nutzenfunktion wird als erklärendes Modell der Präferenz herangezogen. Wir gebrauchen deswegen den Begriff des Präferenzmodells synonym. Es gibt verschiedene Grundtypen von Nutzenfunktionen. Die gängigsten sind das Teilnutzenwertmodell, das Idealvektormodell und das Idealpunktmodell. Diese Funktionstypen werden ausführlich in Abschnitt 3.4.1 besprochen.

Die Nutzenstrukturfunktion bzw. das Präferenzstrukturmodell. Auch diese beiden Begriffe werden synonym gebraucht. Bei der multiattributiven Modellierung von Präferenzen stellt sich immer die Frage, wie der Nutzen der verschiedenen Attribute zu einem Gesamtnutzen kombiniert wird. Um diesen informationsintegrierenden Vorgang modellieren zu können, bedient man sich ebenfalls algebraischer Modelle.

Das Präferenzstrukturmodell ist somit eine Abbildung der Informationsverarbeitungsergebnisse in algebraischer Form. Parameter im Modell sind Nutzenwerte, also skalierte Bewertungen der Attributausprägungen von Objekten. Dabei wird das Modell so gebildet und die Modellparameter so geschätzt, daß die Funktionswerte der Objekte (Objektgesamtnutzen) die empirischen Objektpräferenzen möglichst konsistent abbilden. In dieser Arbeit werden wir uns vornehmlich mit linear-additiven kompensatorischen Präferenzstrukturmodellen befassen.

2.2.2 Notation

In diesem Abschnitt werden die verwendeten Abkürzungen der algebraischen Ausdrücke beschrieben. Sie werden im weiteren Verlauf konsistent gebraucht.

2.2.2.1 Grundbausteine der Präferenzstrukturmodelle

h	Auskunftsperson (A_p) $h. h = 1 \dots H.$
X_i	Objekt $i. i = 1 \dots I.$
x_j	Attribut $j. j = 1 \dots J.$
x_{jk}	Attributstufe k auf dem Attribut $j. k = 1 \dots K_j.$
$u(X_i)$	Nutzen (utility) des Objekts $X_i.$
$u(x_{jk})$	Teilnutzen der Attributstufe $x_{jk}.$

2.2.2.2 Objektdefinition

Ein Objekt X_i ist durch seine Ausprägungen auf J Attributen definiert. Es ist formal beschreibbar durch das n -Tupel $X_i = (x_{1k}, x_{2k}, \dots, x_{jk}, \dots, x_{Jk})$.

2.2.2.3 Das Präferenzintegrationsmodell

Der Gesamtnutzen eines Objekts ist eine Funktion des durch die Attribute gestifteten Teilnutzens eines Objekts. Formal läßt sich dies folgendermaßen ausdrücken:

$$u(X_i) = f(u(x_1), u(x_2), \dots, u(x_J)) \quad (2.1)$$

Gilt die Annahme, daß die Attribute des Objekts voneinander präferenzunabhängig sind, d.h. gilt für jedes Attribut, daß die Teilpräferenz für die Attributaus-

prägung eines Attributs unabhängig von der Ausprägung des Objekts auf einem beliebigen anderen Attribut ist, dann kann ein kompensatorisches additives Präferenzstrukturmodell geeignet sein, die Präferenzstrukturen angemessen zu repräsentieren. Es ist geeignet, wenn zusätzlich gilt, daß es sich um risikolose Beurteilungen handelt, die Attributstufen nutzenskaliert sind und alle Attributstufen auf dieselbe Nutzenskala abgebildet werden können.

Der Gesamtnutzen eines Objekts X_i ergibt sich aus der Summe seiner Teilnutzen.

$$\begin{aligned} u(X_i) &= \sum u(x_j) \\ &= u(x_1) + u(x_2) + \dots + u(x_J) \end{aligned} \quad (2.2)$$

Dieses Modell (2.2) gibt das *Präferenzintegrationsmodell* der Objektklasse wieder, der das Objekt X_i angehört.

2.2.2.4 Das Präferenzmodell

Der Teilnutzen eines Attributs $u(x_j)$ ist wiederum eine Funktion der Ausprägung des Attributs:

$$u(x_j) = f(x_{jk}). \quad (2.3)$$

Die Funktion in Gleichung (2.3) ist das *Präferenzmodell* des Attributs x_j . Dieses Präferenzmodell kann für jedes Attribut ein anderes sein. Eine ausführliche Diskussion von Präferenzmodellen soll hier nicht angestrebt werden (s. Abschnitt 3.4.1). Ein mögliches und sehr allgemeines Präferenzmodell ist das Teilnutzenwertmodell. Es soll an dieser Stelle jedoch nur exemplarisch dargestellt werden:

$$u(x_j) = \sum b_{jk} * x_{ijk} \quad (2.4)$$

mit

$u(x_j)$ Teilnutzen des Attributs x_j .

b_{jk} Teilnutzenwert der Attributstufe k des Attributs x_j .

$x_{ijk} = \begin{cases} 1 & \text{falls ein Objekt } X_i \text{ die Attributausprägung } k \text{ auf Attribut } x_j \text{ besitzt} \\ 0 & \text{in anderen Fällen.} \end{cases}$

2.2.2.5 Das Präferenzstrukturmodell

Aus den Gleichungen (2.2) und (2.4) folgt ein additives kompensatorisches *Präferenzstrukturmodell (PSM)* der Form

$$u(X_i) = \sum_j b_{jk} \cdot x_{ijk} \quad (2.5)$$

mit

$u(X_i)$ (Gesamt-)Nutzen des Objekts X_i .

b_{jk} Teilnutzenwert der Attributstufe k des Attributs x_j .

$x_{ijk} = \begin{cases} 1 & \text{falls Objekt } X_i \text{ die Attributausprägung } k \text{ auf Attribut } x_j \text{ besitzt} \\ 0 & \text{in anderen Fällen.} \end{cases}$

Das Präferenzstrukturmodell (PSM) faßt das Präferenzmodell und das Präferenzintegrationsmodell zusammen und bestimmt den Gesamtnutzen eines Objekts X_i . Bei diesem additiven PSM berechnet sich der Gesamtnutzen eines aus den Attributen definierbaren Objekts durch einfache Addition der entsprechenden Teilnutzenwerte. Der nächste Abschnitt beschäftigt sich mit den grundlegenden Verfahrenstypen, diese Teilnutzenwerte zu bestimmen.

2.3 Nutzenschätzung bei additiven Modellen

Liegt ein Set von Attributen und deren Stufen fest und wird ein additives Präferenzstrukturmodell angenommen, dann lassen sich die Teilnutzenwerte des Modells prinzipiell über zwei verschiedene Verfahrensklassen bestimmen: Die Nutzenparameterschätzung kann über direkte Verfahren oder durch indirekte Verfahren erfolgen.

2.3.1 Grundüberlegung

Direkte Verfahren verlangen von den Befragten eine realistische Einschätzung darüber, wie ihre Urteile und Entscheidungen zustande kommen - die Befragten betreiben also eine Art Introspektion oder Metakognition. Direkte Verfahren sind oft so konzipiert, daß sie die Beurteilungsaufgabe in mehrere Teilaufgaben zerlegen. Dadurch soll eine realistische Beurteilung unterstützt werden. Die Ver-

fahren verlangen von den Entscheidern eine direkte Zuordnung von numerischen Werten zu den Attributen und Attributstufen bzw. den Konsequenzen die mit diesen verbunden sind. Diese Klasse von Verfahren nennt man auch *kompositorische Modellbildung oder explizite Modellbildung*, weil der Nutzen von Objekten aus den explizit eingeschätzten Teilnutzen ihrer Attribute komponiert werden.

Von den direkten Verfahren sind die *indirekten Verfahren* der Nutzenparameterschätzung abgrenzbar. Diesen Verfahren ist gemeinsam, daß Personen mehr oder weniger komplexe Entscheidungsobjekte zur Beurteilung vorgelegt werden. Die Objekte bestehen aus mindestens zwei Attributen. Die Personen beurteilen die Objekte entsprechend ihrer Präferenz. Aus diesen Gesamtpräferenzurteilen werden dann die Teilnutzenwerte der einzelnen Attribute dekomponiert. Man nennt diese Vorgehensweise deshalb auch *dekompositorische Modellbildung*. Die CA ist eine solche Methode.

Bei den Verfahren der kompositorischen Modellbildung erfragt man von den Apn Nutzenwerte, die intervallskaliert und über die Attribute vergleichbar sein sollten, direkt. Es wurden eine Reihe von Vorgehensweisen vorgeschlagen, die diese schwierige Aufgabe für die Befragten erleichtern sollten (z.B. Edwards, 1977; Hoepfl & Huber, 1970; Huber, 1974a; Huber, Daneshgar & Ford, 1971; Srinivasan, 1988; Wilkie & Pessemier, 1973). Näher beschrieben wird an dieser Stelle exemplarisch das SMART-Verfahren (Edwards, 1977). In Abschnitt 3.7.1 werden weitere kompositorische Verfahren dargestellt und diskutiert.

2.3.2 SMART als Beispiel für ein direktes Nutzenschätzverfahren

Der Methode SMART (*s*imple *m*ultiattribute *r*ating *t*echnique) von Edwards (1977) liegt die Annahme eines linearen Vektorpräferenzmodells und eines additiven Präferenzintegrationsmodells zugrunde (s. Abschnitt 3.4). Das lineare Vektormodell ist ein Spezialfall des Teilnutzenwertmodells (Gleichung 2.5). Es geht von quantitativen Attributen aus. Zu bestimmen ist bei diesem Modell lediglich die relative Gewichtung der Attribute. Veranschaulicht man die Nutzenfunktionen graphisch, dann ist dieses relative Gewicht die Steigung des Graphen der linearen Nutzenfunktion bei jedem Attribut. Das PSM hat folgende Form:

$$u(X_i) = \sum w_j * x_{ij} \quad (2.6)$$

Mit w_j als relatives Gewicht des Attributs j , und x_{ij} als die Ausprägung des Objekts X_i auf Attribut j , kann mittels Gleichung (2.6) der Gesamtnutzen des Objekts $u(X_i)$ bestimmt werden.

Die relative Wichtigkeit eines Attributs ist von seiner potentiellen minimalen und maximalen Ausprägung nicht unabhängig. Bittet man z.B. einen Beurteiler anzugeben, was ihm wichtiger erscheint: Die Entfernung eines Urlaubshotels zum Strand oder der Preis des Hotelzimmers, so könnte dieser zunächst zu der Einschätzung kommen, daß die geringe Entfernung ihm wichtiger ist. Teilt man ihm dann aber mit, daß am Urlaubsort seiner Wahl alle Hotels eine maximale Entfernung von 300 Metern zum Strand aufweisen (das Attribut 'Entfernung' hat die niedrigste Stufe 20m und die höchste Stufe 300m), und daß die Preise zwischen 20 DM und 500 DM schwanken (Attribut 'Preis': geringste Stufe 20 DM, höchste Stufe 500 DM), dann könnte er schnell zu einer anderen Einschätzung seiner Attributgewichtung gelangen.

Um die Abhängigkeit der relativen Attributwichtigkeit von den Attributstufen zu berücksichtigen, definiert Edwards die relative Attributwichtigkeit als Nutzen-differenz. Diese Differenz erhält man, wenn innerhalb eines Attributs der Nutzen der am niedrigsten präferierten Attributstufe vom Nutzen der am höchsten präferierten Attributstufe abgezogen wird.

Um die Beurteilungsaufgabe durch Teilschrittbildung zu erleichtern, werden die Attribute entsprechend ihrer Wichtigkeit zunächst in eine Rangordnung gebracht. Dann wird dem unwichtigsten Attribut willkürlich eine relative Wichtigkeit von 10 zugeordnet. Alle anderen relativen Wichtigkeiten müssen nun von den Beurteilern im Vergleich hierzu eingeschätzt werden. Die zugeordneten Zahlen sollten dabei die Verhältnisse der relativen Wichtigkeiten repräsentieren.

Anschließend werden die relativen Wichtigkeiten so normiert, daß ihre Summe 1 ergibt. Soll der Nutzen eines Objekts bestimmt werden, so bedarf dies der Spezifikation des Objekts auf allen Attributen. Hierzu wird eine Skala von 0 bis 100 verwendet. Die am wenigsten präferierte Attributstufe erhält den Wert 0, und der am höchsten präferierte Stufe wird der Wert 100 zugeordnet. Gemäß Gleichung (2.6) kann der Nutzen eines jeden Objekts berechnet werden.

2.3.3 Ein indirektes Nutzenschätzverfahren: Das Conjoint Measurement

Anders als die direkten Verfahren der kompositorischen Präferenzstrukturmodellierung gehen indirekte Verfahren vor. Mit Conjoint Measurement (CM) wird eine Klasse von Verfahren der dekompositorischen Präferenzstrukturmodellierung bezeichnet. Das axiomatisch fundierte CM wurde erstmals ausführlich von Debreu (1960) und Luce und Tukey (1964) im Bereich der mathematischen Psychologie dargelegt. Der Ansatz verfolgt das Ziel, auf der Grundlage ordinalskalierte Beurteilungsdaten multiattributiver Objekte (abhängige Variable) simultan den Nutzen jeder unabhängigen Variable (Attribute) so zu skalieren, daß die empirischen Relationen der Beurteilungsdaten durch das PSM bestmöglich reproduziert werden können. Die Beurteilungen werden beim CM anhand einer Menge von Objekten gewonnen, deren Attributausprägungen systematisch kombiniert werden. In der Regel handelt es sich um ein vollfaktorielles Design.

Das CM hat zum Ziel, PSM zu schätzen und deren Angemessenheit zu überprüfen. Hierzu benötigt man ordinalskalierte Präferenzurteile über Objekte, die auf allen Attributen spezifiziert sind. Diese können z.B. dadurch entstehen, daß Auskunftspersonen (Apn) multiattributive Objekte entsprechend ihrer Präferenz in eine Rangreihe bringen. Werden die Beurteilungsdaten mit Hilfe eines linear-additiven Präferenzintegrationsmodells abgebildet, spricht man auch vom additiven CM. Die Grundlagen des additiven CM und das polynomiale Conjoint Measurement werden z.B. bei Krantz, Luce, Suppes und Tversky (1971) und Tversky (1967a) ausführlich besprochen.

Das additive CM hat u.a. den Vorteil, daß die Menge der Beurteilungsobjekte relativ klein gehalten werden kann. Zur Schätzung der Modellparameter reichen beim additiven CM reduzierte orthogonale Designs (Addelman, 1962a; Green, 1974; Green, Carroll & Carmone, 1978) aus.

Die Besonderheit beim CM besteht darin, daß die Beurteilungen der Apn lediglich auf Ordinalskalenniveau oder sogar auf Kategorialskalenniveau (Carroll, 1969) abzugeben sind. Sind die Annahmen des CM haltbar, dann können die ordinalskalierten Präferenzdaten durch die Teilnutzenwerte der Objektattribute reproduziert werden. Die Teilnutzenwerte und der Gesamtnutzen der Objekte sind approximativ intervallskaliert.

Die Teilnutzenwerte sind Parameter des Präferenzstrukturmodells. Sie werden für jede Auskunftsperson individuell geschätzt und besitzen über alle

Attribute dieselbe Skaleneinheit. Die Differenzen der Teilnutzenwerte verschiedener Attributstufen sind damit direkt miteinander vergleichbar.

Die Ausgangssituation einer Präferenzstrukturanalyse nach dem CM-Ansatz kann folgendermaßen beschrieben werden (vgl. auch Thomas, 1979):

Gegeben: Mindestens kategorial skalierte, i.d.R. ordinalskalierte Werte der Gesamtpräferenz für bestimmte Beurteilungsobjekte. Diese Gesamtpräferenzwerte werden als das Ergebnis der verbundenen (joint) Wirkungen der Objektattribute interpretiert.

Gegeben: Eine Objektmenge, deren Elemente durch Merkmalskombinationen konstruiert wurden. Dieses faktorielle Design kann vollständig oder orthogonal reduziert sein.

Gesucht: Eine Verknüpfungsregel bzw. ein Präferenzstrukturmodell, das geeignet ist, die kognitive und affektive Verarbeitung der wahrgenommenen Merkmalskombinationen (Objekte) zu repräsentieren.

Gesucht: Intervallskalierte Skalenwerte mit derselben Maßeinheit für den Teilnutzen aller Attribute bzw. deren Stufen und für den Objektgesamtnutzen. Dieser Gesamtnutzen sollte in einer monotonen Beziehung zu den empirischen Gesamtpräferenzen stehen.

Das CM ist ein axiomatisches System. Es formuliert notwendige und hinreichende Bedingungen, die eine empirisch erhobene Ordnung erfüllen muß, damit auf der Basis eines spezifizierten Präferenzstrukturmodells metrische Skalen für den Objektnutzen und die Objektteilnutzen der Attribute existieren und konstruiert werden können. Ansätze der axiomatischen Fundierung des CM sind z.B. bei Debreu (1960), Krantz (1964), Krantz et al. (1971), Luce (1966), Luce und Tukey (1964) zu finden.

2.4 Wo ist die Conjoint Analyse anzusiedeln?

Neben der Modellierung klinischer Urteilsbildungsprozesse fielen die Gedanken des CM insbesondere in der Marketingforschung auf fruchtbaren Boden. Die erste Veröffentlichung, die das CM in die Marketingliteratur einführt, ist unseres Wissens der Aufsatz von Green und Rao (1971) im *Journal of Marketing Research*. Die erste Monographie, die das CM und verwandte Techniken marketingorientiert behandelt, stammt von Green und Wind (1973). Es folgten wei-

tere Aufsätze zu Einsatzmöglichkeiten des CM in diesem Bereich, z.B. von Green, Carmone und Wind (1972), Green und Wind (1975), Johnson (1974), Srinivasan und Shocker (1973c) und Westwood, Lunn und Beazly (1974).

Da die Beurteilung von Produkten durch Konsumenten eine zentrale Frage der Absatzforschung ist, fand die Modellierung dieser Beurteilungsvorgänge einige Beachtung. Flankierend zu diesen Vorgängen ist im Marketing eine Entwicklung des veränderten Produktbegriffs zu erkennen. Das Produkt wurde von Lancaster (1971) nicht mehr als Gut verstanden, sondern als Bündel von Eigenschaften, die ein Anbieter zusammenstellt, um die Wünsche und Bedürfnisse tatsächlicher und potentieller Abnehmer zu befriedigen. Aus Sicht des Anbieters sind diese Eigenschaften Gestaltungselemente der objektiven Beschaffenheit des Produkts. Aus der Sicht des Nachfragers werden diese Elemente durch den Prozeß der subjektiven Wahrnehmung und Bewertung mit verschiedenen Nutzenerwartungen verknüpft. Diese, und nicht die objektiven Produkteigenschaften, bestimmen das Verhalten des Nachfragers.

2.4.1 Conjoint Analyse und Conjoint Measurement

Im Bereich der Konsumentenforschung wurde von der Forschergruppe um Green (Carmone, Green & Jain, 1978; Green & DeSarbo, 1978; Green & Srinivasan, 1978) der Begriff der *Conjoint Analyse* geprägt. Das CM betrachtet vornehmlich die Bedingungen, unter welchen gleichzeitige numerische Repräsentationen der abhängigen und unabhängigen Variablen existieren, wenn gegeben ist, daß die unabhängigen Variablen gemeinsame Effekte auf die abhängige Variable bei einer angenommenen Integrationsregel haben. Typische Ansätze des CM verwenden Techniken der axiomatischen Analyse, um zu überprüfen, ob die notwendigen Bedingungen des CM erfüllt sind. Im ersten Überblicksartikel zur CA schlagen Green und Srinivasan (1978) vor, Ansätze nach dem Grundgedanken des CM, die aber im Gegensatz zum CM nicht die Prüfung der axiomatischen Annahmen, sondern die Qualität der Parameterschätzungen bei der Transformation subjektiver Antworten über multiattributive Beurteilungsobjekte auf multidimensionale Nutzenskalen in das Zentrum ihrer Betrachtung stellen, *conjoint analysis* zu nennen. Dieser Vorschlag wurde zumindest in der Marketingliteratur weitgehend angenommen.

Ansätze der CA stellen den Skalierungsaspekt in den Vordergrund. Es wird genau wie beim CM zunächst ein Modell der Präferenzstruktur formuliert. Die Modellparameter werden dann unter Verwendung verschiedener Algorithmen so geschätzt, daß das Präferenzstrukturmodell möglichst gut den empirischen Präferenzdaten angepaßt ist. Es geht also in erster Linie nicht um die Prüfung von Voraussetzungen der Anwendbarkeit der Methode, sondern um die Maximierung der Anpassungsgüte (goodness of fit) eines Modells an empirische Daten.

2.4.2 Conjoint Analyse und kompositorische Modellierung

Wie das CM verfolgt auch die CA die dekompositorische Strategie der Präferenzstrukturmodellierung. Trotz einer großen Menge von Literatur zur kompositorischen Modellierung von Konsumentenpräferenzen mittels Erwartungs-Wert-Einstellungsmodellen (Fishbein, 1967; Rosenberg, 1956) in den siebziger Jahren, konnte sich die CA als dekompositorisches Verfahren der Präferenzmodellierung behaupten. Obwohl CA-Verfahren nach ihrer Einführung im Marketingbereich häufig in Forschung und Anwendung zum Einsatz kamen, ist doch ein ständiger und kritischer Vergleich zur kompositorischen Modellbildung zu verzeichnen (vgl. Abschnitt 3.7.1). Dieser Vergleich bezieht sich einerseits auf konzeptionelle Fragen, wie z.B. der realitätsgerechten Erklärung von Beurteilungsvorgängen, der die CA eine eher untergeordnete Bedeutung beimißt und dafür der möglichst genauen Vorhersage von Präferenzurteilen die Priorität einräumt. Insbesondere aber betrachten diese vergleichenden Studien die kriterienorientierte Validität der beiden Ansätze. Die CA 'schlägt sich damit zwar auf eigenem Territorium', von einem klaren Gewinn dieses Wettstreits kann aber keineswegs gesprochen werden (vgl. Abschnitt 3.7.1.4).

Seit Mitte der achtziger Jahre ist ein Trend zur Kombination beider Verfahren der Präferenzstrukturmodellierung feststellbar. Oft laufen diese Ansätze weiterhin unter dem Label 'Conjoint Analyse', obwohl die Beurteilungsaufgabe und Parameterschätzung zum Teil nicht auf dem Gedanken der Verbundmessung basieren. Meist wird durch kompositorische (explizite) Modellbildung eine Vorschätzung der Modellparameter vorgenommen. Die nachfolgende 'echte' CA hat dann z.B. die Funktion der Feinschätzung der Parameter. Wir

werden in Abschnitt 3.7.2 ausführlich auf diese Entwicklung und ihre Begründung eingehen.

2.4.3 Conjoint Analyse und Functional Measurement

Ein der CA ähnliches Verfahren zur Analyse beurteilender und bewertender menschlicher Informationsverarbeitungsprozesse ist das Functional Measurement (FM). Es wurde von Anderson (1970) vorgeschlagen und kann als methodisches Pendant zur Informationsintegrationstheorie (Anderson, 1981, 1982) angesehen werden. Das FM ist ein dekompositorischer Ansatz der Modellierung multiattributiver Wahrnehmungsprozesse und weist damit viele Parallelen zur CA auf. Unterschiede bestehen dahingehend, daß das FM eine intervallskalierte abhängige Variable (z.B. Ratingaufgabe) annimmt und voraussetzt. Entsprechend kommen parametrische Schätzverfahren (z.B. ANOVA) zum Einsatz. Die Menge der Beurteilungsobjekte stellen meist ein vollfaktorielles Design dar. Wie die CA bemüht sich auch das FM um eine möglichst angepaßte Schätzung der Modellparameter. Daneben verfolgt es jedoch das Ziel, angenommene Informationsintegrationsmodelle zu testen. Eine typische Fragestellung des FM kann z.B. darin bestehen, ob zwei Attribute miteinander in Wechselwirkung stehen, oder ob sie sich lediglich additiv überlagern. Während die CA solche Annahmen meist unüberprüft trifft und sich dabei auf die oft gefundene Angemessenheit und Robustheit linear-additiver Integrationsmodelle verläßt, steht beim FM gerade die Überprüfung des Informationsintegrationsmodells, also auch erklärende Aspekte, im Zentrum des Bemühens.

3 Prozeß der Conjoint Analyse

In Abschnitt 2 wurde das Grundprinzip der multiattributiven Präferenzstrukturmodellierung durch die CA dargestellt. Abschnitt 3 widmet sich der ausführlichen Darstellung des Ablaufs einer CA-Untersuchung. Hierzu wird der Ablauf einer CA in Schritte eingeteilt. Diese Art der Ordnung des Forschungsfeldes orientiert sich an den Problemen, die sich einem Anwender der CA sukzessive stellen. Bei jedem dieser Schritte bestehen verschiedene Möglichkeiten der Vorgehensweise und der einsetzbaren Verfahren. Die Schritte werden beschrieben, die möglichen Verfahren dargestellt und vor dem Hintergrund vorliegender Forschungsbefunde kritisch diskutiert.

Die gewählte Technik der Einteilung des Prozesses der CA in Schritte bedeutet nicht, daß die Schritte und die dabei gewählten Verfahrensweisen von den Verfahren der anderen Untersuchungsschritte unabhängig sind. Vielmehr ist die Vorgehensweise notwendig, um die Komplexität des Forschungsfeldes handhabbar zu machen. Querverweise sind daher unumgänglich.

Das Ziel dieses Abschnitts besteht darin, den aktuellen Forschungsstand zur CA umfassend zu beschreiben und Forschungslücken aufzuzeigen bzw. Vorschläge zu deren theoretischen Ergänzung zu machen. Um die Forschungsbefunde darzustellen, werden Kenntnisse über Aggregationsmethoden (Abschnitt 4) und über die Beurteilung der Meßgüte einer CA (Abschnitt 5) vorausgesetzt.

Auf der Grundlage des aktuellen Forschungsstandes werden wir die Charakteristika der (aus unserer Sicht) derzeitig besten conjointanalytischen Vorgehensweise ableiten. In Abschnitt 6 soll dann durch eigene Simulationsstudien und empirische Untersuchungen versucht werden, ein derzeit 'optimales' CA-Verfahren weiter zu optimieren, indem einer der offenen Forschungsfragen nachgegangen wird.

Tabelle 3.1 gibt einen Überblick über den schrittweisen Ablauf einer CA und nennt typische Fragen, die sich ein Untersucher bei der Konstruktion einer CA-Untersuchung stellt. Da Abschnitt 3 prozeßhaft gegliedert ist, skizziert Tabelle 3.1 gleichzeitig den Aufbau des Abschnitts 3. In der letzten Spalte ist der Unterabschnitt genannt, der die jeweilige Thematik behandelt.

Tabelle 3.1: Der Prozeß der Conjoint Analyse, typische Fragen und Probleme und die Gliederung des Abschnitts 3.

Problembereich	Typische Fragen/ Probleme	Abschnitt
Untersuchungsziel	Was modellieren? Präferenzen oder Entscheidungsverhalten (Kaufverhalten, ...)?	3.1
Untersuchungssetting	Welche Auskunftspersonen? Korrespondenz der Untersuchungssituation mit der Entscheidungssituation?	3.1
Attribute, Attributstufen	Gewinnung von Objektattributen? Welche Attribute in der Untersuchung verwenden? Untersuchungsobjekte durch welche Attributstufen repräsentieren?	3.2 3.3
Präferenzstrukturmodell	Welchen Nutzenverlauf innerhalb der Attribute zulassen? Wie integrieren Beurteiler und Entscheider die Teilnutzen der Attribute?	3.4
Beurteilungsobjekte	Profilstärke der Beurteilungsobjekte? Welche Beurteilungsobjekte (Design)? Art der Objektpräsentation?	3.5
Abhängige Variable	Erhebung der abhängigen Variablen (Rangreihung, Paarvergleich)? Skalenniveau der abhängigen Variablen?	3.6
Verfahren der Nutzenschätzung	Welche Verfahren gibt es? Vergleichende Forschungsbefunde?	3.7
Parameterschätzverfahren	Welches Schätzverfahren bei welcher Erhebungsform der abhängigen Variablen?	3.8

Einen vertiefenden Überblick über die Vorgehensmöglichkeiten und einsetzbaren Verfahren gibt Tabelle A im Anhang. Sie kann auch als ausführliche Zusammenfassung des Abschnitts 3 angesehen werden.

3.1 Ziel und Setting einer CA-Untersuchung

Ähnlich wie in der Einstellungsforschung das Konstrukt der Einstellung menschliches Verhalten vorhersagen soll, soll auch die Präferenzmodellierung meist zur Vorhersage späteren Verhaltens dienen. Um diesem ultimativen Untersuchungsziel möglichst nahe zu kommen, müssen verschiedene Dinge beachtet werden. In der Literatur zur CA wurden die Inhalte dieses Abschnitts bisher weitgehend nicht problematisiert. Wir gehen von Forschungsergebnissen zum Zusammenhang von Einstellung und Verhalten aus und versuchen, Modelle und Erkennt-

nisse aus diesem Bereich für die CA fruchtbar zu machen. Hierzu sollen zunächst die Begriffe Einstellung, Präferenz, Verhaltensabsicht und Verhalten näher betrachtet und zueinander in Beziehung gesetzt werden.

3.1.1 Einstellung, Präferenz, Verhaltensabsicht und Verhalten

Die CA kam bisher überwiegend im Kontext produktpolitischer Entscheidungen zum Einsatz. Es ist sicher nicht zuletzt diesem pragmatisch orientierten Einsatzgebiet zuzuschreiben, daß die Konzeptualisierung der zu modellierenden kognitiven Vorgänge weitgehend nicht thematisiert wurden. Meist ist in der Literatur zur CA von Präferenzen die Rede, ohne dieses Konzept genauer zu spezifizieren oder in einen theoretischen Rahmen einzuordnen.

Hier sollen vier mögliche Untersuchungsziele einer CA getrennt werden. Wie Tabelle A verdeutlicht, unterscheiden wir die Modellierung von Einstellungen, von Präferenzen, von Verhaltensabsichten und von Verhalten. Die klare Trennung dieser Konzepte ist deswegen wichtig, weil sie sich einerseits in ihrer Nähe zu tatsächlichem, situationsgebundenem Verhalten und andererseits in der Art ihrer theoretischen Präzisierung unterscheiden.

Interessiert menschliches Verhalten am Markt, dann ist dieses z.B. in Form von Kaufentscheidungen direkt beobachtbar. Einstellungen, Präferenzen und Verhaltensabsichten dagegen sind am Markt nicht direkt beobachtbar, sondern können entweder indirekt erfragt werden (Präferenzen und Verhaltensabsichten) oder dienen als erklärende Konzepte zwischen beobachtbaren Sachverhalten, z.B. objektiven Objekteigenschaften und Kaufverhalten (Einstellungen). Das Konzept, das die größte Aufmerksamkeit in der Forschung erfahren hat, ist das der *Einstellung*. Dies bedeutet aber nicht, daß es auch theoretisch klar und eindeutig formuliert ist. Rosenberg und Hovland (1960) definieren Einstellungen als "predispositions to respond to some class of stimuli with certain classes of response" (S. 3). In ihrem Dreikomponentenmodell der Einstellung sind diese Reaktionsklassen affektiver Art (positive oder negative Gefühle gegenüber dem Einstellungsobjekt), kognitiver Art (Meinungen über ein Einstellungsobjekt) oder verhaltensnah (Verhaltenstendenzen, Verhaltensabsichten oder beobachtbares Verhalten). Gegensätzlich zu dieser Sichtweise wurden andere Konzeptualisierungen vorgeschlagen. Einige Autoren sehen die bewertende Komponente als das zentrale oder sogar alleinige Bestimmungsmerkmal von

Einstellungen. Fishbein und Ajzen (1975) und Petty und Cacioppo (1981) betrachten Einstellungen als ein allgemein gehaltenes und überdauerndes positives oder negatives Gefühl gegenüber dem Einstellungsobjekt. Gemäß dieser Sichtweise ist der zentrale Aspekt von Einstellungen die affektive Komponente des Modells von Rosenberg und Hovland (1960). Wir schließen uns dieser engen Fassung des Einstellungsbegriffs an, benutzen jedoch das Rosenberg-Hovland-Modell um die Konzepte Präferenz, Verhaltensabsicht und Verhalten einzuordnen.

Unter der *Präferenz* sei hier eine relativ überdauernde und relationale Beurteilung eines Objekts verstanden. Diese Beurteilung setzt sich sowohl aus affektiven als auch aus kognitiven Bestandteilen zusammen. Der Begriff ähnelt damit dem eng gefaßten Einstellungsbegriff: Präferenzen sind ebenfalls objektgebunden und zeitlich überdauernd. Er umfaßt die affektiv-bewertenden Reaktionen der Einstellung und geht darüber hinaus, da in Präferenzen auch elaborierte Meinungen über ein Objekt mit eingehen. Zusätzlich, und hier sehen wir den zentralen Unterschied zum Einstellungsbegriff, beinhalten Präferenzen *relationale Beurteilungen zwischen verschiedenen Objekten*. Präferenzen können nie losgelöst gegenüber nur einem Objekt bestehen. Sie bilden sich gegenüber den Objekten einer Menge vergleichbarer Objekte heraus. Ein Beispiel soll unsere Sichtweise verdeutlichen: Wenn jemand die Bekleidung eines bestimmten Herstellers als generell positiv beurteilt, dann sprechen wir von einer Einstellung. Typische Antworten auf die Frage, warum die Person zu diesem Urteil kommt, könnten z.B. sein: 'Gefällt mir eben sehr gut' oder 'ich habe mit diesen Sachen immer nur gute Erfahrungen gemacht und bin deswegen sehr zufrieden'. Affektiv-kognitive und relationale Präferenzen dagegen könnten so formuliert werden: 'X finde ich gut. Ich bin damit immer sehr zufrieden gewesen, weil die Bekleidung der Firma X immer auf dem neuesten modischen Stand ist, die Kleidung qualitativ hochwertig ist und auch oftmaliges Waschen fast ohne Spuren übersteht. Welche Bekleidungsunternehmen können mir das bieten? Selbst bei Y ... [relationale Beurteilung]'.

Die verhaltensmäßige Einstellungskomponente von Rosenberg und Hovland (1960) beinhaltet sowohl Verhaltensabsichten als auch offenes Verhalten. Die Beziehungen zwischen dem eng gefaßten und hier favorisierten Einstellungsbegriff und den verhaltensbezogenen Einstellungen versucht das Einstellungs-Verhaltens-Modell von Fishbein und Ajzen (1975) zu erklären. In

diesem Modell setzen *Verhaltensabsichten* positive Einstellungen gegenüber dem Objekt und bestimmte Präferenzen des Objekts unter vergleichbaren Objekten voraus. Dies ist jedoch nicht die einzige Voraussetzung, um Verhaltensabsichten zu generieren. Auch der Kontext des angestrebten Verhaltens spielt eine bedeutende Rolle. Wichtige absichtsbeeinflussende Faktoren sind soziale Normen und andere situative Gegebenheiten, wie z.B. persönliche Ressourcen, zeitliche Umsetzbarkeit etc. In noch stärkerem Maße bestimmen diese Faktoren das tatsächliche *Verhalten*.

Tabelle 3.2: Modellierungsziele einer Untersuchung und die jeweiligen kognitiven Repräsentationen der Objekte und der Situation als Beurteilungsgrundlage.



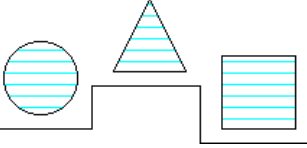
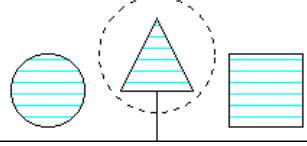
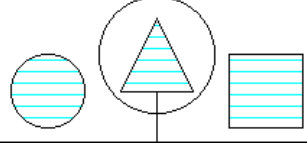
<p>Ausgangslage: objektive Eigenschaften der Objekte (z.B. Produkte, Produktkonzepte) und objektive Eigenschaften der Beurteilungs-/ Entscheidungssituation</p>	
<p>Untersuchungsziel: Modellierung von</p>	<p>Beurteilungsgrundlage:</p>
<p>Einstellungen</p>	<p>kognitive Repräsentation eines Objekts</p> 
<p>Präferenzen</p>	<p>relationale kognitive Repräsentation eines Objekts</p> 
<p>Verhaltensabsichten</p>	<p>relationale kognitive Repräsentation der Objekte und kognitive Repräsentation einer antizipierten Situation</p> 
<p>Entscheidungsverhalten</p>	<p>relationale kognitive Repräsentation der Objekte und kognitive Repräsentation einer vorhandenen Situation</p> 

Tabelle 3.2 veranschaulicht verschiedene Möglichkeiten von Modellierungszielen conjointanalytischer Untersuchungen und die jeweiligen kognitiven Repräsentationen der Objekte und der Situation. Im allgemeinen ist das ultimale Untersuchungsziel einer marketingorientierten CA-Untersuchung die Modellierung tatsächlichen Entscheidungsverhaltens über reale Objekte in gegebenen Situationen. Demgegenüber stehen die oft begrenzten Möglichkeiten, bei empirischen Untersuchungen (1) reale Objekte als Entscheidungsgrundlage bereitzustellen (z.B. bei Produktkonzepttests) und (2) die Entscheidungssituation der definierten Zielsituation anzugleichen, also z.B. Kaufverhaltenssituationen mit den darin vorkommenden begrenzten persönlichen Ressourcen, wie Zeit, Geld, ... und der begrenzten Verfügbarkeit von Kaufobjekten zu simulieren. Aus diesem Grunde begnügen sich conjointanalytische Untersuchungen meist mit der Modellierung von Präferenzen. Auf der Grundlage dieser Erkenntnisse wird dann mehr oder weniger explizit auf Verhaltensabsichten oder tatsächliches Verhalten geschlossen.

3.1.2 Präferenz und Verhalten

Es ist sicherlich hochgradig problematisch, ohne weitere Überlegungen von Präferenzen auf Verhalten zu schließen. Einige Autoren sind der Meinung, daß die Präferenz ein guter Prädiktor für das Kaufverhalten ist (z.B. Banks, 1950). Dagegen stellte man in vielen Untersuchungen im Bereich der Einstellungsforschung einen sehr geringen Zusammenhang zwischen Einstellung und Verhalten fest. Betrachtet man die Ähnlichkeit der Begriffe 'Einstellung' und 'Präferenz', dann sollten die Probleme der Vorhersage von Verhalten durch Einstellungen auch bei der Vorhersage durch Präferenzen auftreten.

Ajzen und Fishbein (1977) führen den geringen Zusammenhang zwischen Einstellung und Verhalten auf die geringe Korrespondenz hinsichtlich der Spezifität von Einstellungs- und Verhaltensmaßen zurück. Demzufolge werden Einstellungen meist sehr global erhoben (Wie stehen Sie zu türkischen Gastarbeitern?), während Verhaltensweisen sehr spezifisch beobachtet werden (Verhalten gegenüber einem türkischen Kellner bei einem Rendezvous mit einer begehrten Dame in einem erstklassigen Hotel). Es überrascht kaum, daß die geringe Korrespondenz der Meßspezifität beider Variablen zu einem nicht vorhersagbaren Zusammenhang dieser Maße führt.

Ajzen und Fishbein (1977) klassifizieren Einstellungen und Verhaltensweisen nach vier Aspekten:

- ❑ Handlung. Um welche Art von Handlungen/ Verhalten soll es gehen (etwas zum Eigengebrauch kaufen, etwas zum Verschenken kaufen, helfen, ...)?
- ❑ Ziel. Auf welches Ziel ist die Einstellung/ Handlung gerichtet (ein neues Produkt, ein guter Freund, eine Fremde Person, ...)?
- ❑ Kontext. In welcher Situation wird die Einstellung/ das Verhalten gezeigt (öffentlich, privat, mit leerem oder gefülltem Portemonnaie, ...)?
- ❑ Zeit. Zu welchem Zeitpunkt wird die Einstellung geäußert/ das Verhalten gezeigt (jetzt, innerhalb der nächsten zwei Jahre, ...)?

Die Korrespondenzhypothese der Autoren besagt, daß eine enge Beziehung zwischen Einstellung und Verhalten dann zu finden ist, wenn Einstellungs- und Verhaltensmaße in ihrem Spezifikationsgrad hinsichtlich dieser vier Aspekte übereinstimmen.

Das Modell von Ajzen und Fishbein (1977) ist einfach auf die Vorhersage von Verhalten aufgrund von Präferenzmessungen übertragbar. Werden bei der Präferenzmessung die Beurteilungsaufgaben sehr global gehalten (Tabelle A im Anhang: einfache Präferenz), dann sollte das Verhalten nur dann aus den Präferenzen vorhersagbar sein, wenn auch die Verhaltensmaße global gehalten sind. Dies kann z.B. dadurch erreicht werden, in dem das Verhalten in verschiedenen Situationen, mit unterschiedlichen Zielen und über einen längeren Zeitraum gemittelt erhoben wird. Interessiert dagegen ein spezifisches Verhalten in bekannten Situationen mit klar definierten Verhaltenszielen, dann fordert die Korrespondenzhypothese auch eine entsprechend spezifische Präferenzmessung, um daraus eine Vorhersage ableiten zu können.

Tabelle A im Anhang zeigt beispielhaft einige Möglichkeiten, das Untersuchungssetting so zu gestalten, daß Präferenzen spezifisch meßbar werden. Besonders wichtig bei der Konstruktion der Beurteilungs- und Entscheidungsaufgaben ist die klare Definition des entsprechenden Untersuchungssettings. Das Untersuchungssetting bestimmt die Entscheidungssituation. Diese kann durch die vier Aspekte der Entscheidungshandlung, des Entscheidungsziels, der Entscheidungssituation und der zeitlichen Gegebenheiten der Entscheidung charakterisiert werden. Korrespondieren die Entscheidungssituation der Unter-

suchung (Präferenz) mit der realen Entscheidungssituation (Verhalten), dann sollte die Vorhersagbarkeit des Verhaltens durch die Präferenzen zunehmen.

3.1.3 Auskunftspersonen

Die Vergleichbarkeit der Urteils- und Entscheidungssituation mit der realen Situation ist nur ein Aspekt der ökologischen Validität des Untersuchungssettings. Darüber hinaus müssen die Apn der Untersuchung mit tatsächlichen Entscheidern vergleichbar sein. Insbesondere ist kritisch zu prüfen, ob diese Vergleichbarkeit hinsichtlich der Auskunftsfähigkeiten und -kenntnisse sowie der Auskunftsmotivation gegeben ist (Trost, 1995). Bei allen Arten von Einstellungsmessungen ist das Problem bekannt, daß Apn meist auch ohne Kenntnisse des Objektbereichs spontan ein Einstellungsurteil abgeben. Auch bei fehlender Aufgabenmotivation ist mit einer Beantwortung der Fragen aufgrund von Effekten der sozialen Erwünschtheit zu rechnen.

Um das geforderte Niveau der Fähigkeit und Motivation der Apn sicherzustellen, sind eine Vielzahl von Interventionsmöglichkeiten denkbar. Entsprechend der Klassifikation von Interventionsstrategien nach Pawlik (1976) sind Selektionsstrategien und Modifikationsstrategien sowohl auf der Personen- als auch der Bedingungebene denkbar. Um die *Vergleichbarkeit der Fähigkeiten der Apn* mit den tatsächlichen Entscheidern herzustellen, wird in der Regel eine Personenselektion erfolgen müssen. Auch Bedingungsmodifikationsstrategien (z.B. Vorgabe eines bestimmten Interviews entsprechend den Kenntnissen einer Ap) oder Personenmodifikationsstrategien (z.B. Schulung der Apn) sind denkbar. Um die Vergleichbarkeit der Motivation der Apn mit tatsächlichen Entscheidern zu erreichen, wird normalerweise die *Anhebung der Motivation in der Interviewbedingung* notwendig sein. Diese kann über eine Personenselektion erfolgen. Sie kann u.U. auch nachträglich durchgeführt werden, indem Apn mit einer niedrigen Modellanpassung bei der Datenanalyse gesondert behandelt werden. In aller Regel wird man eine Bedingungsmodifikationsstrategie anwenden. Diese besteht z.B. in einer anregenden Gestaltung des Interviews oder in der Sicherstellung eines hohen Commitments, indem die Apn persönliche positive Konsequenzen durch eine sorgfältige Bearbeitung der Interviewaufgaben erwarten.

3.2 Attribute

Unabhängig davon, welches Verfahren zur Nutzenschätzung herangezogen wird (Abschnitt 3.7), besteht das Prinzip der Präferenzstrukturmodellierung darin, daß über mehrere uniattributive (bei kompositorischen Verfahren) oder multiattributive (bei dekompositorischen Verfahren) Objekte zunächst Präferenzurteile gebildet werden müssen. Ein grundlegender Schritt bei der Präferenzstrukturmodellierung ist daher die Festlegung dieser Objektdimensionen, der Attribute.

Die schrittmachenden theoretischen Arbeiten zur CA im Marketing (Green & Srinivasan, 1978; Green & Srinivasan, 1990; Green & Wind, 1973) und die Überblicksarbeiten zum kommerziellen Einsatz der CA (Cattin & Wittink, 1982; Wittink & Cattin, 1989; Wittink, Vriens & Burhenne, 1994) nennen folgende Problembereiche: Art des Präferenzmodells, Art der Beurteilungsstimuli, Art der Präferenzurteile, Variante der Nutzenschätzung, eingesetzte Parameterschätzmethoden, Reliabilitäts- und Validitätsschätzung und Wahlsimulationen. Dagegen werden die Methoden zur Gewinnung der Attribute und der Attributstufen weitgehend vernachlässigt oder höchstensfalls kurz als problematisch angesprochen. Diese Vernachlässigung ist insofern besonders kritisch zu beobachten, da die Auswahl der Attributmenge sowohl die nachfolgenden Schritte der CA beeinflusst, insbesondere aber auch die Meßgüte ihrer Ergebnisse von ihr abhängen (Cattin & Weinberger, 1980). In einer Bestandsaufnahme der Anwendung der CA in der Marktforschung bemängelt auch Schubert (1991) das mangelnde Problembewußtsein bei der Attributgewinnung und Attributauswahl. Fast die Hälfte (44,5%) der von ihm zusammengestellten 128 Untersuchungen machen keine Angaben darüber, wie es zur Gewinnung und Auswahl der untersuchten Attribute gekommen ist. Auch Wittink et al. (1994) kommen zu dem Schluß, daß es nützlich wäre, im Zusammenhang mit der CA zu fundamentalen Forschungsfragen, wie z.B. der Auswahl und Definition der Attribute und Attributstufen zurückzukehren. Dies ist um so schwerwiegender, da bereits seit über 15 Jahren diese Problematik erkannt und veröffentlicht ist (vgl. z.B. Acito, 1977; McCullough & Best, 1979; Sheluga, Jaccard & Jacoby, 1979).

Vergleichsweise ausführliche Darstellungen und Lösungsvorschläge zu diesem Problem finden sich in einigen deutschsprachigen Veröffentlichungen zur CA. Schweikl (1985) und Schubert (1991) systematisieren Attributarten und

beschreiben Methoden zur Gewinnung und Auswahl der Untersuchungsattribute. Diese Ansätze sollen hier dargestellt und erweitert werden.

3.2.1 Grundlegende Anforderungen an das Attributset

Bei conjointanalytischen Untersuchungen wird ein nicht unerheblicher Aufwand bei der Datenerhebung, Datenanalyse und -weiterverarbeitung betrieben. Die Grundbausteine der gebildeten PSM sind die Attribute. Es ist aus diesen Gründen nicht zu rechtfertigen, daß die Basis der Präferenzstrukturmodellierung, die präferenzrelevanten Attribute, eine untergeordnete Bedeutung in der konzeptionell und pragmatisch orientierten Forschung erfahren.

Die Anforderungen an das Attributset einer Untersuchung kann aus mehreren Überlegungen abgeleitet werden:

- ❑ Die gewählten Attribute sollten den *axiomatischen Annahmen des additiven CM* (vgl. Krantz et al., 1971) entsprechen. So fordern diese, daß die Attribute unabhängig voneinander realisierbar sind (empirische Unabhängigkeit) und ihre Wirkung auf die Präferenz unabhängig voneinander entfalten (Präferenzunabhängigkeit). Darüber hinaus sollten die Attribute dem Axiom der 'Bedeutsamkeit' genügen (Krantz et al., 1971). Diese Voraussetzungen des CM werden in die Forderungen nach determinanten, empirisch unabhängigen und präferenzunabhängigen Attributen eingehen.
- ❑ Die Anzahl der bei Beurteilungen und Entscheidungen tatsächlich verwendeten Attribute ist durch die *beschränkte Informationsverarbeitungskapazität* des Menschen begrenzt. Diese vielfach belegte Tatsache führt zum Konzept der relevanten Attribute.
- ❑ *Forschungsökonomische Gegebenheiten* zwingen oft dazu, die Menge der Attribute, die in eine Untersuchung eingehen, gegenüber der Menge der relevanten Attribute zu reduzieren. Diese Reduktion kann anhand der Zielgebundenheit der Forschung erfolgen. Beispielsweise könnte die Umsetzbarkeit von Forschungsergebnissen ein Forschungsziel sein. Diese Überlegung führt zum Konzept der maßgeblichen Attribute.

Im Bereich des Marketing kann über die beiden letzten Punkte die zusammenfassende Forderung erhoben werden, daß die Attributmenge sowohl die Attribute enthalten sollte, die Konsumentenentscheidungen determinieren, als auch solche, die in Managemententscheidungen umsetzbar sind. Es kann sich dabei

z.B. um produkt-, preis-, kommunikations- oder distributionspolitische Entscheidungen handeln (vgl. Cattin & Wittink, 1981).

Im folgenden möchten wir sinnvoll unterscheidbare Arten von Attributen definieren. Danach werden wir einige Verfahren der Gewinnung von Attributen und deren Selektion in Anlehnung an Schweikl (1985) und Schubert (1991) darstellen und ordnen. Der Abschnitt schließt mit der Diskussion der Repräsentation und Definition der Attribute und Fragen zum Attributset. Eine Übersicht gibt Tabelle A.

3.2.2 Art der Attribute

Wir unterscheiden sechs Arten von Attributen. Die potentiell relevanten Attribute stellen die größte Attributmenge dar. Nachfolgende Attributarten sind jeweils Teilmengen der vorhergehenden Art. Letztendlich gehen die maßgeblichen Attribute in die Untersuchung ein. Ein Stufenverfahren, welches die Attributmenge von der Menge der potentiell relevanten Attribute zur Menge der maßgeblichen Attribute reduziert, ist in Abschnitt 3.2.4 dargestellt. Zunächst soll aber definiert werden, wie wir diese Konzepte verstehen.

1. potentiell relevante Attribute
2. determinante, saliente und wichtige Attribute
3. empirisch unabhängige Attribute
4. präferenzunabhängige Attribute
5. relevante Attribute
6. maßgebliche Attribute

3.2.2.1 *Potentiell relevante Attribute*

Eine Grundannahme der multiattributiven Entscheidungsmodelle ist, daß jede Alternative/ jedes Objekt auf einer Menge von Attributen beschrieben werden kann. Im Rahmen der Modellierung von Präferenzen sind die potentiell relevanten Attribute die Menge aller Objektmerkmale, auf welchen sich die Objekte einer Objektklasse unterscheiden können und welche damit einen potentiellen Einfluß auf die Präferenzen zwischen den Objekten ausüben. Die potentiell relevanten Attribute umfassen auch solche Dimensionen, auf denen sich vorhandene Objekte einer Objektklasse zu einem gegebenen Zeitpunkt nicht unterscheiden. Dies kann z.B. beim Einsatz der CA zur Produktinnovation bedeutsam sein, bei welchen sich Beurteilungsobjekte auf präferenzwirksamen Attributen unterscheiden, obwohl diese Unterschiede im empirischen Objektraum bisher nicht vorhanden sind.

3.2.2.2 *Determinante Attribute*

Die determinanten Attribute (Myers & Alpert, 1968) sind eine Teilmenge der potentiell relevanten Attribute. Ihr Wesensmerkmal besteht nicht mehr nur in der Unterscheidbarkeit von Objekten, sondern darin, daß ein Unterschied auf diesen Attributen tatsächlich eine unterschiedliche Objektpräferenz determiniert.

Einige Autoren grenzen von den determinanten Attributen die salienten ('salient attributes') und die wichtigen Attribute ('important attributes') ab (z.B. Alpert, 1971, 1980; Myers & Alpert, 1968, 1977; Pras, 1975). Mit dem Begriff *saliente Attribute* werden diejenigen Attribute bezeichnet, die bei einer Befragung von potentiellen Entscheidern als erstes genannt werden (vgl. Abschnitt 3.2.3 Elicitationstechnik), also leicht aus dem Gedächtnis abrufbar sind. Saliente Attribute sind jedoch nicht mit wichtigen Attributen gleichzusetzen, da die Leichtigkeit des Gedächtnisabrufs auch von anderen Faktoren, wie z.B. der Zeitspanne der letztmaligen Konfrontation der Apn mit dem Attribut oder dem Befragungskontext abhängig ist (z.B. Kahneman & Tversky, 1973; Loftus, 1974). *Wichtige Attribute* sind diejenigen determinanten Attribute, die von Beurteilern als besonders präferenzrelevant erachtet werden. In unserem System sind die salienten und wichtigen Attribute eine spezielle Untermenge der determinanten Attribute.

Ein anschauliches Beispiel zur Unterscheidung dieser Begriffe stammt von Schubert (1991). Fragt man Konsumenten, welche Produkteigenschaften ihnen beim Kauf von Autoreifen besonders wichtig sind, so wird am häufigsten das Merkmal 'Sicherheit' genannt. Läßt man gleichzeitig mehrere Reifenmarken bezüglich des Merkmals 'Sicherheit' beurteilen, so stellt man fest, daß sich die Marken hinsichtlich der Einschätzung auf dieser Dimension kaum unterscheiden. In diesem Fall sollte das Attribut die Präferenz für verschiedene Reifenmarken nur wenig beeinflussen. Andere Attribute, bei welchen große Unterschiede zwischen den Marken vorhanden sind, determinieren die Beurteilung wesentlich stärker. Dies könnte z.B. der Preis sein. Das Attribut 'Sicherheit' ist somit ein salientes Attribut, weil es so häufig als Hauptkriterium genannt wurde und ein determinantes Attribut, weil es beurteilungsrelevant ist und Unterschiede zwischen den Objekten subjektiv wahrgenommen werden. Es ist jedoch kein wichtiges Attribut, weil die Merkmalsunterschiede zwischen den Objekten nur gering sind, und es damit zu geringen Beurteilungsunterschieden führt.

3.2.2.3 Empirisch unabhängige Attribute

Eine Menge empirisch unabhängiger Attribute liegt dann vor, wenn durch die Ausprägung eines Objekts auf einem Attribut nicht seine Ausprägung auf einem anderen Attribut vorhergesagt werden kann. Die Attributmenge ist damit nicht redundant. Ein Beispiel für empirisch abhängige Attribute wäre die Motorleistung eines Kraftfahrzeugs und die Beiträge für die Haftpflichtversicherung.

3.2.2.4 Präferenzunabhängige Attribute

Eine Attributmenge ist dann präferenzunabhängig, wenn der Präferenzbeitrag einer bestimmten Ausprägung eines Objekts auf einem Attribut unabhängig von der Ausprägung des Objekts auf allen anderen Attributen ist. Eine Attributmenge, die die Attribute 'Farbe eines Autos' und 'Ausführung eines Fahrzeugs' enthält, wäre z.B. nicht präferenzunabhängig, wenn Mittelklassewagen mit der Farbe 'gelb' eine Präferenzminderung, Sportwagen mit derselben Farbe aber eine Präferenzsteigerung erführen. Die Präferenzunabhängigkeit der Attribute ist eine Voraussetzung des additiven CM. Bei der Verwendung von nicht

ausschließlich additiven Präferenzintegrationsmodellen entfällt die Forderung der Präferenzunabhängigkeit.

3.2.2.5 *Relevante Attribute*

Eine Attributmenge enthält dann ausschließlich relevante Attribute, wenn alle Attribute wichtige determinante Attribute sind, die zueinander in einem empirisch unabhängigen und präferenzunabhängigen Verhältnis stehen. Der Begriff des relevanten Attributs umfaßt damit alle vorgenannten Eigenschaften. Eine notwendige Bedingung für diese Attributart ist damit, daß alle relevanten Attribute in einem PSM statistisch signifikant zur Erklärung der beobachteten Objektpräferenzen beitragen.

3.2.2.6 *Maßgebliche Attribute*

Die bisher genannten Attributformen sind alle aus der Perspektive der Objektentscheider definiert. Die Population der Objektentscheider besteht aus denjenigen, die über die Objekte einer Objektklasse urteilen oder entscheiden. Im Bereich des Marketing sind dies vornehmlich Konsumenten, Abnehmer und Absatzmittler. Mit dem Begriff 'maßgebliche Attribute' möchten wir die Perspektive vom Objektentscheider zum Programmentscheider wechseln. Ein Programm-entscheider versucht die Beurteilungen und Entscheidungen von Objektentscheidern durch die Bereitstellung eines Objektprogramms zu beeinflussen. Im absatzwirtschaftlichen Kontext sind dies zumeist die Anbieter (bzw. deren Manager), die z.B. Produkte oder Dienstleistungen mit bestimmten Eigenschaften anbieten.

Sind aus Sicht der Objektentscheider die relevanten Attribute einer Objektklasse festgelegt, dann ist es Aufgabe der Programmentscheider, daraus die Teilmenge der maßgeblichen Attribute zu entnehmen. Die maßgeblichen Attribute zeichnen sich vornehmlich dadurch aus, daß sie der Beeinflußbarkeit oder dem Beeinflussungswillen der Programmentscheider unterliegen. Ein Beispiel soll dies verdeutlichen. Es hat sich bei der Objektklasse 'Mikrocomputer' herausgestellt, daß die Rechengeschwindigkeit ein relevantes Attribut für potentielle Käufer (Objektentscheider) darstellt. Ein Computerhersteller (Programmentscheider) weiß aber, daß der Einkaufspreisunterschied zwischen unterschiedlichen am Markt befindlichen Prozessoren nur sehr gering ist. Er

wird also ohnehin den schnellsten verfügbaren Prozessor in seine Produkte einbauen. Die Entwicklung noch schnellerer Prozessoren ist technisch noch nicht möglich. Damit ist aus seiner Sicht keine Beeinflußbarkeit des Attributs vorhanden. Eine mögliche Modellierung der Präferenzstruktur mit diesem Attribut hätte für ihn keinen Nutzen. Der Begriff der maßgeblichen Attribute ist vergleichbar mit der Schnittmenge der relevanten Attribute und den Attributen, die Shocker und Srinivasan (1974) als 'actionable attributes' bezeichnen.

Die Bildung der Teilmenge der maßgeblichen Attribute aus der Menge der relevanten Attribute ist meist aus forschungsökonomischen Gründen notwendig, da bei einer CA mit der Anzahl der untersuchten Attribute entweder die Beurteilungsobjekte überproportional komplex wird oder die Interviewdauer überproportional zunimmt.

3.2.2.7 *Fazit*

Letztendlich stellen sich bei der Attributauswahl jeder Untersuchung zur multiattributiven Präferenzstrukturmodellierung zwei Fragen:

1. Wie können potentiell relevante Attribute gefunden werden?
2. Wie kann die Menge der potentiell relevanten Attributen gemindert werden, um den axiomatischen Annahmen des additiven CM, der begrenzten Informationsverarbeitungskapazität des Menschen und forschungsökonomischen Gegebenheiten Rechnung zu tragen und die CA sinnvoll anwenden zu können?

Abschnitt 3.2.3 behandelt die Methoden zur Gewinnung potentiell relevanter Attribute. In Abschnitt 3.2.4 werden Selektionsmöglichkeiten dargestellt.

3.2.3 Verfahren zur Gewinnung potentiell relevanter Attribute

Eine einfache Möglichkeit, potentiell relevante Attribute ausfindig zu machen, ist die Analyse von Dokumenten. Denkbar ist dabei die Durchsicht von Produktbeschreibungen, die Analyse von Testberichten, Werbeanzeigen, Prospekten und die Heranziehung von Fachliteratur. Ergänzend zu diesen anfänglichen Bemühungen sollte in jedem Falle eine Befragung von Experten erfolgen. Vorzugsweise sollten sowohl Objektentscheider (Konsumenten, Absatzmittler) als auch Programmentscheider (Anbieter, Hersteller, Manager) befragt werden.

Diese Vorgehensweise stellt sicher, daß keine potentiell relevanten Attribute übersehen werden.

Obwohl die extensive Sammlung von Attributen mittels verschiedener Datenquellen einsichtig ist, wird diese Vorgehensweise eher selten angewandt. In ihrer Untersuchung zur kommerziellen Anwendung der CA kommen Cattin und Wittink (1982) zu dem Schluß, daß bei weitem nicht in allen Fällen die Sammlung der Attribute unter Einbeziehung der Konsumenten, also der eigentlichen späteren Objektentscheider, erfolgt. Hierfür dürften hauptsächlich Kostengründe verantwortlich sein. Schubert (1991) berichtet in seiner sekundäranalytischen Untersuchung von 128 veröffentlichten und anders zugänglichen CA. Von diesen machten 44,5% keine Angaben zur Herkunft der Attribute. In 20,3% der Fälle wurden die Attribute vom Produktmanagement (teilweise unter Zuhilfenahme von literarischen Quellen) festgelegt. Lediglich 35% der Studien bestimmten die Attribute unter Einbeziehung der Objektanwender.

In den folgenden Abschnitten beschreiben wir einsetzbare Befragungsverfahren, die zur Ermittlung potentiell relevanter Attribute geeignet sind (vgl. auch Quack, 1980; Pras, 1975; Grunert, 1985). Grundsätzlich können alle nachfolgend aufgeführten Verfahren sowohl bei Programm- als auch Objektentscheidern eingesetzt werden. Zur besseren Übersicht unterteilen wir die Verfahren in direkte und indirekte Verfahren. Bei indirekten Verfahren werden die Apn nicht direkt nach Dimensionen gefragt. Vielmehr werden diese erschlossen. Einen Überblick gibt Tabelle A im Anhang.

3.2.3.1 *Direkte Befragung*

Bei dieser einfachen Methode werden objektklassenkundige Apn gebeten, Merkmale zu nennen, die Unterschiede zwischen den Objekten kennzeichnen. Man kann dabei auch direkt nach Merkmalen fragen, die Bevorzugen oder Kaufentscheidungen beeinflussen. Die Antworten werden kategorisiert. Die relative Häufigkeit der Nennungen kann als Indikator für die Wichtigkeit eines Merkmals herangezogen werden (Myers & Alpert, 1968; zur Problematik der Unterscheidung von wichtigen Attributen und salienten Attributen siehe Abschnitt 3.2.2.2). Kritische Variablen bei der direkten Befragung sind die Aus-

kunftswilligkeit der Apn und die Auskunftsfähigkeit der Apn für z.T. hochautomatisierte Informationsverarbeitungsvorgänge.

3.2.3.2 *Elicitationstechnik*

Bei diesem Verfahren werden objektklassenkundige Apn mit einem Stimulus konfrontiert. Sie haben die Aufgabe, alle Eigenschaften zu nennen, die ihr zum Stimulus einfallen (z.B.: 'An was denken Sie, wenn ich 'Käse' sage?'). Diese Methode ist der direkten Befragung ähnlich. Sie stellt die Fragen aber weniger direktiv.

3.2.3.3 *Gruppeninterview und Gruppendiskussion*

Das Gruppeninterview stellt eine Alternative zu Interviews mit Einzelpersonen dar. Sie haben den Vorteil, daß Gruppen mit offenem Klima kreativer sein können als Einzelpersonen, weil manche Redebeiträge neue Ideen bei anderen Gruppenmitgliedern hervorrufen können. Darüber hinaus sind sie meist ökonomischer, da die Mehrfachnennung einzelner Attributvorschläge gegenüber mehreren Einzelinterviews unterbleibt. Kritisch dagegen sind die kaum kontrollierbaren Prozesse des sozialen Einflusses zu beurteilen. So kann z.B. die Auskunftswilligkeit in Gruppen eingeschränkt sein. Nach der Diskussion ist kein objektiver Indikator für die Wichtigkeit der Attribute durch die relative Häufigkeit der Nennung vorhanden. Wird die Wichtigkeit in einer Gruppendiskussion ermittelt, ist zu erwarten, daß sich die Meinung weniger Mitglieder der Gruppe durchsetzt.

3.2.3.4 *Kreativitätstechniken*

Der Einsatz von Kreativitätstechniken (z.B. Brainstorming, Tagebuchtechnik; vgl. Glover, Ronning & Reynolds, 1989; Schlicksupp, 1985) ist sowohl bei Einzel- als auch Gruppeninterviews denkbar. Insbesondere wenn es sich um Objektinnovationen handelt, kann ihr Einsatz indiziert sein.

3.2.3.5 *Repertory Grid Technik*

Die Repertory Grid Technik stammt aus der Theorie der persönlichen Konstrukte (Kelly, 1955) und hat zum Ziel, Bewertungsdimensionen auf individueller Ebene zu erfassen. Das Verfahren geht dabei folgendermaßen vor: Üblicherweise werden die Objekte einer Objektklasse auf jeweils einer Karte notiert. Die Objekte werden entweder vorgegeben oder von der Ap generiert. Aus der Gesamtheit der Objektkarten werden einer Ap jeweils drei zufällig ausgewählte Karten vorgelegt. Die Ap hat die Aufgabe, die zwei Objekte auszuwählen, die sich am ähnlichsten sind. Anschließend soll sie den Grund der Ähnlichkeit mit einem Begriff benennen (Konstrukt-pol) und den Unterschied der beiden ähnlichen Objekte zum dritten Objekt angeben (Kontrast-pol). Der verbal benannte Konstrukt- und Kontrast-pol wird schriftlich als Bewertungsdimension festgehalten. Dann wird das nächste Kartentriple gezogen, der Ap die gleiche Aufgabe gestellt usw., bis keine neuen Dimensionen mehr benannt werden können. Auf diese Weise hat man eine Menge der individuellen Bewertungsdimensionen erstellt¹.

Alle Objekte und Dimensionen werden dann in einem Gitter, daher der Name der Technik, abgetragen. Die Objekte werden in die Zeilen und die Dimensionen, definiert durch Konstrukt- und Kontrast-pol, in die Spalten eingetragen. Die Ap hat nun die Aufgabe, alle Objekte auf den Dimensionen auf einer Punkteskala einzustufen. Ein Beispiel dieser Aufgabe gibt Tabelle 3.3. Im Beispiel werden die Objekte auf einer 5-Punkteskala eingestuft. Ein niedriger Wert bedeutet, daß das Objekt nahe dem Konstrukt-pol wahrgenommen wird. Anschließend kann die Bewertungsmatrix einer Faktorenanalyse unterzogen werden. Als Ergebnis der Faktorenanalyse werden im Regelfalle ähnliche Bewertungsdimensionen zu Faktoren zusammengefaßt. Die Faktoren stehen zueinander orthogonal. Detaillierte Beschreibungen der Interview- und Datenanalysetechniken sind bei Raeithel (1993), Riemann (1991), Scheer (1993) zu finden. Software zur computergestützten Auswertung von Repertory Grid Untersuchungen stellen z.B. Mitterer und Adams-Webber (1988) und Riemann (1991) bereit. Bringmann (1992) und Willutzki und Raeithel (1993) besprechen angebotene Software.

¹ Die weitere Vorgehensweise der Repertory Grid Technik soll hier dargestellt werden, obwohl es sich streng genommen bereits um eine Attributselektionstechnik (vgl. Abschnitt 3.2.4) handelt.

Tabelle 3.3: Ein fiktives Beispiel des Objekte-Beurteilungsdimensionen-Gitters.

Bewertungsdimension	sportlich - normal	teuer - erschwinglich	Qualität gut - weniger gut	viel Raum - wenig Raum	Benzinverbrauch hoch - niedrig
Objekt					
BMW 320	2	2	2	2	2
VW Golf	4	4	2	3	4
Mercedes 240 D	5	2	2	1	3
Porsche 911	1	1	1	5	1
Honda Civic	4	5	4	4	5
Opel Kadett	4	4	4	3	4
Fiat Panda	5	5	5	4	4
BMW Z1	1	1	1	5	1
Mercedes 190	2	2	2	2	3

Die gewinnbringende Einsatzfähigkeit der Technik im Zusammenhang mit wirtschaftlichen Fragestellungen kann mittlerweile als gesichert gelten (z.B. Jankowicz, 1990; Müller-Hagedorn & Vornberger, 1979). Obwohl die Repertory Grid Technik zur Gewinnung von Attributen bei der CA immer wieder vorgeschlagen (z.B. Green & Srinivasan, 1978) und als leistungsfähig beurteilt wurde (Jain, Mahajan & Malhotra, 1979), wird sie eher selten tatsächlich eingesetzt (eine der wenigen Untersuchungen, die dies tun, ist z.B. Jain, Acito, Malhotra & Mahajan, 1979). Cattin & Wittink (1982) kommen zu dem Ergebnis, daß zwar 7 ihrer 17 befragten Unternehmen Erfahrungen mit der Technik zur Attributgewinnung besitzen, diese aber nur selten tatsächlich eingesetzt haben.

3.2.3.6 *Lautes Denken*

Die folgenden Verfahren werden zu den indirekten Verfahren gezählt. Bei der Methode des lauten Denkens werden die Apn z.B. gebeten, bei einer realen oder vorgestellten Kaufentscheidungssituation alle Gedanken laut auszusprechen, die ihnen in den Sinn kommen. Die lauten Gedanken werden protokolliert (z.B. durch einen Tondatenträger). Die Technik stellt somit primär eine Prozeßverfolgungsmethode dar. Aus den Protokollen werden von Experten

potentiell relevante Beurteilungs- und Entscheidungsdimensionen entnommen bzw. erschlossen. Diese Methode kommt nur selten zum Einsatz. Ihre Anhänger scheinen sie jedoch stark zu favorisieren (Cattin & Wittink, 1982).

3.2.3.7 Tiefeninterview

Diese Technik setzen z.B. Anttila, van den Heuvel und Möller (1980) zur Gewinnung potentiell relevanter Attribute ein. Die Datengewinnung erfolgt hierbei anhand eines unstrukturierten Interviews, mit dem Ziel, schwierig verbalisierbare Ursachen des Konsumentenverhaltens zu erfahren. Wie immer bei unstandardisierten Befragungstechniken ist bei diesem Verfahren mit vergleichsweise hohen Interviewereinflüssen zu rechnen. Der Erhebungsaufwand ist, gemessen an anderen Verfahren, hoch.

3.2.3.8 Projektive Verfahren

Der Klasse der projektiven Verfahren ist gemeinsam, daß die Fragen an die Apn möglichst wenig direktiv gestellt werden. Dies kann z.B. dadurch erreicht werden, daß einer Ap alle möglichen Freiheiten bei der Antwort gelassen werden (z.B. Wortassoziationen, Satzergänzungen, Anfertigen von Cartoons), oder sie die Möglichkeit bekommt, Stimuli zu kommentieren, deren Bedeutungshaltigkeit vage gehalten ist (z.B. Geschichten zu vorgegebenen mehrdeutigen Bildern erfinden). Schweikl (1985) kommt zu er Erkenntnis, daß diese Art der Verfahren im Zusammenhang mit der CA weniger eingesetzt wird.

3.2.3.9 Multidimensionale Skalierung

Eine multidimensionale Skalierung (MDS; Coombs, 1950; Kruskal, 1964a, 1964b; Tucker & Messick, 1963) findet ihren Ausgangspunkt meist in Ähnlichkeitsurteilen über Objekte einer bestimmten Objektklasse. Wie bei der CA werden damit Globalurteile über komplexe Objekte abgefragt. Im Gegensatz zur CA sind bei der MDS jedoch keine Bewertungsdimensionen vorgegeben. Vielmehr wird die Dimensionalität der Bewertungen aus den Daten geschätzt. Hat man sich für eine bestimmte Anzahl von Bewertungsdimensionen entschieden, bleibt dem Forscher noch die semantische Interpretation der Dimensionen. Diese können als Hinweise dienen, welche Attribute als bewertungsrelevant

angesehen werden. Wie auch die Repertory Grid Technik ist die MDS nicht nur ein Verfahren zur Gewinnung von potentiell relevanten Attributen sondern, es erfüllt gleichzeitig Selektionsaufgaben. Beispielsweise können die gewonnenen Bewertungsdimensionen orthogonal angelegt sein, was in präferenzunabhängigen Attributen resultieren sollte. Eine anwendungsorientierte Beschreibung der MDS kann man bei Backhaus, Erichson, Plinke und Weiber (1994) finden.

3.2.4 Verfahren zur Selektion der Attribute

Sind die potentiell relevanten Attribute einer Objektklasse sorgfältig ausgewählt worden, kann mit der Selektion dieser Attribute zur Durchführung einer CA begonnen werden. Meist wird die Menge der potentiell relevanten Attribute sehr groß sein, so daß die dekompositorische Schätzung der erforderlichen Parameter mit einem unvermeidbaren Aufwand verbunden wäre. Aus diesem Grund wird die Attributmenge reduziert.

3.2.4.1 Grundüberlegungen

Genau wie bei der Gewinnung der Attribute sind auch zur Selektion der Attribute Experten des Objektbereichs erforderlich. Alle vorherigen Überlegungen zur Auswahl der Personen gelten auch hier. Die Urteile der Experten können einzeln durch Befragung erhoben und anschließend statistisch weiterverarbeitet werden. Eine andere Möglichkeit ist die Gewinnung der Information in Gruppeninterviews oder Gruppendiskussionen. Um solche Befragungen strukturieren zu können, soll im folgenden Abschnitt ein Stufenverfahren vorgeschlagen werden.

3.2.4.2 Stufenverfahren

Auf der Grundlage des in Abschnitt 3.2.2 vorgeschlagenen Attributkategorisierungssystems ist ein Stufenverfahren zur Selektion der potentiell relevanten Attribute anwendbar. Dieses Stufenverfahren dient zur Strukturierung der Selektionsprobleme und gibt Experten(-gruppen) eine Möglichkeit an die Hand, schrittweise die Menge der potentiell relevanten Attribute sinnvoll zu verringern.

Im *ersten Schritt* werden die potentiell relevanten Attribute gesammelt. Dies erfolgt z.B. mit den in Abschnitt 3.2.3 dargestellten Verfahren. Im *zweiten Schritt* reduziert man diese Attributmenge auf die determinanten Attribute. Alpert (1971) untersucht verschiedene direkte und indirekte Methoden zur Identifikation determinanter Attribute. Er kommt zu dem Schluß, daß direkte Befragungsmethoden eher geeignet sind, diese Attribute zu identifizieren als indirekte Methoden, die sich Regressionstechniken (z.B. multiple Regression, Diskriminanzanalyse) bedienen. Er gibt insbesondere der Technik der 'dualen Frage' (dual questioning, Myers & Alpert, 1968) den Vorzug. Bei dieser Technik werden die Apn gefragt, (1) für wie wichtig sie ein Attribut bei der Wahlentscheidung erachten, und (2) für wie groß sie den Unterschied zwischen konkurrierenden Objekten auf dem jeweiligen Attribut einschätzen. Beide Einschätzungen erfolgen auf einer numerischen Skala und werden miteinander multipliziert. Attribute, die wichtig sind und auf denen die Objekte einen Unterschied zeigen, werden als determinante Attribute gewählt.

In Abhängigkeit von Technik zur Identifikation determinanter Attribute können zwei Probleme auftreten: Das Konfundierungsproblem und das Redundanzproblem (vgl. Pras, 1975). Das Konfundierungsproblem besteht darin, daß bei der Anwendung von Regressionstechniken als Methode zur Identifikation determinanter Attribute diese mit den nachfolgenden dekompositorischen Verfahren der Präferenzstrukturmodellierung, das die Attribute verarbeitet, konfundiert ist. Dies soll durch ein Beispiel verdeutlicht werden. Werden die geschätzten Standardpartialregressionskoeffizienten eines multiplen linearen Regressionsmodells als Indikatoren der Attributwichtigkeit herangezogen, dann werden durch diese Technik determinante Attribute mit nichtlinearen Nutzenfunktionen herausgefiltert. Andererseits verbleiben weniger wichtige, aber lineare Attribute zur nachfolgenden linear-additiven Präferenzstruk-

turmodellierung. Da die CA der Regressionstechnik, die zum Auffinden der wichtigen oder determinanten Attribute verwendet wurde, sehr ähnlich ist, bestätigt die CA immer die Wichtigkeit der Attribute.

Bei Fragemethoden (z.B. direkte Befragung oder duale Frage) tritt das Problem der Konfundierung nicht auf. Beide Methoden haben jedoch das Redundanzproblem gemeinsam. Da die Abhängigkeiten der Attribute zueinander nicht bekannt sind, können in der Menge der potentiell relevanten Attribute Redundanzen auftreten. Bei Regressionstechniken stellt diese Multikollinearität ein grundsätzliches Problem dar, da eine der Grundvoraussetzungen für die Anwendbarkeit der Methode verletzt ist. Bei Fragemethoden zur Identifikation determinanter Attribute werden solche Abhängigkeiten nicht entdeckt. Bei unreflektierter Anwendung z.B. der dualen Fragetechnik resultiert möglicherweise eine Menge mehrerer determinanter, aber weitgehend redundanter Attribute.

Zur Lösung dieser zwei Probleme schlägt Pras (1975) vor, eine besondere direkte Fragetechnik zur Identifikation determinanter Attribute anzuwenden. Die Pras'sche Technik kann als Erweiterung der dualen Fragetechnik nach Myers und Alpert (1968) aufgefaßt werden:

- Bei mehreren Apn werden zunächst für jedes Attribut Einschätzungen der Wichtigkeit (z.B. fünfstufige Ratingskala) erhoben. Dann gewinnt man für jedes Attribut Einschätzungen darüber, wie sehr sich die Objekte des empirischen Objektraums auf dem jeweiligen Attribut unterscheiden (z.B. fünfstufige Ratingskala). Das gemittelte Produkt der Wichtigkeits- und Unterschiedseinschätzungen stellt den *Determinanzwert eines Attribut* dar. Attribute, die eine definierte Schwelle überschreiten, sind determinante Attribute. Die Vorgehensweise entspricht bis dahin dem Prinzip der dualen Fragetechnik.
- Aus den determinanten Attributen wird eine Matrix erstellt. Die Apn haben nun die Aufgabe, eine Beurteilung der Bedeutungsähnlichkeit jeweils zweier Attribute vorzunehmen. Jede Matrixzelle erhält so einen *Wert der Bedeutungsähnlichkeit* durch die Mittelung über die Apn. Überschreiten zwei Attribute eine definierte Ähnlichkeitsschwelle, wird das Attribut mit dem geringeren Determinanzwert aus der Menge der determinanten Attribute eliminiert.

Ist die nichtredundante Menge der determinanten Attribute spezifiziert, muß in einem *dritten Schritt* überprüft werden, ob die Attribute empirisch unabhängig sind. Eine spezielle Technik hierzu ist mir nicht bekannt. Es ist jedoch denkbar, daß einer Auswahl von Experten eine systematisch aus den determinanten Attributen kombinierte Menge von Objekten vorgelegt wird und diese in Einzel- oder Gruppenurteilen auf ihre empirische Unabhängigkeit einschätzen. Die Objekte können dabei aus zwei oder mehreren Attributen bestehen. Sind die Attribute empirisch nicht unabhängig, müssen Attribute aus der Attributmenge eliminiert werden.

Der *vierte Schritt* besteht in der Überprüfung der Präferenzunabhängigkeit der verbleibenden Attributmenge. Hierzu können Expertenurteile dienen. Eine stringenter Überprüfungsmaßnahme kann in einer Voruntersuchung durch die Anwendung spezieller Tests für Nichtadditivität (z.B. Tukey, 1949), die Prüfung der Axiome des CM (Nygren, 1986) oder nach dem Ansatz des FM erfolgen (Anderson, 1981, 1982; Lerman & Louviere, 1978; Lynch, 1985). Diese Überprüfungsmaßnahmen sind jedoch aufwendiger als Expertenurteile und nur unter gewissen Bedingungen einsetzbar. Einfacher ist die Durchführung einer CA-Voruntersuchung und die Begutachtung des 'fits' eines additiven PSM. Der Modellfit ist eine notwendige Bedingung für Präferenzunabhängigkeit. Ist er hoch, kann davon ausgegangen werden, daß die Attribute weitgehend präferenzunabhängig sind. Es gibt jedoch keine allgemeingültigen Regeln, so daß von Fall zu Fall entschieden werden muß, ob die Präferenzunabhängigkeit der Attribute genügt. Bei pragmatisch orientierten Untersuchungen, bei welchen nicht die Abbildung der realen Beurteilungsstrukturen sondern die Vorhersage von Präferenzen im Vordergrund stehen, ist eine nicht zu ausgeprägte Präferenzabhängigkeit der Attribute unproblematisch.

Wird eine ausgeprägte Präferenzabhängigkeit zwischen Attributen festgestellt, können bestimmte Attribute eliminiert werden. Eine andere Möglichkeit besteht darin, die fraglichen Attribute zu 'Superattributen' zusammenzufassen. Die Stufen dieser Attribute sind dann Kombinationen aus den Stufen der ursprünglichen präferenzabhängigen Attribute.

Hat man die Attributmenge auf Präferenzunabhängigkeit geprüft und eventuelle Maßnahmen zur Herstellung derselben ergriffen, verbleibt die Menge der relevanten Attribute. Der letzte und *fünfte Schritt* besteht in der Bildung der maßgeblichen Attribute aus der Menge der relevanten Attribute. Die maßgeb-

lichen Attribute werden aus Sicht des Programmentscheiders festgelegt. Sie kann, in Übereinstimmung mit den Unternehmens- und Untersuchungszielen, per Expertenurteil in Einzel- oder Gruppenurteilen gefällt werden. Der Grad der Dezimierung der relevanten Attribute hängt bei diesem Schritt auch weitgehend von Überlegungen zur Forschungsökonomie ab (z.B. Aufwand der Untersuchung, Zumutbarkeit der Interviewlänge, Auskunftsmotivation der Apn). Allgemeingültige Verfahrensvorschläge können hier nicht gegeben werden. Die Menge der tatsächlich verwendeten maßgeblichen Attribute einer Untersuchung wird auch als Attributset bezeichnet.

Über die Anzahl der Attribute eines Attributsets können keine allgemeingültigen Aussagen getroffen werden. Sie sollten im allgemeinen nicht über 6-12 liegen. Schubert (1991) fand in seiner Untersuchung zu den Merkmalen durchgeführter CA-Studien folgendes: 18,7% der Studien verwendeten weniger als 4 Attribute, 29,0% 4 Attribute, 38,2% zwischen 5 und 7 Attributen, 9,4% zwischen 8 und 10 Attributen und 1,6% der Studien bezogen mehr als 10 Attribute ein. Der Median der Anzahl der verwendeten Attribute geben Cattin & Wittink (1982) mit 6 bis 7 an. Wittink und Cattin (1989) beziffern diesen auf 8 Attribute.

In ihrer Studie zur Bewertung von Discountkarten untersuchen Green und Wind (1973) die Stabilität der Parameterschätzungen über das Attributset. Hierzu wurde eine CA mit fünf Attributen und eine CA mit einer Teilmenge aus dreien dieser Attribute durchgeführt. Die Ergebnisse deuten auf eine hohe Stabilität der Parameter hin. Vor einer vorschnellen Verallgemeinerung dieses Befundes muß aber gewarnt werden.

Schweickl (1985) hat eine Methode entwickelt, bei der die Befragung der Apn nur mit den jeweils individuell wichtigen Merkmalen durchgeführt wird. Auch bei der adaptiven Conjoint Analyse (ACA) mit dem Programm ACA 4.0 (Sawtooth Software, 1994) ist es möglich, den dekompositorischen Teil der Befragung nicht mit allen, sondern mit den individuell wichtigsten Attributen durchführen. Die Wichtigkeit der Attribute wird vorab über eine kompositorische Methode der Nutzenmessung erhoben.

3.2.5 Repräsentation der Attribute

Sind die maßgeblichen Attribute einer CA bestimmt, dann stellt sich die Frage, in welcher Form man diese Attribute den Apn zur Beurteilung vorlegt. Diese Frage betrifft zum einen die Art der Objektpräsentation in z.B. verbaler, schriftlicher

oder bildlicher Form, die in Abschnitt 3.5.4 diskutiert wird, zum anderen die Art der Objektrepräsentation, mit deren Möglichkeiten sich dieser Abschnitt beschäftigt.

In seiner Theorie des Konsumentenbewertungsprozesses neuer Produktkonzepte unterscheidet Finn (1985) die objektive Repräsentation eines Attributs von der subjektiven Repräsentation. Finn versteht unter objektiven Repräsentationen z.B. die konkrete Beschreibung physikalischer Eigenschaften. Ein Beispiel bei Kraftfahrzeugen wäre die Beschreibung 'Motorleistung 115 PS'. Ein Hauptvorteil dieser Repräsentationsform besteht darin, daß Programmmentseider nach der Untersuchung relativ einfach Interventionsmaßnahmen ergreifen können, indem die objektiven Merkmale der Objekte verändert werden. Die objektive Repräsentation hat jedoch auch Nachteile. Die objektiv repräsentierte Eigenschaft 'Motorleistung 115 PS' ist für die Apn für sich genommen noch nicht nutzenstiftend. Erst die mit ihr verbundenen Eigenschaften, wie z.B. 'gute Beschleunigung beim Überholen', 'Höchstgeschwindigkeit 200 km/h' oder 'Haftpflichtversicherungsbeitrag y', sind nutzenstiftend. Bei einer objektiven Repräsentation des Attributs überläßt man es den Apn, aus den Attributen nutzenstiftende Eigenschaften abzuleiten. Diese Inferenzen können aber zwischen den Apn beträchtlich variieren. Aus diesem Grunde kann man in Frage stellen, ob objektive Eigenschaften tatsächlich von den Apn gleich wahrgenommen werden.

Alternativ zur objektiven Repräsentation der Attribute besteht auch die Möglichkeit der subjektiven Repräsentation. Hier werden den Apn direkt subjektive Interpretationsmöglichkeiten der objektiven Eigenschaften vorgegeben (z.B. 'Ein Auto mit einer eleganten Form'). Es läßt sich derzeit keine allgemeingültige Aussage darüber ableiten, welche Art der Attributrepräsentation ähnlichere Vorstellung bei den Apn hervorrufen. Aus Sicht des Programmmentseiders dagegen können Untersuchungsergebnisse, die mit objektiven Repräsentationen arbeiten, zweifellos besser in Managemententscheidungen umgesetzt werden.

Eine andere Typologie von Objekt- und Attributrepräsentationsmöglichkeiten stammt von Myers & Shocker (1981). Sie unterscheiden die Repräsentation durch 'physical characteristics', 'benefits' und 'imagery'. Die characteristics beziehen sich dabei auf objektive (z.B. physikalisch, chemische) Objekteigenschaften. Als Beispiel kann hier bei Autos die Eigenschaft 'Benzinverbrauch 6l/ 100km' dienen. Mit benefits bezeichnen die Autoren sub-

jektive, funktionale Nutzenerwartungen. Das genannte objektive Merkmal könnte durch 'niedrige laufende Kosten durch niedrigen Benzinverbrauch' in eine benefits-Repräsentation transformiert werden. Mit dem Begriff imagery meinen Myers & Shocker den symbolischen Nutzen, der sich durch die Relation zu sozialen Gruppen ergibt. Der symbolischen Nutzen kann z.B. 'ökologische Bewußtheit beim Fahrzeugkauf' oder noch deutlicher 'ich ordne mich durch meine Fahrzeugwahl der Gruppe ökologisch orientierter Menschen zu' sein. Insbesondere bei Modeartikeln wird der Nutzen der Objekte schwerwiegend erst durch den sozialen Kontext hervorgerufen. Finn (1985) stellt elf Typologisierungsmöglichkeiten der Objekt- und Attributrepräsentation gegenüber und kommt zu dem Schluß, daß die Typologie von Myers & Shocker (1981) die getroffenen Unterscheidungen am besten erfaßt.

Untersuchungen, die nach dem Prinzip der CA vorgehen, können alle drei Möglichkeiten der Objektrepräsentation verwenden. Sieht man veröffentlichte CA-Untersuchungen durch, bemerkt man jedoch die ausgeprägte Tendenz, die Attribute durch objektive Eigenschaften (characteristics) zu repräsentieren (vgl. Rao, 1977; Schubert, 1991; Schweikl, 1985). Manche Autoren sehen die Repräsentation "physikalisch-chemischer Merkmale" sogar als definitorische Eigenschaft einer CA an (Böcker, 1986, S. 564). Schubert (1991) findet in seiner Studie über die Merkmale einer Stichprobe von zwischen 1972 und 1989 durchgeführten CA-Studien, daß 77,3% characteristics-Repräsentationen verwendeten, nur 1,6% arbeiteten mit benefits und bei 18,0% kam eine Kombination der Repräsentationsmöglichkeiten zum Einsatz. Schweikl (1985) stellt fest:

Aus den genannten Gründen werden für die CA i.d.R. nur objektive Produktmerkmale verwendet bzw. von subjektiven Merkmalen nur diejenigen, die einfach in physikalische Merkmale zu übertragen sind. (S. 96)

Die Gültigkeit dieser Aussage kann nicht angezweifelt werden. Fraglich ist jedoch die Haltbarkeit der angesprochenen Gründe. Neben der bereits erwähnten Umsetzbarkeit der CA-Ergebnisse in Managemententscheidungen, wird als Begründung folgendes angeführt:

Es ist nicht gewährleistet, daß sich jede Versuchsperson unter den Ausprägungen des subjektiven Merkmals das gleiche vorstellt. (S. 96)

Diese Gewährleistung ist auch bei der sogenannten objektiven Repräsentation nicht gegeben, da die individuellen Erwägungen der Apn zu den benefits und dem imagery nicht kontrollierbar sind und stark variieren können.

Ob die rein objektive Repräsentation der Attribute sinnvoll ist, muß nach einem neueren Befund von Lefkoff-Hagius und Mason (1993) weiter in Frage gestellt werden. Die Autorinnen untersuchten die Wirkung von Attributen, die in Form von characteristics, benefits oder imagery definiert sind, auf Ähnlichkeits- und Präferenzurteile. Sie kommen zu dem Schluß, daß insbesondere Attribute, die benefits repräsentieren, Präferenzurteile beeinflussen. Hingegen wirken sich besonders characteristics-Attribute, aber auch image-Attribute, auf Ähnlichkeitsurteile aus. Aus diesem Befund folgt: Wenn Attribute durch objektive Eigenschaften repräsentiert sind und Apn Präferenzurteile abgeben sollen, werden diese Eigenschaften weitgehend zunächst in subjektive Nutzen transformiert. Auf der Grundlage dieser Interpretationen werden dann die Präferenzurteile abgegeben. Ob die Interpretationen über die Apn vergleichbar sind, ist jedoch keinesfalls gewährleistet.

Ebenfalls gegen die allgemeine Verwendung von objektiven Attributen argumentieren MacLachlan, Mulhern und Shocker (1988). Ausgehend von der Forderung, die Befragungssituation möglichst ökologisch valide bzgl. realen Beurteilungs- und Entscheidungssituationen zu gestalten, halten MacLachlan et al. benefits-/ costs-Repräsentationen der Attribute für oft besser als die Repräsentation durch characteristics.

Reynolds, Fiedler und Gutman (1985) sprechen sich gegen eine gemischte Repräsentation der Attribute in einer CA aus, um die Vergleichbarkeit der Attribute auf numerischer Ebene zu erhalten.

Unter Berücksichtigung dieser Befunde und Argumente läßt sich zusammenfassend festhalten, daß eine sinnvolle Attributrepräsentation folgende Punkte beachten sollte:

- ❑ Die Attribute und deren Stufen müssen eindeutig definiert sein.
- ❑ Es ist damit zu rechnen, daß Attribute unabhängig von ihrer tatsächlichen Repräsentation weitere Interpretationen bei den Beurteilern hervorrufen. Um die Vergleichbarkeit zwischen den Apn zu erhalten, sollten aus diesem Grunde die Interpretationen vom Untersucher möglichst kontrolliert bzw. über die Apn vergleichbar gemacht werden.

Der ersten Forderung wird z.B. ein Verfahren gerecht, das die Attribute und deren Stufen kurz und unmißverständlich auf der objektiven Ebene definiert. Um der zweiten Forderung genüge zu tun, sollten den Apn zu Beginn eines CA-

Interviews alle Attribute und Attributstufen ausführlich vorgestellt werden. Hierbei ist ein besonderer Wert auf die gewünschte 'Übersetzung' der objektiven Eigenschaften in subjektive Eigenschaften zu legen. Diese Übersetzungen sollten sowohl in Form des utilitarischen Nutzens (benefits) als auch in Form des symbolischen Nutzens (imagery) vor sich gehen. Dies kann der Untersucher beispielsweise über einen ausführlichen Instruktionstext oder einen Film (vgl. Abschnitt 3.5.1) erledigen. Auf diese Art können die Vorteile einer hohen ökologischen Validität der Beurteilungssituation und der Vergleichbarkeit der Apn mit den Vorteilen der kurzen Definition und der objektiven Repräsentation der Attribute und dessen Stufen verbunden werden. Diese kurze Definition ist bei schriftlich vorgegebenen Beurteilungsobjekten, insbesondere bei Einsatz der Vollprofilmethode, unumgänglich.

3.3 Attributstufen

Um die Ausprägung der Objekte auf den Attributen den Apn zur Beurteilung vorzugeben, definiert man auf den Attributen mehrere Attributstufen. Bei der Festlegung der Stufen hat man zunächst folgende Fragen zu beantworten:

1. Wird der Nutzenverlauf eines Attributs als stetig angenommen, dann handelt es sich um ein quantitatives Attribut. Besitzt ein Attribut nur für diskrete Werte einen Nutzen, so spricht man von einem qualitativen Attribut. Bei qualitativen Attributen sind die Stufen meist durch die interessierenden nominalen Ausprägungen vorgegeben (schwarz-weiß Druckfähigkeit - 16 Farben Druckfähigkeit - 256 Farben Druckfähigkeit). Bei qualitativen Attributen interessiert weiter:
2. Wird der Definitionsbereich eines Attributs als begrenzt oder als unbegrenzt angenommen?
3. Wo werden die Endpunkte der Attributstufen gesetzt (Endpunkteproblem)?
4. Wie viele Abstufungen zwischen diesen Endpunkten sollen vorgenommen werden (Abstufungsproblem)?

Die Stufen quantitativer (stetiger) Attribute lassen sich mindestens auf einer Intervallskala darstellen. Ein typisches Beispiel hierfür ist der Preis eines Objekts. Bei qualitativen (diskreten) Attributen ist die Stufendefinition lediglich kategorial (Nominalskala). Qualitative Attribute sind insofern unproblematisch,

als sie wesentlich weniger Freiheitsgrade in ihrer Verwendung bei einer CA zulassen. Es ist nur ein Präferenzmodell denkbar (Teilnutzenwertmodell; Abschnitt 3.4.1) und die Attributstufen sind durch die Beeinflussungsmöglichkeiten des Programmenscheiders begrenzt. Bei quantitativen Attributen dagegen tritt das Endpunkteproblem und das Abstufungsproblem auf.

3.3.1 Endpunkteproblem

In der Regel ist der theoretische Definitionsbereich quantitativer Attribute nicht oder nur nach einer Seite begrenzt. Nimmt man das Attribut 'Breite eines Autos' als Beispiel, dann ist die theoretische Breite nach unten durch die 0 und nach oben nicht begrenzt. Bei der Definition der kleinsten und größten Attributstufe macht es nun wenig Sinn, diesen gesamten Definitionsbereich abzudecken. Wird der oberste Endpunkt mit 'Breite = 50m' definiert, dann ist von vorne herein klar, daß der Nutzen dieser Attributausprägung gleich Null ist. Der Aufwand einer CA wäre nicht zu rechtfertigen, da wesentlich einfachere und ökonomischere Methoden gestatten, diesen Sachverhalt abzubilden. Dasselbe gilt analog für sehr kleine Ausprägungen der Breite. Viel interessanter dagegen ist der Einsatz der CA in einem Definitionsbereich, dessen Nutzenfunktion nicht ohne weiteres bekannt ist.

Mehrere Verfahren sind denkbar, um diesen Bereich aufzufinden. Man kann die Apn einer Voruntersuchung z.B. um uniattributive Nutzenschätzungen auf Ratingskalen bitten. Eine andere, indirekte Methode besteht darin, die Attributausprägungen der tatsächlich vorkommenden Objekte einer Objektklasse, z.B. die am Markt befindlichen Produkte einer Klasse, zu bestimmen und deren Variation als Anhaltspunkt zur Definition der Endpunkte heranzuziehen. Man könnte z.B. die tatsächlich niedrigste und höchste Ausprägung der Produkte als Endstufen verwenden. Diese Endpunkte könnten anschließend, dem aktuellen Forschungsproblem entsprechend, korrigiert werden (bei der geplanten Einführung eines extremen Niedrigpreisprodukts muß z.B. die tatsächlich vorkommende niedrigste Stufe unterschritten werden). Green und Srinivasan (1978) empfehlen aus meßtheoretischen Gründen, die Merkmals-spannweite größer anzulegen, als sie in der Realität vorkommt. Dadurch soll die Genauigkeit der Parameterschätzwerte gesteigert werden.

Nicht nur die Festlegung der Endpunkte innerhalb eines Attributs stellt ein Problem dar, sondern auch die relative Merkmalsspannweite der Attribute im Vergleich mit anderen Attributen. Will man aufgrund der Ergebnisse einer CA Aussagen über die relative Wichtigkeit von Attributen machen, dann sollte die Festlegung der Endpunkte bei den verschiedenen Attributen vergleichbar sein. Eine allgemeingültige Empfehlung zum Umgang mit dem Endpunkteproblem kann nicht gegeben werden. Systematische Untersuchungen dazu gibt es unseres Wissens nicht.

3.3.2 Abstufungsproblem

Sind die Endpunkte eines quantitativen Attributs festgelegt, stellt sich die Frage, ob und wie viele Abstufungen zwischen diesen Endpunkten gewählt werden sollen. Je mehr Stufen man wählt, desto mehr Objekte müssen von den Apn bei Verwendung eines Teilnutzenwertmodells beurteilt werden, da von der Anzahl der Stufen direkt die Anzahl zu schätzender Parameter abhängt. Grundsätzlich kann daraus die Forderung nach möglichst wenigen, aber relevanten Ausprägungen abgeleitet werden.

Acito und Olshavsky (1980) kommen bei ihrer Untersuchung zu dem Ergebnis, daß die kriterienorientierte Validität einer CA höher ist, wenn ein Attribut statt drei Attributstufen nur zwei Stufen hat. Dieser Befund ist durch die geringere Anzahl zu schätzender Parameter bei zweistufigen Attributen erklärbar. Er sollte allerdings nicht generalisiert werden, da denkbar ist, daß der Effekt vom tatsächlichen Nutzenverlauf abhängt. Ist dieser linear, dann dürften zwei Attributstufen bessere Ergebnisse erbringen. Ist er jedoch nicht linear, dann benötigt man mehr Attributstufen, um diese Nichtlinearität fassen zu können. Darüber hinaus hängt es vom eingesetzten Präferenzmodell ab, ob mehr Attributstufen zu einem erhöhten Aufwand bei der Parameterschätzung führen. Wird ein Teilnutzenwertmodell verwendet, dann steigt mit der Anzahl der Stufen auch die Anzahl zu schätzender Parameter, während dies bei einem linearen Vektormodell nicht der Fall ist.

Eine nicht geringe Beachtung im Zusammenhang mit dem Abstufungsproblem sollte dem 'Anzahl-der-Stufen-Effekt' zuteil werden. Der Effekt wurde erstmals von Currim, Weinberg und Wittink (1981) thematisiert (number-of-levels-effect). In dieser Studie ergab sich eine Zunahme der relativen Wichtigkeit

(= Differenz zwischen dem Nutzen der am höchsten und der am niedrigsten präferierten Attributstufe) eines Attributs mit der Anzahl seiner Stufen.

Der Effekt wurde experimentell von Wittink, Krishnamurthi und Nutter (1982) weiter belegt. Diese Autoren fanden z.B., daß die relative Wichtigkeit des Produktpreises um 7% zunahm, wenn das Attribut bei gleichbleibenden Endpunkten auf zwei weiteren Stufen definiert wurde. Der Anzahl-der-Stufen-Effekt tritt bei verschiedenen Varianten der CA auf (Wittink, Krishnamurthi & Reibstein, 1989; Untersuchung bei Trade-off-Ansätzen, Vollprofilrangordnung und Vollprofilpaarvergleich). In einer neueren experimentellen Untersuchung stellen Wittink, Huber, Fiedler und Miller (1991, zit. in Wittink, Huber, Zandan & Johnson, 1992) fest, daß der Effekt bei der Vollprofilmethode etwas ausgeprägter zu sein scheint, als bei der adaptiven Conjoint Analyse (ACA; s. Abschnitt 3.7.2.4). Durch die Steigerung der Attributstufen von zwei auf vier, nahm die relative Wichtigkeit der Attribute bei der Vollprofilmethode um 9-12%, bei der ACA nur um 5-6% zu.

Noch strittig ist die theoretische Erklärung des Effekts. Green & Srinivasan (1990) favorisieren eine Erklärung über Aufmerksamkeitsprozesse. Die Zunahme der Wichtigkeit der Attribute sollten demzufolge durch die gesteigerte Aufmerksamkeit entstehen, die einem oft variierten Attribut zuteil wird. Noch verstärkt werden könnte der Effekt durch Vermutungen der Apn, daß oft variierte Attribute dem Versuchsleiter besonders wichtig sind (demand characteristics; Orne, 1962). Dieser kognitiven Hypothese steht die Vermutung von Currim et al. (1981) entgegen. Diese Autoren favorisieren die Erklärung, daß der Effekt durch die Parameterschätzalgorithmen zustande kommt.

3.4 Präferenzstrukturmodell

In diesem Abschnitt werden wichtige Präferenzstrukturmodelle (PSM) vorgestellt. PSM bestehen aus den Präferenzmodellen der jeweiligen Attribute und dem Präferenzintegrationsmodell. Die Präferenzmodelle werden in Abschnitt 3.4.1 und Präferenzintegrationsmodelle in Abschnitt 3.4.2 besprochen. In Abschnitt 3.4.3 gehen wir dann kurz auf sequentielle Urteilsmodelle ein.

3.4.1 Präferenzmodell

Das Präferenzmodell (synonym: Nutzenfunktion) beschreibt den Zusammenhang zwischen einem Objektmerkmal und dem Nutzen des Merkmals für ein beurteilendes Subjekt. Es ist ein Modell für die beobachtbaren Präferenzen, die eine Ap bzgl. eines Attributs äußert. Wie in Abschnitt 2.2.2 bereits im Zusammenhang mit der Festlegung der verwendeten Notation dargestellt, läßt sich für die Ap h der Nutzen eines Attributs als reellwertige Funktion der Ausprägung eines Attributs folgendermaßen darstellen:

$$u_h(x_j) = f_h(x_{jk}) \quad (3.1)$$

Die Nutzenfunktion ist auf individueller Ebene definiert. Im Zusammenhang mit der CA werden insbesondere drei Nutzenfunktionen diskutiert (Böcker, 1986; Green & Srinivasan, 1978, 1990): Das lineare Vektormodell, das Idealpunktmodell und das Teilnutzenwertmodell.

3.4.1.1 Lineares Vektormodell

Das *lineare Vektormodell* (auch: Idealvektormodell) nimmt einen linearen Zusammenhang zwischen der Ausprägung eines Attributs und dem Nutzen an. Es kann bei quantitativen Attributen eingesetzt werden.

$$u_h(x_j) = b_{h0} + b_{hj} * x_{jk} \quad (3.2)$$

b_{h0} und b_{hj} sind zu schätzende Parameter. Für $b_{hj} > 0$ ($b_{hj} < 0$) gilt: Je höher (niedriger) Ausprägung des Attributs x_j , desto größer ist dessen Teilnutzen. Die Nutzenfunktion ist linear, d.h. die Nutzenänderung je Merkmalsänderung ist konstant. Der Index h gibt an, daß die Parameter der Nutzenfunktion auf individueller Ebene geschätzt werden.

Da die Nutzenwerte b_{hj} nicht ratioskaliert sind, ihr Nullpunkt also beliebig wählbar ist, wird die additive Komponente der Nutzenfunktion in der Regel auf Null gesetzt². Die Anzahl der zu schätzenden Parameter des linearen Vektormodells reduziert sich damit auf einen Parameter je Attribut. Das Modell nimmt die Form

$$u_h(x_j) = b_{hj} * x_j \quad (3.3)$$

an. Dabei ist

$u_h(x_j)$	der Teilnutzen des Attributs x_j für die Ap h .
b_{hj}	zu schätzender Parameter des Attributs x_j bei der Ap h .
x_j	Ausprägung des Attributs j .

3.4.1.2 Idealpunktmodell

Ein weiteres, häufig diskutiertes Modell ist das *Idealpunktmodell*. Es kann bei quantitativen Attributen eingesetzt werden. Bei dieser Modellklasse wird angenommen, daß der Nutzen bei einer bestimmten Attributausprägung maximal ist und mit der Entfernung von diesem Idealpunkt abnimmt. Die Nutzenfunktion läßt sich folgendermaßen darstellen:

$$u_h(x_j) = b_{hj} * |x_j - x_{j*}|^r \quad (3.4)$$

Dabei ist

² Dies trifft auch bei allen weiteren Modellen zu. Aus diesem Grund wird die additive Komponente der Modelle nicht weiter erwähnt.

- $u_h(x_j)$ der Teilnutzen des Attributs x_j für die Ap h.
- b_{hj} zu schätzender Parameter des Attributs x_j bei der Ap h.
- x_j Ausprägung des Attributs j.
- x_{j^*} Idealpunkt des Attributs j.
- r Minkowski-Parameter. Er wird im folgenden auf 2 gesetzt. Damit handelt es sich um gewichtetes euklidisches Idealpunktmodell.

Ein Beispiel für die eventuelle Angemessenheit eines Idealpunktmodells wäre die Präferenzmodellierung bei der Einschätzung der Meerwassertemperatur für einen Badeurlaub, da anzunehmen ist, daß sowohl zu kaltes als auch zu warmes Wasser auf Ablehnung stößt. Im euklidischen Idealpunktmodell ist ein Parameter zu schätzen.

3.4.1.3 Teilnutzenwertmodell

Das *Teilnutzenwertmodell* ist ein allgemeines Modell, welches sowohl das lineare Vektormodell als auch das Idealpunktmodell umfaßt. Dieses Modell kann auch nichtlineare Nutzenverläufe abbilden und ist bei quantitativen und qualitativen Attributen einsetzbar. Es kann formal folgendermaßen dargestellt werden:

$$u_h(x_j) = \sum_k b_{hjk} * x_{ijk} \tag{3.5}$$

Dabei ist

- $u_h(x_j)$ der Teilnutzen des Attributs x_j für die Ap h.
- b_{hjk} Teilnutzenwert der Attributstufe k des Attributs x_j bei der Ap h. b_{hjk} sind die zu schätzenden Modellparameter.
- $x_{ijk} = \begin{cases} 1 & \text{falls ein Objekt } X_i \text{ die Attributausprägung } k \text{ auf Attribut } x_j \text{ besitzt} \\ 0 & \text{in anderen Fällen.} \end{cases}$

Für jedes Attribut x_j sind K_j Parameter zu schätzen. Da die Teilnutzenwerte b_{hjk} nicht ratioskaliert sind, kann ein Parameter auf 0 gesetzt werden. Dies ist in der Regel der Parameter mit dem niedrigsten Wert. Es resultieren beim Teilnutzenwertmodell $K_j - 1$ zu schätzende Parameter b_{hjk} . Die Modellkomponente x_{ijk} ist eine Dummy-Variable. Die K_j Stufen des Attributs x_j werden durch die $K_j - 1$ Dummy-Variablen repräsentiert. Der Flexibilität des Modells steht die vergleichsweise hohe Anzahl an zu schätzenden Parametern gegenüber.

Formal ist die Teilnutzenwertfunktion nur für die diskreten Attributstufen definiert. Bei quantitativen Attributen werden die Definitionslücken in der Regel linear interpoliert.

3.4.1.4 Anwendungen und Fazit

Innerhalb eines PSM ist natürlich ein gemischter Einsatz der Präferenzmodelle möglich. Neben den erwähnten Präferenzmodellen wurden z.B. auch quadratische Vektormodelle (Pekelman & Sen, 1979a, 1979b) vorgeschlagen. Bei diesen geht die Attributausprägung quadriert in die Nutzenfunktion ein.

Über die differentielle Häufigkeit des Einsatzes der verschiedenen Präferenzmodelle gibt es unseres Wissens keine Untersuchungen. Cattin und Wittink (1982) behaupten zwar "The most common model used by the survey respondents is the part worth model." (S. 46), legen aber keine Häufigkeitsverteilung vor. Aufgrund der weiten Verbreitung des Programms ACA (Sawtooth Software, 1985, 1994) bei neueren kommerziellen CA-Untersuchungen (Wittink & Cattin, 1989; Wittink et al., 1994), ist aber zu vermuten, daß das Teilnutzenwertmodell die anderen Modelle weit dominiert, da ACA 4.0 nur die Spezifikation eines Teilnutzenwertmodells zuläßt.

Fraglich ist, ob die standardmäßige Anwendung des Teilnutzenwertmodells sinnvoll ist. Pekelman und Sen (1979b) erhalten eine höhere kriterienorientierte Validität der PSM bei Verwendung von Vektor- und Idealpunktmodellen statt Teilnutzenwertmodellen. Dieser Vorteil ergibt sich aus der niedrigeren Anzahl zu schätzender Parameter bei den erstgenannten Modellen. Allgemein kann sicherlich das Fazit gezogen werden, daß sich aus Sicht der kriterienorientierten Validität der Einsatz der flexibleren Teilnutzenwertmodelle nur lohnt, wenn tatsächlich mit Nutzenverläufen gerechnet werden muß, die durch einfachere Präferenzmodelle nur ungenügend abgebildet werden können. Besteht das Ziel einer CA jedoch nicht nur in der Maximierung der prädiktiven Validität, sondern z.B. auch in der Bestimmung von präferenzwirksamen Schwellen (Preisschwellen o.ä.), dann sind Teilnutzenwertmodelle unverzichtbar.

3.4.2 Präferenzintegrationsmodell

Präferenzmodelle beschreiben den Zusammenhang zwischen der Attributausprägung eines Objekts und dem Teilnutzen des jeweiligen Attributs. Da die

Objekte aus mehreren Attributen bestehen, stellt sich die Frage, wie die Teilnutzen der Attribute zu einem Gesamtnutzenwert kombiniert werden. Das Präferenzintegrationsmodell (synonym: Nutzenintegrationsfunktion) beschreibt diesen Zusammenhang. Man kann die Funktion darstellen mit

$$u_h(X_i) = f(u_h(x_1), u_h(x_2), \dots, u_h(x_j), \dots, u_h(x_J)) \quad (3.6)$$

In der Literatur zur Entscheidungstheorie und zur CA werden eine Reihe von Integrationsmodellen diskutiert.

3.4.2.1 Kompensatorische Präferenzintegrationsmodelle

Nahezu alle Ansätze der Präferenzstrukturmodellierung verwenden kompensatorische Präferenzintegrationsmodelle. Unter diesen wiederum dominiert das additive Integrationsmodell. Allen kompensatorischen Integrationsmodellen ist gemeinsam, daß die Teilnutzenwerte eines Attributs durch andere Attribute kompensiert werden können. Diese Annahme wird bei fast allen Ansätzen der Präferenzstrukturmodellierung getroffen. Sie ist dann angemessen, wenn die Feinbeurteilung konkurrierender Objekte im Zentrum der Aufmerksamkeit steht.

Multiplikative Integrationsmodelle sind bei der kompensatorischen Modellbildung selten. Multiplikative Modelle sind additiven Modellen formal äquivalent, da die Modelle sich durch eine strukturerhaltende logarithmische Transformation ineinander überführen lassen. Sie haben die Form

$$u_h(X_i) = \square u_h(x_{ij}). \quad (3.7)$$

Polynomiale Präferenzintegrationsmodelle betrachten sowohl Haupteffekte als auch Wechselwirkungen zwischen den Attributen. Beispiele für diese Art der Modelle sind z.B. bei Green (1973), Green und Devita (1974), Green und Devita (1975) und Thomas (1983) zu finden. Es soll beispielhaft das polynomiale Integrationsmodell von Green und Devita (1975) dargestellt werden. Es umfaßt die Haupteffekte und Zweifachwechselwirkungen zwischen den Attributen. Formal dargestellt hat es die Form

$$u_h(X_i) = \square u_h(x_{ij}) + \square u_h(x_{ij_1}) * u_h(x_{ij_2}) \quad (3.8)$$

Polynomiale Präferenzmodelle können komplexe Bewertungsprozesse abbilden. Damit verbunden ist jedoch auch die Notwendigkeit der Schätzung relativ vieler Modellparameter. Green (1984) konnte empirisch zeigen, daß Präferenzintegrationsmodelle, die Interaktionsterme beinhalten, oft eine niedrigere kriterienorientierte Validität haben als einfache additive Modelle. Die höhere Realitätsangepaßtheit der Modelle durch Berücksichtigung von Attributwechselwirkungen ist offensichtlich meist gering im Vergleich zur Verminderung der Vorhersagegenauigkeit durch die Hinzunahme zusätzlicher Parameter.

Die bekannte Robustheit additiver Modelle, die keine Wechselwirkungen betrachten, und die schwierigere Handhabbarkeit polynomialer Integrationsmodelle haben dazu geführt, daß in den Jahren 1986 - 1991 diese Art der Modelle bei der Präferenzstrukturmodellierung in nur ca. 10% der Anwendungen eingesetzt wurden (Wittink et al., 1994).

Additive Präferenzintegrationsmodelle kommen im Rahmen der CA weit überproportional zum Einsatz. Die Gründe dürften in der einfachen Verfügbarkeit von Software zur Schätzung der Parameter dieser Modelle, der einfachen Verständlichkeit und der damit verbundenen einfachen Ergebnisinterpretierbarkeit solcher Untersuchungen liegen, die additive Modelle verwenden. Diese Modelle haben die Form

$$u_h(X_i) = \sum u_h(x_{ij}). \quad (3.9)$$

Der Gesamtnutzen eines Objekts setzt sich additiv aus der Summe der Teilnutzen der Attribute zusammen, die das Objekt konstituieren.

Faßt man das Präferenzmodell und das Präferenzintegrationsmodell zusammen, entsteht das *Präferenzstrukturmodell (PSM)*. Das am häufigsten verwendete PSM bei der CA ist das additive Teilnutzenwertmodell. Es hat die Form

$$u_h(X_i) = \sum b_{hjk} * x_{ijk} \quad (3.10)$$

mit

$u_h(X_i)$ (Gesamt-)Nutzen des Objekts X_i bei der Ap h .

b_{hjk} Teilnutzenwert der Attributstufe k des Attributs x_j .

$$x_{ijk} = \begin{cases} 1 & \text{falls Objekt } X_i \text{ die Attributausprägung } k \text{ auf Attribut } x_j \text{ besitzt} \\ 0 & \text{in anderen Fällen.} \end{cases}$$

Die in diesem Modell zu schätzenden Parameter b_{hjk} werden auch Teilnutzenwerte (TNW) genannt.

3.4.2.2 *Paramorphe Repräsentation*

In der entscheidungstheoretischen Literatur ist seit einiger Zeit bekannt, daß linear-additive Modelle bei der Vorhersage von Kriteriumsvariablen aus mehreren Prädiktoren leistungsfähig sind (Meehl, 1954). Diese Arbeiten stellten das Verhältnis von statistischer und klinischer Urteilsbildung in das Zentrum ihrer Aufmerksamkeit und fanden im allgemeinen, daß linear-additive Modelle hohe statistische 'fits' zeigten. Eine weitere Steigerung dieser Modellanpassung durch Wechselwirkungskomponenten und nichtlineare Zusammenhänge scheint kaum des Aufwandes wert zu sein, sofern das Ziel der Modellierung die Vorhersage ist (Holbrook, 1981, Slovic, Fischhoff & Lichtenstein, 1977). Dies ist auch dann der Fall, wenn die Struktur der Modelle der empirischen Zusammenhangsstruktur eigentlich unangemessen sind. Hoffman (1960) nannte solche Modelle eine 'paramorphe Repräsentation'. In ihrer Überblicksarbeit kommen Dawes und Corrigan (1974) ebenfalls zu dem Schluß, daß paramorphe Repräsentationen zur Vorhersage überaus erfolgreich sind (vgl. auch Dawes, 1982).

Ähnliche Befunde erbrachten auch Forschungsarbeiten zur CA. Zusammenfassend kann festgehalten werden, daß das additive kompensatorische Präferenzintegrationsmodell nahezu alle Entscheidungsprozesse gut approximiert und die Entscheidungsergebnisse sehr gut abbilden kann (z.B. Carmone, Green & Jain, 1978; Lynch, 1985; Olshavsky & Acito, 1980). Dies gilt insbesondere dann, wenn flexible Teilnutzenwertpräferenzmodelle die Attributnutzen modellieren (Green & Srinivasan, 1978).

Hagerty (1986) untersuchte empirisch, ob Interaktions-PSM genauere Vorhersagen zulassen, als additive Haupteffektmodelle, weil sie den realen Urteilsprozeß angemessener abbilden könnten. Dieser möglichen besseren Angemessenheit steht jedoch der Nachteil entgegen, daß bei komplexen Modellen mehr Parameter zu schätzen sind. Hagerty kommt zu folgenden Befunden: Wird das PSM, wie bei der CA üblich, auf der individuellen Ebene

gebildet, und das vorhergesagte Kriterium ebenfalls auf der individuellen Ebene gemessen, dann sind additive Haupteffektmodelle in ihrer Vorhersagefähigkeit polynomialen Haupteffekt-Wechselwirkungsmodellen überlegen (vgl. auch Akaah & Korgaonkar, 1983; Anderson & Donthu, 1988; Elrod, Louviere & Davey, 1992). Der Grund liegt in der geringeren Anzahl zu schätzender Parameter bei den einfacheren Haupteffektmodellen. Dieser Befund dreht sich aber um, wenn das Kriterium auf aggregierter Ebene erhoben wird. Wird die Vorhersagefähigkeit der Modelle daran gemessen, wie gut sie in der Lage sind, Wahlanteile vorherzusagen, dann zeigen sich Modelle überlegen, die Haupteffekte und Wechselwirkungen modellieren. Ist der Schluß von Hagerty (1986), daß Haupteffekt-Wechselwirkungsmodelle bei der Vorhersage aggregierter Kriterien eine höhere Validität besitzen, generalisierbar? Dies bleibt zumindest fraglich, da Anderson und Donthu (1988) eine Überlegenheit (oder zumindest keine Unterlegenheit) von einfachen Haupteffektmodellen bei der Vorhersage von individuellen und aggregierten Kriterien finden. Diese Forschungsfrage ist demnach keinesfalls abgeschlossen.

In einer Simulationsstudie prüft Teichert (1994), inwieweit additive Modelle eine simulierte 'wahre' Nutzenstruktur abbilden können. Er verwendet wahre Strukturen mit und ohne Attributwechselwirkungen. Erwartungsgemäß sind die geschätzten TNW stärker fehlerhaft, wenn die wahre Struktur Interaktionen enthält. Im Mittel weichen die geschätzten Nutzenwerte in dieser Bedingung 48,2% von den wahren Werten ab, während die Abweichung ohne Interaktion nur 15,0% beträgt. Ähnliche Fehlerniveaus finden sich jedoch auch bei wahren interaktionsfreien Strukturen, deren TNW sehr inhomogen sind (53,8%). Eine genauere Betrachtung der Bedingungen, wann additive Haupteffektmodelle eine gute Anpassung zeigen, unternahmen Carmone und Green (1981). Sie kommen zu dem Schluß, daß additive Haupteffektmodelle sehr robust gegen Verletzungen ihrer Annahmen sind. Sie zeigen in dieser Simulation auch bei Vorliegen von Wechselwirkungen noch gute Modellfits. Dies ist insbesondere dann der Fall, wenn lediglich Wechselwirkungen vom Typ 'non-cross-over' auftreten (s. auch Green, 1973). Bei dieser Art von Wechselwirkungen bestehen zwar Interaktionen zwischen den Attributen, die ordinale Ordnung der Präferenzen innerhalb eines Attributs ist jedoch nicht von anderen Attributen abhängig.

3.4.2.3 *Strukturangemessene Repräsentation*

Modellfit und Kriterienvolidität sind zwar notwendige aber keine hinreichende Bedingungen dafür, daß ein geschätztes Modell auch empirische Strukturen realitätsgerecht abbildet. Strukturangemessenheit der Modelle spielt in vielen Bereichen der Marketingforschung oft eine nur untergeordnete Rolle, solange die Vorhersagbarkeit von Entscheidungen und Präferenzen gegeben ist. Aber auch innerhalb des Marketing und auf anderen Forschungsgebieten gibt es eine Reihe von Fragestellungen, innerhalb derer die Entdeckung der 'wahren' Beurteilungs- und Entscheidungsstruktur wichtig ist. Dies gilt bei allen Grundlagenforschungsansätzen und bei einigen angewandten Forschungsbemühungen, wie z.B. bei der Marktsegmentierung auf der Basis von Präferenzstrukturen.

Lerman und Louviere (1978) schlagen vor, daß vor der Parameterschätzung eines angenommenen PSM durch die CA eine Prüfung der Angemessenheit des Modells anhand einer kleinen, nicht repräsentativen Stichprobe mit Methoden des FM (s. Abschnitt 2.4.3) erfolgen sollte. Anschließend können dann die Parameter an einer repräsentativen Stichprobe durch die CA geschätzt werden. Bei dieser 'großen Untersuchung' stellt man an das Design und das Skalenniveau der Antwortvariablen geringere Anforderungen. Auf diese Weise soll die Strukturangemessenheit und Validität der PSM mit vertretbarem Aufwand erreicht werden.

Anwendungen von Techniken der Informationsintegrationstheorie sind bei der Präferenzstrukturmodellierung im Bereich der Marketingforschung eher selten. Beispiele hierfür sind Lerman und Louviere (1978), Timmermans (1982) und Troutman und Shanteau (1976).

Ein ähnlicher Vorschlag des stufenweisen Vorgehens stammt von Olshavsky und Acito (1980). Entsprechend diesen Autoren kann es sinnvoll sein, vor der tatsächlichen CA zunächst unter Laborbedingungen und mit einer kleinen Stichprobe das tatsächliche Integrationsmodell möglichst genau zu bestimmen. Sie empfehlen hierzu insbesondere Prozeßverfolgungstechniken durch eine Protokollanalyse. Die CA sollte zumindest teilweise durch Protokollanalysen überprüft werden. Dies scheint insbesondere deswegen sinnvoll, da vermutlich die Entscheidungsregeln durch die CA-Prozedur beeinflußt werden. Denkbare Einflußfaktoren sind die Anzahl der Objekte, der Attribute und der Attributstufen, der Erhebungsmodus der Antwortvariablen etc. Ein Fernziel der Forschung zur CA könnte darin bestehen, die Informationsverarbeitungsregeln

der Interviewsituation mit denen der echten Entscheidungssituation durch die Manipulation dieser Einflußfaktoren zur Deckung zu bringen (Bettman, 1979; Olshavsky & Acito, 1980).

Auch Nickerson, McClelland und Petersen (1990) sprechen sich für die Prüfung der Angemessenheit der Integrationsmodelle aus. Sie verwenden hierfür die Prüfung einiger Axiome des CM (vgl. Abschnitt 2.3.3). Sie prüfen in ihrer Untersuchung die einfache und gemeinsame Unabhängigkeit und das Axiom der doppelten Kürzbarkeit (vgl. Krantz et al., 1971). Die einfache Unabhängigkeit nimmt bei additiven Integrationsmodellen an, daß die Präferenzrangordnung der Stufen innerhalb eines Attributs invariant über alle Kombinationen mit allen anderen Attributen ist. Die gemeinsame Unabhängigkeit bedeutet, daß die Ordnung von Attributstufenkombinationen als unabhängig von der Verknüpfung der Kombination mit anderen Attributen angenommen wird. Die Annahme der doppelten Kürzbarkeit kann nicht so einfach verständlich beschrieben werden, ist jedoch in aller Regel nicht verletzt, wenn die Unabhängigkeitsannahmen halten. Es gibt eine Reihe von Computerprogrammen (vgl. Nickerson et al., 1990), die eine Axiomprüfung vornehmen können (CONJOINT, Holt & Wallsten, 1974; PCJM, Ullrich & Cummins, 1973; PCJM2, Ullrich, Cummins & Walkenbach, 1978; CPCJM, Ullrich & Wilson, 1990; COMESCAL, Lukas, 1985; CMSCAL, Nygren, 1986).

In der Forschung und bei Anwendungen der CA spielt die Prüfung der Strukturangemessenheit eine untergeordnete Rolle. In aller Regel werden additive Teilnutzenwertmodelle ungeprüft eingesetzt. Bei angewandten CA-Untersuchungen werden in den letzten Jahren hauptsächlich die Software ACA (Sawtooth Software; 42%), CONJOINT (SPSS) und CONJOINT ANALYZER (Bretton-Clark) eingesetzt (Wittink et al., 1994). Bei ACA und CONJOINT ANALYZER sind nur Teilnutzenwertmodelle spezifizierbar, während CONJOINT auch die Spezifikation eines linearen Vektormodells zuläßt. ACA und CONJOINT beschränken sich auf additive Haupteffektmodelle. Wechselwirkungen sind nur über Umwege modellierbar. CONJOINT ANALYZER läßt auch die Aufnahme von Interaktionen durch die Spezifikation geeigneter Kreuzproduktterme zu. Allerdings muß zur Schätzung der Interaktionseffekte auch ein geeignetes Design vorgegeben werden. Bretton-Clarks Software zur Bildung reduzierter Designs (CONJOINT DESIGNER) kann diese Aufgabe nicht

leisten. Allen Softwarepaketen ist gemeinsam, daß sie über keine Prozeduren zur direkten Prüfung der Angemessenheit des geschätzten PSM verfügen.

3.4.2.4 *Fazit*

Steht bei der multiattributiven Präferenzstrukturmodellierung die Schätzung eines vorhersagekräftigen PSM im Vordergrund, sind additive Haupteffektmodelle meist ausreichend. Ihre Ergebnisse sind einfach zu interpretieren und es sind leistungsfähige Softwaresysteme verfügbar, die bei der Planung, Durchführung und Datenanalyse unterstützen. Der Forscher muß sich jedoch vergegenwärtigen, daß er, trotz einer eventuell hohen Validität der geschätzten Modelle, paramorphe Repräsentationen geschaffen haben könnte.

Wittink et al. (1994) geben eine Anzahl von Forschungsgebieten an, in welchen ihrer Meinung nach weitere Forschung zur CA geboten ist. Hierzu zählt auch die Modellierung von Interaktionen. Da bei dieser Arbeit der pragmatische Aspekt der kriterienorientierten Vorhersage im Vordergrund steht, beschränken wir uns jedoch auf additive Modelle.

3.4.3 **Prozeßmodell**

Lussier und Olshavsky (1979) gehen davon aus, daß nur bei einfachen Wahl-/ Beurteilungsaufgaben kompensatorische Urteilsprozesse angewandt werden. Sind die Aufgaben komplexer, kommen andere Urteilsprozesse zum Einsatz. Einige Ansätze der Präferenzmodellierung nehmen an, daß vor einer Feinbeurteilung der Objekte eine Grobbeurteilung vorgenommen wird (Böcker, 1986). Zunächst werden nicht akzeptable Objekte durch nichtkompensatorische Urteilsprozesse von der eingehenderen Betrachtung ausgeschlossen. Danach erfolgt erst die aufwendige Feinbeurteilung der Objekte durch kompensatorische Urteilsprozesse. Modelle, die einen solchen Entscheidungsprozeß abbilden, versuchen sowohl der ersten Stufe der Ausscheidung von Alternativen durch nichtkompensatorische Regeln als auch der zweite Stufe, die durch kompensatorische 'trade-offs' zwischen den multiattributiven Alternativen gekennzeichnet ist, gerecht zu werden.

Srinivasan (1988) wendet diese Denkweise in seinem kompositorischen Ansatz der Präferenzstrukturmodellierung an. Sein konjunktiv-kompensatorisches Modell fordert die Ap zunächst auf, nicht akzeptierbare Attribut-

stufen zu kennzeichnen. Diese Attributstufen definieren sich dadurch, daß sie ein Objekt so unattraktiv machen, daß es immer zurückgewiesen würde, ungeachtet seiner Eigenschaften auf allen anderen Attributen. Srinivasan (1988) fand in seiner Studie zur Berufswahl mit MBA-Studenten, daß keine der 54 Apn eine Arbeitsstelle antrat, die ein Merkmal besaß, das sie Monate zuvor als nicht akzeptierbar eingestuft hatte. In Einklang mit diesem Befund berichten Bucklin und Srinivasan (1991), daß weniger als 1% des Verkaufsvolumens auf Kaffee-Produkte mit nicht akzeptierbaren Attributstufen entfielen.

Aber nicht alle Forschungsbefunde rechtfertigen die Angemessenheit eines konjunktiv-kompensatorischen Urteilsprozesses. Klein (1986) stellt erstens in Frage, ob Apn zwischen nicht akzeptierbaren Attributstufen (im strengen Sinne ihrer Definition) und unerwünschten Attributstufen (Stufen mit einem niedrigen Nutzen) unterscheiden können. Zweitens bezweifelt sie, daß selbst bei einer möglichen positiven Antwort auf die erste Frage die kriterienorientierte Validität konjunktiv-kompensatorischer Modelle gegenüber einem einfachen additiven, kompensatorischen Modell ansteigt. Klein fand, daß 15% ihrer Apn insofern inkonsistent antworteten, als sie im Verlauf der Untersuchung Objekte mit nicht akzeptierbaren Attributstufen gegenüber vollständig akzeptierbaren Objekten bevorzugten. Bezüglich der Validität stellte sie fest, daß beide Modelle äquivalente Vorhersagen und Ergebnisse erbrachten. Sie zieht den Schluß, daß die Elimination von nicht akzeptierbaren Attributstufen zwar keine Validitätssteigerung hervorrufen, aber geeignet sind, um die Datensammlung zu vereinfachen, da der Attributraum dadurch verringert wird. Allerdings empfiehlt Klein, daß die Elimination konservativ erfolgen sollte, indem die Apn genau über das Konstrukt der Nichtakzeptierbarkeit instruiert werden und vor der Entscheidung alle Stufen der anderen Attribute kennenlernen. Auch Olshavsky und Acito (1980) fanden keinen signifikanten Validitätsunterschied zwischen ausgewählten nichtkompensatorischen und kompensatorischen Modellen.

Green, Krieger und Bansal (1988) behaupten, daß Attributstufen auch dann kompensatorisch verarbeitet werden, wenn Apn diese als nicht akzeptierbar bezeichnen. Ihnen haftet lediglich ein geringer Teilnutzen an. Zum selben Schluß kommen auch Dorsch und Teas (1992). Sie finden in ihrer vergleichenden Untersuchung zwischen der konjunktiv-kompensatorischen expliziten Modellbildung nach Srinivasan (1988) und traditionellen CA-Ansätzen

eine beträchtliche Inkongruenz dieser beiden Verfahren, was die nicht akzeptierbaren Attributstufen betrifft.

Das Programm ACA 4.0 (Sawtooth Software, 1994) greift diese Forschungsbefunde auf, indem die Abfrage nicht akzeptierbarer Attributstufen vor der Modellbildung optional möglich ist. Dadurch wird der Attributraum der nachfolgenden Parameterschätzung gemindert. Die Modellierung erfolgt jedoch durch ein kompensatorisches additives Teilnutzenwertmodell. Die nicht akzeptierbaren Attributstufen erhalten per Voreinstellung einen niedrigen Teilnutzenwert. Die Höhe des Werts ist vom Untersucher beeinflussbar.

3.5 Beurteilungsobjekte

In diesem Abschnitt werden die Möglichkeiten der Konstruktion der Beurteilungsobjekte besprochen, die die Apn zu bearbeiten haben. Diese betreffen a) die Art der Definition der Objekte auf den Attributen (Abschnitt 3.5.2). Insbesondere die Vollprofilmethode und der Trade-off-Ansatz sind bei der CA weit verbreitet. Des weiteren bestimmt das Untersuchungsdesign, b) welche Beurteilungsobjekte den Apn vorgegeben werden (Abschnitt 3.5.3). Darüber hinaus stellt sich dem Untersucher das Problem, c) auf welche Art und Weise die Objekte den Apn präsentiert werden können (Abschnitt 3.5.4).

Bevor jedoch mit der eigentlichen Befragung begonnen wird, sollten verschiedene 'Aufwärmaufgaben' durchgeführt werden.

3.5.1 Aufwärmphase

Um die ökologische Validität (vgl. Abschnitt 3.1) einer conjointanalytischen Untersuchung zu steigern, sollte vor dem eigentlichen Befragungsbeginn die Apn sorgfältig mit dem Objektbereich, den möglichen realen Entscheidungssituationen und den Attributen vertraut gemacht werden. Der Teil einer Untersuchung, in dem dies geleistet wird, bezeichnen wir als Aufwärmphase. Hierbei ist besonderer Wert auf die Wahl des Untersuchungssettings (vgl. Abschnitt 3.1.2) und die Auswahl und Entwicklung der Apn (vgl. Abschnitt 3.1.3) zu legen.

Bei der Wahl des Untersuchungssettings sollte man zunächst die relevante reale Urteils- und Entscheidungssituation identifizieren und diese hinsichtlich ihrer Spezifität einschätzen. Hierbei können die vier Aspekte der Spezifität von

Einstellungs- und Verhaltensmaßen hilfreich sein (vgl. Abschnitt 3.1.2; Art der Handlung, Art des Ziels, Kontextcharakteristika, zeitliche Gegebenheiten; Ajzen & Fishbein, 1977). Das Untersuchungssetting sollte eine möglichst hohe Spezifitätskorrespondenz zwischen der Untersuchung und der realen Urteils- und Entscheidungssituation herstellen.

Die Apn sollten so ausgewählt bzw. trainiert werden, daß sie realen Entscheidungen möglichst entsprechen. Dies betrifft sowohl ihre den Objektbereich betreffenden Fähigkeiten und Kenntnisse als auch ihre Urteils- und Entscheidungsmotivation. In aller Regel wird es notwendig sein, die Auskunftsmotivation der Apn in der Untersuchung zu steigern, um sie der Motivation bei realen Entscheidungen anzugleichen. Die Entscheidungsmotivation wird durch mehrere Faktoren moderiert (Slovic et al., 1988). Das Involvement, die Wichtigkeit der Entscheidungskonsequenzen und die wahrgenommene Entscheidungskompetenz wirken sich auf die Motivation besonders aus. Die reale Entscheidungssituation und das Untersuchungssetting sollten daher hinsichtlich dieser Faktoren vergleichbar sein.

Um das Untersuchungssetting zu beeinflussen, sind mehrere Techniken denkbar. Die am weitesten verbreitete Technik ist sicherlich die verbale Instruktion. Diese kann schriftlich erfolgen, auch Präsentationen oder Filme können hilfreich sein. Die Objektbereichskenntnisse der Apn sollten durch Vorstellung der Attribute und deren Stufen vertieft werden.

Eine interessante Hypothese formulieren in diesem Zusammenhang Huber, Wittink, Fiedler und Miller (1993). Sie behaupten, daß eine kompositorische Urteilsaufgabe vor einer dekompositorischen Nutzenmessung als Aufwärmübung angesehen werden kann. Diese Form der Aufwärmphase kann zu einer Validitätssteigerung der dekompositorischen CA führen. Huber et al. (1993) belegen diese Vermutung zunächst mit einer Zusammenstellung empirischer Untersuchungen. In Studien, die die Validität von kompositorischen und dekompositorischen Ansätzen der Nutzenschätzung vergleichen, belegen Akaah und Korgaonkar (1983), Cattin et al. (1982) und Green, Goldberg und Wiley (1983) eine höhere Validität der CA im Vergleich zu kompositorischen expliziten Modellbildungsansätzen. Zum umgekehrten Schluß kommen dagegen Green, Krieger und Agarwal (1991) und Wright und Kriewal (1980), während Leigh, McKay und Summers (1984) keinen Validitätsunterschied zwischen einem expliziten Modellbildungsverfahren und der CA finden. Ein Unterschied zwischen

diesen Studien besteht darin, daß bei den drei erstgenannten Untersuchungen der kompositorische Teil vor der CA erledigt wurde, während bei den letzteren die Reihenfolge umgekehrt war. In einem eigenen Experiment kontrollierten Huber et al. (1993) diese Reihenfolge. Die Ergebnisse zeigen eine Überlegenheit von ACA³ über eine Vollprofil-CA im Sinne der kriterienorientierten Validität. Dieser Effekt verschwindet jedoch, wenn auch vor der Vollprofilmethode eine explizite Modellbildung als Aufwärmübung zum Einsatz kommt. Diese Befunde veranlassen die Autoren zu dem Schluß, daß Aufwärmphasen conjointanalytische Nutzenschätzungen verbessern. Eine mögliche Vorgehensweise besteht in der Verwendung von expliziten Modellbildungsaufgaben als Aufwärmübung.

Zusammenfassend kann man sagen, daß bisherige Forschungsbe-mühungen der Aufwärmphase bei CA nur wenig Beachtung geschenkt haben, obwohl die Ausgestaltung dieses Untersuchungsteils wahrscheinlich einen nicht unerheblichen Einfluß auf die Ergebnisse der CA ausüben.

³ Wie in Abschnitt 3.7 noch näher besprochen wird, verwendet ACA eine kompositorische Nutzenschätzung vor dem eigentlichen conjointanalytischen Befragungsteil.

3.5.2 Definition der Beurteilungsobjekte

Dekompositorische Ansätze der multiattributiven Präferenzstrukturmodellierung schätzen die Modellparameter auf der Basis von Urteilen oder Entscheidungen der Apn über Objekte, die auf mehreren Attributen spezifiziert sind. Es gibt verschiedene conjointanalytische Erhebungsformen, die sich dadurch unterscheiden, *auf wie vielen Attributen* ein Beurteilungsobjekt definiert ist. Am weitesten verbreitet sind die Vollprofilmethode, der Trade-off-Ansatz und in jüngerer Zeit die ACA-Teilprofilmethode.

3.5.2.1 Vollprofilmethode

Der Vollprofilansatz wird auch häufig 'traditionelle CA' genannt. Bei dieser Erhebungsmethode werden den Apn Objekte zur Beurteilung vorgelegt, die auf allen maßgeblichen Attributen definiert sind ('Konzeptbeurteilung'). Die Begründung für diese ganzheitlich-simultane Objektdefinition ist die, daß den Apn eine möglichst realistische Vorstellung von den Objekten gegeben wird. Darüber hinaus ist eine bild- oder modellhafte Darstellung (vgl. Abschnitt 3.5.4) meist nur bei Anwendung der Vollprofilmethode möglich.

Ein übliche Darbietungsart von Vollprofilobjekten ist die schriftliche Beschreibung eines Objekts auf je einer Karte. Meist haben die Apn die Aufgabe, diese Karten gemäß ihrer Präferenz in eine Rangordnung zu bringen (vgl. Abschnitt 3.5.4 und 3.6.2.2).

Die Vollprofilmethode besitzt jedoch auch einige Nachteile. Diese beruhen hauptsächlich auf der kognitiven Überlastung, die entsteht, wenn eine größere Anzahl von Attributen in eine CA eingehen ('information overload'; z.B. Jacoby, 1977; Payne, Bettman & Johnson, 1992). Motivationale Bedenken dagegen befürchten, daß in realen Beurteilungs- und Entscheidungssituationen viel mehr Aufwand bei der Informationsverarbeitung betrieben wird als in der Untersuchungssituation. Aus diesem Grunde könnten die Informationsverarbeitungsergebnisse der Untersuchungssituation und der realen Situation nur bedingt miteinander vergleichbar sein. Aufgrund der kognitiven und der motivationalen Überlegungen ist damit zu rechnen, daß Apn vereinfachende Strategien bei der Urteilsbildung anwenden, wie z.B. die ausschließliche Beachtung einer Teilmenge der Attribute. Diese Teilmenge enthält entweder die individuell wichtig-

sten Attribute, oder sie wird von subtileren Faktoren bestimmt, wie z.B. der seriellen Position eines Attributs auf der Karte (Acito, 1977; Chapman & Bolton, 1985; Johnson, 1981). Green und Srinivasan (1978) sehen aus den genannten Gründen eine obere Grenze der Attributanzahl beim Vollprofilansatz bei fünf bis sechs.

3.5.2.2 Teilprofilmethoden

Aufgrund der Nachteile des Vollprofilansatzes wurden Teilprofilansätze entwickelt. Bei Teilprofilansätzen sind die Objekte nur auf einer Teilmenge der Attribute spezifiziert. Der Grundgedanke liegt darin, daß die Apn nur eine Reihe einfacher Urteile bilden, indem sie bei jedem Urteil nur wenige Attribute beachten müssen. Aus den geschätzten Modellen können dann komplexe Gesamturteile vorhergesagt werden, wenn Menschen mit einer tatsächlichen Entscheidungssituation konfrontiert werden, bei der mehr Attribute gleichzeitig beurteilt werden müssen. In diesen Situationen sind sie motiviert, in die Beurteilung Zeit zu investieren und alle Attribute der Beurteilungsobjekte gegeneinander abzuwägen. Daß dies in einer Interviewsituation nicht der Fall ist, sollte nicht weiter verwunderlich sein. Auch aus diesem Grunde findet der Vollprofilansatz bei wenigen Attributen seine Grenzen.

Die Nachteile des Teilprofilansatzes liegt darin, daß meist nur eine schriftliche oder verbale Objektbeschreibung möglich ist (Abschnitt 3.5.4) und nur bestimmte Erhebungsformen der abhängigen Variablen (Abschnitt 3.6.2) in Frage kommen. Darüber hinaus wirkt die Beurteilungsaufgabe auf die Apn oft künstlich ('wie finden Sie folgendes Auto: rote Farbe - Benzinverbrauch 7l/ 100 km') und es muß die Annahme gemacht werden, daß die Apn keine Inferenzen von den spezifizierten Attributen auf die weggelassenen Attribute bilden.

Die bekannteste Teilprofilmethode ist der Trade-off-Ansatz (two-factor-at-a-time procedure) von R.M. Johnson (1974). Bei dieser Methode werden den Apn

$$\text{Anzahl der Trade-off-Matrizen} = J * (J-1) / 2 \quad (3.16)$$

sogenannte Trade-off-Matrizen zur Beurteilung vorgegeben. Jede Matrix besteht aus K_{j1} Zeilen und K_{j2} Spalten. Auf diese Weise wird jedes Attribut mit jedem der anderen Attribute einmal in einer Trade-off-Matrix kombiniert. Die Ap hat die Aufgabe, die Zellen der Matrix, welche jeweils ein Objekt auf zwei Attributen definieren, in eine Rangreihe zu bringen. Dies geschieht durch einfaches

Ausfüllen der Matrix mit Zahlen: Das höchstpräferierte Objekt bekommt eine 1, das am wenigsten präferierte Objekt wird mit einer ($K_{j_1} * K_{j_2}$) gekennzeichnet. Auf der Basis dieser Daten werden die Parameter des PSM geschätzt (s. Abschnitt 3.8).

Da immer nur zwei Attribute gleichzeitig in die Urteile der Apn eingehen, gibt es keine kognitive Überlastung beim Trade-off-Ansatz. Die Anzahl der Attribute ist lediglich durch die Interviewzeit beschränkt. Diese nimmt jedoch überproportional mit der Attributanzahl zu. Darüber hinaus dürfte die Auskunftsmotivation beim Ausfüllen der Trade-off-Matrizen sehr schnell kritisch werden. Dies könnte zu einem Antwortverhalten nach einem stereotypen Muster führen.

Beim Trade-off-Ansatz ist besonders wichtig, daß die Attribute voneinander präferenzunabhängig sind (R.M. Johnson, 1974). Was die Verzerrung der Beurteilungsdaten durch Inferenzen über fehlende Attributinformation betrifft, ist die Befundlage nicht ganz einheitlich. R.D. Johnson (1987a) findet in seinem Experiment Verzerrungseffekte durch fehlende Information und erklärt diese theoretisch. Darüber hinaus führt er eine Reihe von Aufsätzen an, die seine Befunde bekräftigen. Ähnliche Ergebnisse berichten Huber und McCann (1982). In einer Replikation und Erweiterung dieses Experiments kommen jedoch Lim, Olshavsky und Kim (1988) zu dem Schluß, daß Inferenzeffekte meist nur dann auftreten, wenn die Inferenzschlüsse explizit induziert werden.

3.5.2.3 Vergleich Vollprofilmethode und Trade-off-Ansatz

Die Vollprofilmethode und der Trade-off-Ansatz wurden in einigen Studien bezüglich ihrer Meßgütekriterien verglichen. Für eine Überlegenheit des Vollprofilansatzes plädiert z.B. Safizadeh (1989). In seiner Studie zur Beurteilung von Krankenhäusern (6 Attribute), in dem die Datensammlungsmethode (Vollprofil vs. Trade-off) innerhalb der Apn experimentell manipuliert wurde, zeigte der Vollprofilansatz eine deutliche höhere kriterienorientierte Validität hinsichtlich der Vorhersage von Holdout-Objekten (dabei handelt es sich um Objekte, die von den Apn zusätzlich beurteilt werden; eine ausführliche Beschreibung dieser Technik findet sich in Abschnitt 5.4).

Eine weitere Studie, die die Hypothese von der Überlegenheit der Vollprofilmethode stützt, stammt von Segal (1982). Er untersuchte die Retest-

Reliabilität der beiden Methoden (zwei Maße: Test-Retest-Korrelation der abhängigen Variablen und der geschätzten TNW; Test-Retest-Intervall: 7-10 Tage) und kommt zu dem Schluß, daß beide Methoden eine hohe Reliabilität besitzen. Die Vollprofilmethode schneidet in dieser Untersuchung mit 3 Attributen etwas besser ab.

Jain, Acito, Malhotra & Mahajan (1979) fanden keinen Unterschied zwischen den beiden Datensammlungsmethoden (5 Attribute) bezüglich der kriterienorientierten Validität. Sie empfehlen, die Wahl der Datensammlungsmethode von den Gegebenheiten der Studie abhängig zu machen (z.B. Bevorzugung der Trade-off-Methode bei Fragebogen, die versendet werden). Oppedijk van Veen und Beazley (1977) ermitteln eine hohe Übereinstimmung der beiden Methoden in ihrer Untersuchung mit 3 Attributen.

Das Bild bleibt jedoch nicht ganz so einheitlich, wenn die Ergebnisse zweier weiterer Untersuchungen mitberücksichtigt werden. Alpert, Betak und Golden (1978, zit. in Green & Srinivasan, 1978) und Montgomery, Wittink und Glaze (1977, zit. in Green & Srinivasan, 1978) berichten von einer Überlegenheit der Trade-off-Methode. Diese bezieht sich bei Alpert et al. auf Maße der Modellanpassung (goodness-of-fit; vgl. Abschnitt 5.3) bei 9 Attributen. In der Studie von Montgomery et al. (1977) zeigte die Trade-off-Methode eine höhere prognostische Validität (8 Attribute).

Zusammenfassend läßt sich festhalten, daß beide Datensammlungsmethoden vergleichbar sind. Die Vollprofilmethode zeigt bei einer Attributanzahl von bis zu 6 eine leichte Überlegenheit. Diese könnte dadurch erklärt werden, daß die Apn eine realistischere Aufgabe durchzuführen haben. Steigt die Attributzahl darüber hinaus, zeigen Teilprofilansätze eine höhere Meßgüte. Obwohl die Beurteilungsaufgabe bei Teilprofilansätzen nicht mehr ganz so realistisch ist, hat diese Methode offensichtlich Vorteile, weil sie Überlastungen der Informationsverarbeitungskapazität vermeidet.

Wie häufig kommen die Vollprofilmethode, der Trade-off-Ansatz und andere Methoden der Datensammlung bei kommerziellen CA zum Einsatz? In den USA benutzen zwischen 1971-1980 56% den Vollprofilansatz, 27% den Trade-off-Ansatz und 14% eine Kombination der beiden Methoden (Cattin & Wittink, 1982). Diese Bevorzugung für die Vollprofilmethode verstärkte sich noch in den Jahren 1981-1985. 61% verwendeten einen Vollprofil-Konzeptbeurteilungsansatz und nur 6% Trade-off-Matrizen. Bei 10% ist eine Kom-

bination zu finden (Wittink & Cattin, 1989). Noch deutlicher fällt die Überlegenheit der Vollprofilmethode hinsichtlich ihrer Anwendungshäufigkeit in der Studie von Schubert (1991) aus. Zwischen 1972 und 1989 verwendeten über 91% die Vollprofilmethode, während auf den Trade-off-Ansatz nur 5,5% entfielen. Eine häufigere Verwendung der Vollprofilmethode ist auch bei der Studie über die kommerziellen Anwendung der CA in Europa zwischen 1986-1991 zu finden (Wittink et al., 1994). Die Vollprofilbeurteilung kann aber nur noch 25% auf sich vereinen, während die Verwendung von Trade-off-Matrizen in 15% der Fälle angegeben wird. Der größte Teil der CA-Studien wurde mit dem ACA-Design (42%, vgl. Abschnitt 3.7.2.4) durchgeführt. Dies ist auf die einfache Handhabbarkeit dieses Programms zurückzuführen und darauf, daß auch die Befragung computergestützt durchgeführt werden kann. ACA verwendet eine besondere Teilprofilmethode (vgl. Abschnitt 3.7.2.4). Man kann aus diesem Grund behaupten, daß Teilprofilmethoden die Vollprofilmethode zunehmend verdrängt. Es ist zu vermuten, daß dieser Trend bei zukünftigen kommerziellen Anwendungen anhält und die traditionelle Vollprofil-CA seltener werden wird.

3.5.2.4 Fazit

Vollprofilansätze und Teilprofilansätze der CA erbringen, zumindest hinsichtlich der kriterienorientierten Validität, bei einer mittleren Attributanzahl (5-6) keine großen Unterschiede. Ist das Attributset einer Untersuchung klein, dann sollte der Vorzug dem Vollprofilansatz gegeben werden. Die Befragten nehmen diese Aufgabe als realistischer wahr, und die Vorgehensweise ist mindestens so valide wie alternative Datensammlungsmethoden. Übersteigt die Anzahl der Attribute die Zahl 6, so sollte unbedingt mit Teilprofilmethoden gearbeitet werden. Überlegungen sowohl zur kognitiven Leistungsfähigkeit des Menschen als auch zu motivationalen Gegebenheiten legen dies nahe.

3.5.3 Konstruktion des Objektsets

Durch die Definition der Attribute ergibt sich die Menge der theoretisch konstruierbaren Beurteilungsobjekte. Beim Vollprofilansatz besteht diese Menge aus

$$\text{Anzahl der konstruierbaren Objekte} = \prod K_j \quad (3.17)$$

Objekten (K_j =Anzahl der Attributstufen des Attributs j). Werden alle diese Objekte in einer CA den Apn vorgegeben, dann spricht man von einem *vollfaktoriellen Design*. Vollfaktorielle Designs werden jedoch selten verwendet, weil die Menge der konstruierbaren Beurteilungsobjekte mit der Anzahl der Attribute/-stufen überproportional anwächst. Um die Auskunftsmotivation der Apn nicht ernsthaft zu gefährden, sollte die Anzahl der zu beurteilenden Objekte beim Konzeptbeurteilungsansatz nicht mehr als 30 betragen (Green & Srinivasan, 1978). Da diese Grenze sehr schnell erreicht ist, arbeiten fast alle CA mit reduzierten Designs. Diese Designs stellen eine sinnvolle Teilmenge des vollfaktoriellen Designs dar. Sie lassen die Schätzung bestimmter Modellparameter zu. Die Menge der Objekte, die den Apn zur Beurteilung innerhalb eines Designs vorgelegt wird, bezeichnet man als Objektset.

Die wichtigsten Designs sind das orthogonale Haupteffektdesign (main effect designs), Zufallsdesigns und Paarvergleichsdesigns. Die Frage, welches Design einsetzbar ist, ist nicht unabhängig von der Datensammlungsmethode (Vollprofilansatz, Teilprofilansätze) und der Art der Erhebung der abhängigen Variablen (z.B. Rating, Rangordnung, Paarvergleich).

3.5.3.1 Systematisch reduzierte Designs

Orthogonale Haupteffektdesigns (OHD) kommen insbesondere zum Einsatz, wenn auf Vollprofilkonzepte die Präferenz durch Rangordnung oder Rating erhoben wird. Die Attribute eines OHD sind unkorreliert. Das Design läßt die nicht konfundierte Schätzung aller Attributhaupteffekte bei additiven Präferenzintegrationsmodellen zu.

Ist die Anzahl der Stufen bei allen Attributen gleich, dann spricht man von einem symmetrischen OHD. Eine bekannte Form ist das Lateinische Quadrat (bei drei Attributen mit $K_1=K_2=K_3$ Stufen). Die Konstruktionsrichtlinie dieses symmetrischen OHD findet sich z.B. bei Glaser (1978). Allgemeine Konstruktionsregeln für symmetrische und asymmetrische OHD sind bei Addelman (1962a, 1962b) zu finden. Weitere Konstruktionsregeln für OHD und für reduzierte Designs, die die Schätzung von Haupteffekten oder Haupteffekten mit ausgesuchten Wechselwirkungen zulassen, können bei Green (1974), Green, Carroll und Carmone (1978), Hahn und Shapiro (1966, zit. in Teichert, 1994) eingesehen werden. Wird das Design sehr stark reduziert, so kann dies zu einer

Verminderung der Reliabilität der geschätzten TNW führen (Reibstein, Bateson & Boulding, 1988).

Die Konstruktion systematisch reduzierter OHD ist relativ aufwendig. Seit einiger Zeit ist Software verfügbar, die annähernd orthogonale OHD automatisch erstellen (z.B. die SPSS-Prozedur ORTHOPLAN oder Bretton-Clarks CONJOINT DESIGNER).

Bei der Trade-off-Methode kann die Anzahl der auszufüllenden Matrizen ebenfalls reduziert werden. Dies kann z.B. durch die Verwendung partiell balancierter unvollständiger Blockdesigns (David, 1988; Green, 1974) erfolgen.

3.5.3.2 Zufallsdesigns

Green und Srinivasan (1978) schlagen die Möglichkeit der Bildung von Zufallsdesigns vor. Diese Art der Designs kann auch Korrelationen des empirischen Objektraums abbilden. Unseres Erachtens wurde diese Art der Designbildung jedoch nie systematisch untersucht. Auch eine Anwendung ist mir nicht bekannt. Eine vertiefte Diskussion soll hier deswegen unterbleiben. Green und Srinivasan vermuten, daß systematisch reduzierte Designs validere Ergebnisse erbringen als Zufallsdesigns, sofern die Attributinterkorrelationen des empirischen Objektraums gering sind.

3.5.3.3 Paarvergleichsdifferenzendesigns

Wird die abhängige Variable durch Paarvergleiche erhoben (vgl. Abschnitt 3.6.2.4), dann gibt es zwei grundsätzliche Möglichkeiten, die Modellparameter zu schätzen. Die erste Vorgehensweise besteht darin, zunächst die Paarvergleiche in eine Rangreihe zu transformieren und diese dann mit Hilfe monotoner Regressionsmethoden auszuwerten. Da die Anzahl der notwendigen Paarvergleichsurteile überproportional mit der Anzahl der Objekte anwächst, ist die Verwendung von reduzierten Designs schon bei wenigen Attributen notwendig. Für diese Art der Vorgehensweise können bestimmte reduzierte Designs, z.B. partiell balancierte unvollständige Blockdesigns (David, 1988; Green, 1974), eingesetzt werden.

Die andere Vorgehensweise ist die direkte Schätzung der Parameter aus den Paarvergleichen. Die Basis für die Schätzung sind in diesem Fall nicht die Präferenzwerte der einzelnen Objekte, sondern die Präferenzdifferenzen jeweils

zweier Objekte. Designs dieser Art nennen Hausruckinger und Herker (1992) *Differenzendesigns*. Die Autoren zeigen ausführlich, wie man von unkorrelierten Ausgangsdesigns (z.B. Addelman, 1962a) zu unkorrelierten Differenzendesigns kommen kann. Die vorgeschlagene Konstruktionsregel hat den Vorteil, daß die Anzahl der erforderlichen Paarvergleiche gegenüber herkömmlichen reduzierten Paarvergleichsdesigns deutlich niedriger ist. Darüber hinaus erlauben diese Designs die Anwendung des Tests von Hauser und Shugan (1980). Der Test überprüft die Antworten der Probanden auf Inkonsistenzen. Hierzu wird bei speziellen Paartripeln geprüft, ob die erhobenen Antworten als mindestens intervallskalierte Präferenzdifferenzen aufgefaßt werden können. Die Details des Tests sind anderswo beschrieben (Hauser & Shugan, 1980; Hausruckinger & Herker, 1992; Schweikl, 1985).

Die besondere Art der Differenzendesignbildung des Programms ACA 4.0 (Sawtooth Software, 1994; vgl. Abschnitt 3.7.2.4) ist derzeit von besonderer Bedeutung. Diese resultiert insbesondere aus der weitverbreiteten (und vermutlich noch zunehmenden) Anwendung des Programms und aus der Tatsache, daß die Konstruktion des Designs adaptiv erfolgt. Dies bedeutet, daß die Paare, die die Apn zu beurteilen haben, nicht bei Interviewbeginn festgesetzt sind, sondern diese vielmehr von den jeweils vorausgegangenen Antworten abhängen.

Der genaue Designkonstruktionsalgorithmus der Paarvergleichsphase des Programms ist leider nicht veröffentlicht. Die folgende Darstellung richtet sich nach den im Handbuch (Sawtooth Software, 1994) beschriebenen Konstruktionsprinzipien und nach eigenen Erfahrungen mit dem Programm.

Zunächst hat der Untersucher die Möglichkeit, die Anzahl der paarweise dargestellten Konzepte im Paarvergleichsteil innerhalb bestimmter Grenzen zu bestimmen. Die maximale Anzahl der Paarvergleiche ist durch das Programm auf

$$\text{Anzahl der Paarvergleiche} = 3 * \left(\left(\boxed{} \right) K_j - J - 1 \right) - \boxed{} K_j \quad (3.18)$$

begrenzt. Darüber hinaus ist die Anzahl der Attribute der Beurteilungsobjekte (Profilstärke) einflußbar. Die Profilstärke kann zwischen zwei und fünf Attributen variieren. Die Definition der Objekte ist nur in zunehmender Profilstärke möglich. Im Handbuch wird vorgeschlagen, in der Regel mit Zweierprofilen zu beginnen, dann nach einigen Fragen Dreierprofile zu verwenden usw.

Die Programmautoren sind bestrebt, das ACA-Design möglichst balanciert zu gestalten. Um dies zu leisten, hat man drei Prinzipien in den Designalgorithmus eingebaut:

1. Die Attribute gehen möglichst gleich häufig in die Objekte ein.
2. Die Attributstufen werden innerhalb der Attribute möglichst gleich häufig verwendet.
3. Die Spalten der Designmatrix (diese repräsentieren die Attributstufen) sollen möglichst orthogonal sein.

Bis hier ist die Designwahl nicht adaptiv. Die adaptive Komponente besteht aus drei Prinzipien, die bei der Objektwahl berücksichtigt werden:

4. Besteht die CA aus vielen Attributen, dann kann man in die Paarvergleichsphase nur eine Teilmenge der Attribute aufnehmen. Die Anzahl der aufzunehmenden Attribute ist vom Untersucher beeinflussbar. Ist die Menge der gewählten Attribute kleiner, als die Menge aller maßgeblichen (Untersuchungs-)Attribute, dann werden in dieser Phase nur die individuell wichtigsten Attribute verwendet. Die Wichtigkeitseinschätzungen stammen aus vorhergehenden Phasen von ACA.
5. Bestehen Attribute aus mehr als fünf Stufen, dann werden nur die fünf Stufen mit dem höchsten Teilnutzen in das Differenzendesign aufgenommen. Die Teilnutzenwertschätzungen stammen wieder aus vorhergehenden Phasen von ACA.
6. Es wird, nachdem die Attribute und die möglichen Stufen der jeweiligen Paarvergleichsfrage festgelegt sind (siehe Punkt 1. und 2.), diejenigen Attributstufen gewählt, die einen möglichst gleichen Nutzen der Objekte zur Folge haben. Ausschlaggebend für diesen Nutzenvergleich sind die initialen Nutzenschätzungen in der kompositorischen Phase von ACA bzw. das jeweilige 'update' dieser Schätzungen, das aufgrund der Antwort auf jede Paarvergleichsfrage berechnet wird (vgl. Abschnitt 3.7.2.4).

Dieser letzte Punkt ist 'das eigentlich adaptive' der Differenzendesignwahl bei ACA. Die Überlegung, die zu dieser Vorgehensweise führt, ist wie folgt: Da vor jeder Paarvergleichsfrage bereits (wenn auch grobe) Nutzenschätzungen vorliegen, sollten einer Ap keine Frage gestellt werden, deren Antwortrichtung deutlich vorhergesagt werden kann. Dies wäre z.B. dann der Fall, wenn ein Objekt das andere dominieren würde (Was bevorzugen Sie? Ein billiges, gutes

Auto oder ein teures, schlechtes Auto?). Es besteht mehr Informationsbedarf bei Objekten, die so konstruiert sind, daß ihr Nutzen sehr ähnlich ist. Die Antworten der Apn dienen zur Feinabstimmung der Nutzenschätzung.

Da das geschätzte PSM die Nutzenstruktur der Apn mit zunehmender Interviewdauer immer besser abbilden sollte, stellt das Programm im Laufe des Interviews immer 'schwierigere' Fragen. Dies führt bei den Apn oft zum Eindruck, daß das System sensibel reagiert. Dies sollte, verglichen mit Papier- und Bleistift-Erhebungsmethoden, zu einer höheren Auskunftsmotivation seitens der Befragten führen (Agarwal, 1988b). Dieser Effekt ist jedoch eher anekdotisch überliefert, als durch kontrollierte Untersuchungen.

Schwierig zu beurteilen ist derzeit, inwieweit der Designalgorithmus tatsächlich zu orthogonalen Designs führt. Falls Abweichungen von der Orthogonalität auftreten, wäre von besonderem Forschungsinteresse, in welchem Maße diese Konfundierung die Parameterschätzung beeinflusst. Des Weiteren ist noch unklar, ob die Darbietung von sehr ähnlichen Objekten in der Paarvergleichsphase im Laufe des Interviews nicht zu einer Änderung der psychologischen Qualität der Antwortskala (abgestufter Paarvergleich) führt. Von der statistischen Perspektive aus betrachtet, ist der ACA-Designalgorithmus sicherlich problematisch. Dennoch fanden Huber und Hansen (1986), die experimentell die Nutzenähnlichkeit der Objektpaare manipulierten, eine höhere kriterienorientierte Validität bei Verwendung des ACA-Designs (hohe Nutzenähnlichkeit der Paare) gegenüber einem Design, das kein Ähnlichkeitskriterium berücksichtigte. Sie kommen zu dem Schluß, daß die psychologischen Vorteile der ACA-Vorgehensweise die statistischen Nachteile mehr als ausgleichen.

Bei der Wahl der *Profilstärke der Objekte* besteht bei ACA ein Zielkonflikt. Sowohl die Möglichkeit der Nutzenbalancierung jeweils eines Objektpaares als auch die statistische Effizienz des Verfahrens bei der Parameterschätzung nimmt mit der Objektprofilstärke zu. Aus diesem Grund sind möglichst starke Profile (4 oder 5 Attribute) wünschenswert (statistische Perspektive). Andererseits aber haben starke Profile den Nachteil, daß die Paarvergleichsaufgabe für die Apn kompliziert wird. Dies könnte nicht nur zu einer kognitiven Überforderung und damit zu einer Fehlersteigerung führen. Auch die raschere Abnahme der Antwortmotivation ist zu erwarten. Aus diesen Gründen sind Profile einer nur geringen Stärke (2 Attribute) wünschenswert (psychologische Perspektive).

Dieser Sachverhalt wird noch komplizierter, wenn man davon ausgeht, daß die Antwortzeit mit der Objektprofilstärke zunimmt. Bei konstanter Interviewzeit könnten die Apn mehr Beurteilungen von Objekten mit geringer Profilstärke vornehmen. Dies aber führt zu einer relativen Zunahme der verfügbaren Datenpunkte bei der Parameterschätzung, was zu einer genaueren Schätzung führen sollte. Dieser Sachverhalt schwächt Teile der statistischen Argumente für starke Objektprofile ab. Huber und Hansen (1986) fanden eine Zunahme der Interviewzeit bei 16 Paarvergleichen von Dreier- und Viererprofilen (12,9 Minuten und 13,1 Minuten) im Vergleich zu Objekten der Profilstärke zwei (11,4 Minuten). Darüber hinaus wurde die Aufgabe bei Zweierprofilen als angenehmer empfunden.

In einer Monte-Carlo-Simulationsstudie erhielt Johnson (1987b, zit. in Agarwal, 1989) nur eine sehr gering bessere Wiedererkennung des simulierenden 'wahren PSM', wenn mehr als zwei Attribute je Paarvergleichsobjekt dargeboten wurden. Es gibt auch eine Reihe von empirischen Untersuchungen, die sich der Frage nach der optimalen Wahl der Profilstärke hinsichtlich der kriterienorientierten Validität von ACA widmen. In einer Studie zu Berufswahlpräferenzen bestimmt Agarwal (1989) die kriterienorientierte Validität der geschätzten PSM (Validität zu Holdout-Vollkonzeptrankings und -ratings; Erste-Wahl-Treffer auf individueller Ebene und Wahlanteile auf aggregierter Ebene [Maximum-utility Modell]: Eine ausführliche Erklärung dieser Validierungskonzepte findet sich in Abschnitt 5.4). In der Paarvergleichsphase manipuliert er experimentell die Profilstärke der Beurteilungsobjekte (2 vs. 4). Er kommt zu dem Schluß, daß Zweierprofile den Viererprofilen hinsichtlich der kriterienorientierten Validität auf aggregierter Ebene überlegen sind. Er berichtet von keinem Unterschied, wenn das Kriterium auf individueller Ebene liegt.

Huber und Hansen (1986) führten ein Experiment durch, in welchem die Präferenzen von MBA-Studenten hinsichtlich Studentenwohnungen mit ACA modelliert wurden. Sie variierten die Objektprofilstärke im Paarvergleichsteil von zwei bis vier. Obwohl sie keinen Unterschied bei der kriterienorientierten Validität zwischen diesen Versuchsbedingungen feststellten, empfehlen die Autoren, nur Objekte der Profilstärke 2 oder maximal 3 zu verwenden. Die Begründung liegt sowohl in der besseren Modellanpassung der Paarvergleichsantworten (Modellfit R^2 ; vgl. Abschnitt 5.3) als auch in den oben bereits genannten Vorteilen hinsichtlich der Interviewzeit und einigen Einstellungs-

indikatoren gegenüber der Befragung. Keinen signifikanten Validitätsunterschied zwischen Zweier- und Dreierprofilen erbrachten PSM-Schätzungen mit ACA in einer Studie von Agarwal (1988c, zit. in Agarwal, 1989).

Eine weitere wichtige Frage beim ACA-Design ist die, wie groß die *Anzahl der Paarvergleiche* in der Paarvergleichsphase gehalten werden soll. Per Voreinstellung ist diese Anzahl nach oben begrenzt (Gleichung 3.18). Die Begrenzung ist so gelegt, daß die Gesamtzahl der Fragen im kompositorischen Teil der expliziten Modellbildung und im dekompositorischen Paarvergleichsteil dreimal so hoch ist, wie die Anzahl zu schätzender Parameter. Auch bei dieser Frage muß ein Kompromiß zwischen der statistisch optimalen Vorgehensweise (sehr viele Paarvergleiche) und der psychologisch angemessenen Anzahl geschlossen werden. Da mit der Interviewdauer die Auskunftsmotivation der Apn nachläßt, ist zu vermuten, daß ab einem bestimmten Punkt die Validität des geschätzten PSM abnimmt.

Die Befunde zur Anzahl der Paarvergleiche sind in Abschnitt 3.7.2.4 dargestellt. Es läßt sich vorläufig und zusammenfassend sagen, daß nicht zu viele Paarvergleiche bei ACA eingesetzt werden sollten. Agarwal (1989) gibt als Faustregel das 1,5-fache der Anzahl zu schätzender Parameter an. Darüber hinaus ist eine Wechselwirkung zwischen der Anzahl der Paarvergleiche und der Profilstärke der Beurteilungsobjekte zu erwarten. Agarwal (1988b) belegt diese empirisch, indem er bei Verwendung von Zweierprofilen einen positiven Zusammenhang zwischen der Anzahl der Paarvergleiche und der kriterienorientierten Validität zeigte. Dies war jedoch bei Dreierprofilen nicht der Fall. Das Ergebnis deutet darauf hin, daß die Auskunftsmotivation der Apn bei der komplexeren Aufgabe früher nachläßt.

3.5.3.4 *Differenzendesigns: Fazit*

Zusammenfassend kann festgehalten werden, daß die psychologischen Vorteile von Paarvergleichsdesigns mit geringen Profilstärken (2, höchstens 3) die statistischen Vorteile höherer Profilstärken mehr als ausgleichen. Ähnlich verhält es sich mit Designs, die Alternativen mit sehr ähnlichem Nutzen zum Paarvergleich anbieten. Huber und Hansen (1986) führen hierzu aus:

Both [...] strategies, minimizing affective differences [Alternativen mit ähnlichem Nutzen; Anm. d. Verf.] and decreasing dimensional complexity [geringe Objektprofilstärken; Anm. d. Verf.], have an immediate liability of lessening the

statistical power of the conjoint design, and thus are suboptimal from a statistical perspective. However, both strategies have important positive effects on the respondents that appear to result in either an increase in predictive power or greater speed and a more positive attitude on the part of the respondents. This suggests that the search for optimal conjoint designs may need to focus on the psychological impact of the questions in addition to their strict statistical properties. (S. 163)

Die Frage nach der optimalen Anzahl der Paarvergleiche kann als noch nicht geklärt angesehen werden. Sie steht aber in jedem Falle in Wechselwirkung mit der Qualität der initialen Parameterschätzungen, der Profilstärke, den Charakteristika des Forschungsfeldes, den Fähigkeiten und der Motivation der Apn. Vermutlich ist auch wesentlich, welches Meßgütekriterium bei vergleichenden Untersuchungen verwendet wird.

3.5.4 Objektpräsentation

Steht das Objektset durch die Wahl des Designs fest, stellt sich als nächstes die Frage, wie die Objekte den Befragten präsentiert werden sollen. Es gibt eine Reihe von Objektpräsentationsmodi. Deren Einsatzmöglichkeit ist nicht unabhängig von anderen Vorgehensweisen bei der CA. Insbesondere ist die Entscheidung für einen Präsentationsmodus abhängig von der Definition der Beurteilungsobjekte (Vollprofil, Teilprofil), der Designwahl (feststehendes Design, adaptives Design) und dem Antwortmodus (Rangreihung, Rating, Paarvergleich). Das Verfahren der Objektpräsentation dürfte wichtige Einflüsse auf die ökologische Validität der Untersuchung ausüben.

3.5.4.1 Objektpräsentationsmodus

Die ökologische Validität einer CA-Untersuchung wird grundsätzlich durch eine realitätsnahe Gestaltung der Beurteilungsobjekte unterstützt. Die schützt nicht nur vor der Gefahr, daß die Apn sich unter den Beurteilungsobjekten etwas Falsches vorstellen, sie dürfte im allgemeinen auch das Interesse der Befragten an der Untersuchung steigern und sich damit positiv auf die Auskunftsmotivation auswirken. Man kann zwischen vier grundlegenden Präsentationsmodi unterscheiden: Die schriftliche, verbale, bildhafte und modellhafte Objektpräsentation.

Die *schriftlichen Darbietung* ist dabei die universellste Methode. Sie erlaubt die Präsentation auf Papier, durch Dias oder Tageslichtprojektoren

(Gruppeninterviews) oder durch die computergestützte Darbietung auf einem Bildschirm. Die schriftliche Form ist sowohl bei der Vollprofilmethode als auch bei Teilprofilmethoden anwendbar. Insbesondere erlaubt sie CA-Untersuchungen über Objekte, die noch gar nicht existieren. Diese Eigenschaft ist oft bei Produktkonzepttests erforderlich. Die schriftliche Darstellung kann mehr oder weniger strukturiert erfolgen. Zwei Beispiele, mit verschiedenem Grad der Strukturiertheit, sind in Abbildung 3.1 und 3.2 verdeutlicht.

Wohnung 12	
Entfernung zur Stadtmitte	3 km
Miete (ohne Nebenkosten)	20 DM/ qm
Umgebung	Arbeitsviertel
Größe	65 qm
Zustand	Neubau

Abbildung 3.1: Beispiel einer schriftlichen, hoch strukturierten Objektbeschreibung.

Oft wird bei der CA die hochstrukturierte Version vorgezogen, weil die Attributbeschreibungen kürzer sind. Sie stellt damit im Verlauf der Befragung niedrigere Ansprüche an die Informationsverarbeitungskapazität der Apn und beansprucht weniger Zeit. Bei dieser Vorgehensweise sollte aber darauf geachtet werden, daß in der Aufwärmphase ausführliche Beschreibungen der Kurzformen gegeben und von den Apn richtig verstanden werden. Unter Umständen sollte während der Befragung die Ap jederzeit Zugriff auf ausführliche Attributbeschreibungen haben. Die weniger strukturierte Beschreibung (Abbildung 3; 'paragraph description'; Green & Srinivasan, 1978) ist zwar wesentlich komplexer als die strukturierte, sie kann jedoch z.B. bei der Überprüfung von werblichem Textmaterial notwendig werden, so daß die Apn sowohl die Objekte als auch die werbliche Umsetzung dieser Objekte beurteilen.

Wohnung 12
Die Wohnung befindet sich in einer Entfernung von 3 Kilometern vom Stadtkern in einem Stadtteil, der überwiegend von Arbeitern und deren Familien bewohnt wird.
Sie hat eine Grundfläche von 65 Quadratmetern. Besonders attraktiv an ihr ist, daß es sich um einen Neubau handelt.
Der Mietpreis ohne Nebenkosten beläuft sich auf 20 DM je Quadratmeter.

Abbildung 3.2: Beispiel einer schriftlichen, weniger strukturierten Objektbeschreibung.

Ein weiterer Vorteil der schriftlichen Methode ist, daß die Attributrepräsentationen direkt die objektive Eigenschaften, den subjektiven utilitaristischen Nutzen oder den symbolischen Nutzen der Objekte ansprechen kann (Myers & Shocker, 1981; vgl. Abschnitt 3.2.5). Bei der Festlegung der Attributstufen kann sowohl eine absolute Definition gewählt werden (Mietpreis: 12 DM/ qm - 16 DM/ qm - 20 DM/ qm) als auch eine relationale Definition (Mietpreis: 25% unter Durchschnitt - Durchschnitt - 25% über Durchschnitt). Diese Eigenschaften der schriftlichen Objektpräsentation tragen dazu bei, daß diese Darbietungsart sehr flexibel und damit vielseitig einsetzbar ist.

Die Objekte *verbal* zu *präsentieren*, ist dem schriftlichen Modus grundsätzlich ähnlich. Jedoch setzt die begrenzte menschliche Informationsverarbeitungskapazität für (zwangsläufig) sequentiell dargebotenes auditorisches Material große Einschränkungen. Bei der Vollprofilmethode wird wahrscheinlich die Grenze des Möglichen bereits bei weniger Attributen erreicht, als bei der schriftlichen Darbietung. Auch bei verbal dargebotenen Paarvergleichen dürften Einschränkungen bei der Profilstärke liegen.

Die Rangreihung als Antwortmodus scheidet bei der verbalen Darbietung schon bei einem sehr kleinen Objektset aus. Aus diesem Grunde ist auch der Trade-off-Ansatz bei verbaler Objektvorgabe nicht praktikabel.

Die verbale Präsentation kann der Interviewer entweder bei persönlicher Anwesenheit oder aber, und hier liegt der große Vorteil der Methode, über das Telefon durchführen. Meist ist es erforderlich, Untersuchungen schnell und ökonomisch durchzuführen. Die fernmündliche verbale Objektpräsentation ist aus der Perspektive der Validität meist eine suboptimale Lösung. Sie kann jedoch aufgrund der Untersuchungsrahmenbedingungen indiziert sein.

Die *bildhafte oder modellhafte Objektpräsentation* wird meist aus Gründen der ökologischen Validität angestrebt. Sie hat die Vorteile, daß die Objekte sehr anschaulich und lebensecht sind. Die modellhafte Präsentation kann sogar bis zur Darbietung tatsächlicher Objekte ausgedehnt werden. Manche Attribute, wie z.B. Geschmack, können sogar nur sinnvoll über reale Objekte dargeboten werden. Allerdings unterliegt diese Art der Objektdarstellung einigen Einschränkungen. Meist ist es nicht möglich, die Teilprofilmethode einzusetzen. Auch die Designwahl ist begrenzt, da die Menge der Objekte im Objektset aus Kostengründen meist klein gehalten werden muß. Insbesondere adaptive Designs können nur mit einem großen Aufwand durch Modelle realisiert werden.

Im allgemeinen kann damit gerechnet werden, daß diese Präsentationsmodi vergleichsweise aufwendig und kostspielig sind.

Holbrook und Moore (1981) vergleichen die schriftliche mit der bildhaften Objektpräsentation. Ausgangspunkt dieser Untersuchung war die Vermutung, daß die bildhafte Objektpräsentation eher zu einer ganzheitlich-konfiguralen Informationsverarbeitung führen sollte, als die schriftliche Darbietungsform. Die ganzheitlich-konfigurale Informationsverarbeitung sollte an einem verstärkten Auftreten von Attributwechselwirkungen erkennbar sein. Die Autoren begründen die unterschiedliche Verarbeitung von bildhaftem und schriftlichem Material theoretisch mit der dualen Kodierungstheorie (Paivio, 1971) und der funktionalen Hemisphärenasymmetrie (z.B. Springer & Deutsch, 1988). Holbrook und Moore ließen ihre Apn modische Objekte (Pullover) beurteilen. Die Autoren gingen davon aus, daß solche modische Objekte am wahrscheinlichsten Attributinteraktionen hervorrufen sollten. Diese Annahme erwies sich als richtig, was sich daran zeigte, daß sich die Effektgrößen der Zweifachwechselwirkungen und der Haupteffekte ungefähr entsprachen. Bei der bildhaften Objektpräsentation konnte jedoch kein verstärktes Auftreten von Wechselwirkungen im Vergleich zu schriftlichem Material festgestellt werden.

Teas und Perr (1989) überprüften, ob die Ergebnisse der dekompositorischen Präferenz- und Ähnlichkeitsstrukturmodellierung bei der schriftlichen Darbietung von Objekten und der Darbietung von Prototypen vergleichbar sind. Ihre Untersuchung über die Beurteilung von Hundefutter zeigt eine weitgehende Übereinstimmung der Präferenzurteile zwischen schriftlich beschriebenen Objekten (paragraph descriptions) und Prototypen.

Die Ergebnisse der Untersuchungen zur Anwendungshäufigkeit verschiedener CA-Methoden von Cattin und Wittink (1982), Schubert (1991) und Wittink et al. (1994) sind in Tabelle 3.4 dargestellt.

Sie zeigen eine weite Überlegenheit schriftlicher und verbaler⁴ Präsentationsmethoden. Eine untergeordnete Verwendungshäufigkeit zeigen bildhafte, modellhafte und reale Darstellungsarten. Auch Kombination der Methoden kommen gelegentlich zum Einsatz.

⁴ Diese Kategorie umfaßt die schriftliche und mündliche Art der Präsentation. Welche Anteile auf die beiden, bei mir getrennt behandelten Darstellungsmöglichkeiten entfallen, kann den genannten Untersuchungen nicht entnommen werden. Aufgrund der Verwendung des Begriffs 'verbale Beschreibungen' ist jedoch zu vermuten, daß es sich überwiegend um eine schriftliche Präsentation der Objekte handelt.

Tabelle 3.4: Relative Häufigkeit der Verwendung verschiedener Objektpräsentationsmodi bei der CA in kommerziellen Studien [K] und Studien zu Forschungszwecken [F]. Alle Angaben in %.

Studie	Cattin & Wittink (1982)	Schubert (1991)	Wittink et al. (1994)
Zeitraum	1971 - 1980	1972 - 1989	1986 - 1991
Zweck der Studien	hauptsächlich K	K & F	hauptsächlich K
schriftliche und verbale Darstellung	70	62	78
bildhafte Darstellung	19	3	9
Modelle und reale Objekte	11	19	6
Kombinationen		15	7

3.5.4.2 Unterstützung der Objektpräsentation

In diesem Abschnitt sollen Möglichkeiten besprochen werden, wie die Objektpräsentation unterstützt werden kann. Diese Unterstützungen dienen hauptsächlich dem Zweck, die Informationsverarbeitungsprozesse der Apn zu vereinfachen, um dadurch die Qualität der Beurteilungen zu verbessern oder die Antwortzeit zu verringern.

Die *Telefon-Mail-Telefon Technik (TMT)* scheint erfolgversprechend zur Unterstützung der telefonisch-verbale Objektpräsentation anwendbar zu sein. Hierzu werden die Apn zunächst telefonisch um die Teilnahme an der Befragung gebeten. Nachdem eine Apn zugesagt hat, wird ihr angekündigt, daß man ihr in den nächsten Tagen Interviewmaterial per Post zusenden wird. Ein Termin für das telefonische Interview wird vereinbart. Dieser erste Anruf dient dazu, die Apn auf das schriftliche Material vorzubereiten und sich ihrer Teilnahme an der Untersuchung zu versichern. Das zugesandte Interviewmaterial kann je nach genauem CA-Verfahren sehr unterschiedlich sein. Es hat in erster Linie die Funktion, die verbal gestellten Fragen und die Antwortmöglichkeiten zu visualisieren. Es kann bei einer traditionellen Vollprofil-CA z.B. Papierkarten enthalten, auf welchen die Beurteilungsobjekte spezifiziert sind. Eine weitere Möglichkeit ist die Visualisierung einer (evtl. verankerten) Antwortskala, bei der die Antworten über Zahlen abgegeben werden. Eine besonders gewagte TMT-CA per Telefon beschreibt Stahl (1988). Er führte ACA per Telefon durch. Hierzu sandte er seinen Apn eine Reihe von Unterlagen, die sowohl den kompositorischen Teil von ACA visualisierte als auch den dekompositorischen Paar-

vergleichsteil und das abschließende Konzeptrating. Dies ist besonders schwierig, da das ACA-Design in der Paarvergleichs- und Konzeptratingphase adaptiv gebildet wird. Hierzu waren im Fragebogenmaterial mehrere identische Seiten enthalten, auf denen sowohl die Antwortskala als auch alle Attributstufen (im Paarvergleichsteil je links und rechts) schriftlich festgehalten waren. Der Interviewer bediente ACA und las der Ap die Fragen vor. Die Ap visualisierte sich die Fragen durch umkreisen der fraglichen Attribute auf je einem entsprechenden Fragebogenblatt und gab dann die Antwort ab. Der Interviewer gab diese an ACA weiter. ACA generierte adaptiv die nächste Frage usw. Weitere Erfahrungen mit telefongestützten CA liegen vor (z.B. Chrzan & Grisaffe, 1992). Moore (1992) ordnet und bespricht Aspekte der Untersuchungsdurchführung per Telefon.

In einer abgewandelten Version der TMT kann nach dem ersten Telefongespräch eine Diskette versandt werden. Auf ihr ist das komplette Interviewprogramm abgespeichert. Die Ap kann dieses Interview selbst bearbeiten und die Diskette anschließend zurücksenden. Eine Telefonhotline sollte für Fragen, die beim Programmstart oder während des Interviews auftreten, eingerichtet sein. Sendet eine Ap ihre Diskette nicht innerhalb einer bestimmten Frist zurück, kann sie telefonisch noch einmal gebeten werden, am Interview teilzunehmen.

Eine andere Technik, schriftliche Objektpräsentationen zu unterstützen, besteht darin, den schriftlichen Attributstufenbeschreibungen jeweils ein Symbol zuzuordnen und dieses mit der Attributstufe darzustellen. Der Informationsaufnahme-prozeß beim Lesen der Objektprofile verläuft normalerweise sequentiell (Anderson, 1981) und relativ langsam. Hat eine Ap die Bedeutung der Symbole einmal gelernt, sollte durch symbolische Unterstützung des schriftlichen Materials der Wahrnehmungsprozeß beschleunigt werden. Darüber hinaus könnte eine weitgehend simultane Aufnahme aller Attributinformationen, wie sie z.B. bei der bildhaften Darstellung der Objekte entsteht, erleichtert werden. Es gibt noch keine Untersuchungen, die sich mit dieser Hypothese beschäftigt.

3.5.4.3 *Fazit*

Die überwiegende Mehrheit conjointanalytischer Befragungen arbeitet mit schriftlich dargestellten Objekten. Diese Vorgehensweise ist ökonomisch, flexibel, und die Datenerhebung läßt sich einfach computergestützt durchführen.

Telefonische Befragungen lassen sich über die TMT-Technik mit schriftlichen Objektpräsentationen durchführen. Kritisch ist zu hinterfragen, ob die verbalschriftlichen Beschreibungen auch die 'richtigen' Vorstellungen bei den Apn hervorrufen. Da der Leseprozeß vergleichsweise langsam vor sich geht, und die Informationsintegrationsaufgabe für die Apn nicht einfach ist, könnte die Aufgabe durch symbolische Unterstützung möglicherweise effizienter gestaltet werden.

3.6 Beurteilungsaufgabe

Die Beurteilungsaufgabe der Apn wird einerseits von der Art des abzugebenden Urteils (Präferenz, Ähnlichkeit, ...) und andererseits von der Art der Erhebung der abhängigen Variablen bestimmt (Rangordnung, Rating, Paarvergleich, ...). In Abschnitt 3.6.1 möchten wir uns kurz mit dem ersten Problembereich beschäftigen, Abschnitt 3.6.2 wendet sich dann ausführlicher dem Antwortmodus zu.

3.6.1 Art des Urteils

Die CA wird fast ausschließlich zur Modellierung von Präferenzen eingesetzt. Der Begriff 'Präferenz' ist dabei sehr weit gefaßt und bezieht sich auf alle Arten von Dominanzurteilen (Nützlichkeit, Kaufabsicht, Sympathie, ...). Wie Green und DeSarbo (1978) und Green, Rao und DeSarbo (1978) gezeigt haben, läßt sich die CA auch auf die Modellierung von Ähnlichkeitsdaten übertragen. Diesen Ansatz möchten wir anhand eines Beispiels darstellen. Ein Forscher sei daran interessiert, wie Konsumenten verschiedene Urlaubsorte wahrnehmen, z.B. Las Palmas, Rimini, London, Cannes etc. Können Attribute generiert werden, durch welche sich solche Urlaubsorte beschreiben lassen (Preisniveau, Sehenswürdigkeiten, sportliche Möglichkeiten, Wetter,), dann ist die CA geeignet, Konsumentenpräferenzen zu erheben. Dies geschieht z.B. durch die Rangreihung von Vollprofilobjekten eines Objektsets gemäß den Präferenzen einer Ap. Aus der Modellspezifikation und der Parameterschätzung resultiert das PSM der Ap. In genau derselben Art kann auch jeweils ein Urlaubsort, z.B. Rimini, der Ap vorgegeben werden. Die Ap bekommt dann die Aufgabe, die Rangreihungsaufgabe nicht im Einklang mit ihren Präferenzen, sondern gemäß

der wahrgenommenen Ähnlichkeit der Vollprofil-Objekte mit dem Zielobjekt Rimini vorzunehmen. Es resultiert ein 'Ähnlichkeitsstrukturmodell' aus dem die wahrgenommenen Eigenschaften des Zielobjekts abzulesen sind. Diese kann folgendermaßen interpretiert werden: Die 'relative Wichtigkeit', also die Differenz zwischen dem höchsten und niedrigsten 'Teilähnlichkeitswert' innerhalb eines Attributs, entspricht der 'Attributsalienz'. Ist diese hoch, dann trägt das Attribut überdurchschnittlich zur Wahrnehmung des Zielobjekts bei. Stellt man eine hohe Variabilität dieser Attributsalienz über die Apn fest, so kann dies bedeuten, daß das Zielobjekt auf diesem Attribut sehr unsicher wahrgenommen wird. Dies kann, gemeinsam mit den PSM, Implikationen für Kommunikationskampagnen haben (Green & Srinivasan, 1978).

Begutachtet man veröffentlichte Untersuchungen, welche die Methode der CA anwenden, dann stellt man fest, daß mit der CA fast ausschließlich Präferenzen modelliert werden (vgl. auch Wittink et al., 1994 bei kommerziellen Untersuchungen). Wir beschränken unsere Darstellung ebenfalls auf Präferenzurteile, obwohl die Möglichkeit der Modellierung anderer Urteile durchaus denkbar ist.

3.6.2 Antwortmodus: Erhebung der abhängigen Variablen

Bei der Erhebung der Untersuchungsdaten sind verschiedene Verfahren einsetzbar. Allen diesen Verfahren ist gemeinsam, daß die Apn die Aufgabe haben, Urteile oder Entscheidungen gemäß ihrer Präferenz über Beurteilungsobjekte abzugeben. Die Erhebungsmodi unterscheiden sich darin, wie schwierig ihre Bearbeitung für die Apn ist und haben meist eine direkte Auswirkung auf das nachfolgende Parameterschätzverfahren. In diesem Abschnitt wollen wir die Erhebungsverfahren kurz darstellen und wichtige Implikationen diskutieren.

3.6.2.1 Kategorienzuordnung

Aus Sicht der Apn besteht die einfachste Antwortart in der Zuordnung der Objekte zu Kategorien. Die Benennung der Kategorien kann eine Präferenzordnung beinhalten (finde ich ... nicht gut - gut - sehr gut). Das kategoriale Conjoint-Measurement-Modell von Carroll (1969, zit. in Carroll, 1973) verlangt eine polytome Antwortvariable.

Das kategoriale CM ist nur von theoretischer Bedeutung, da uns keine Anwendung des kategorialen Antwortmodus' bekannt ist. Die Ursache hierfür ist sicher in der Tatsache zu suchen, daß bei einer Antwort mit nur wenigen Kategorien sehr viel Information verschenkt wird. Wird die Kategorienanzahl erhöht (z.B. Einstufung von 1 bis 10), handelt es sich um eine Rating-Antwort (Abschnitt 3.6.2.3). Der Übergang ist dabei fließend.

Die Kategorienzuordnung kommt jedoch oft gewinnbringend als Vorstufe zur Rangreihung zum Einsatz. Die prinzipielle Vorgehensweise sieht dabei so aus, daß die Ap zunächst die Objekte geordneten Kategorien zuordnet. Bestehen die Objekte z.B. aus schriftlichen Vollprofilbeschreibungen auf Papierkarten, dann kann die Kategorienzuordnung durch die Bildung von fünf Kartestapeln erfolgen. Der erste Stapel enthält sehr schlecht beurteilte Objekte, der fünfte Stapel enthält die Objekte, die die Ap sehr gut findet. Danach bekommt die Ap die Aufgabe, eine Rangreihe der Objekte innerhalb der Stapel zu bilden. Die letzte Aufgabe besteht dann darin, eventuell notwendige Umordnungen an den Kategoriengrenzen vorzunehmen.

3.6.2.2 Rangreihung

Die Bildung von Rangreihen wird bei der Trade-off-Methode regelmäßig und bei der Vollprofilmethode häufig angewandt. Die Aufgabe ist psychologisch sinnvoll und fällt den Apn vergleichsweise leicht, sofern bei der Vollprofilmethode nicht zu viele Attribute gleichzeitig zu beurteilen sind. Sind viele Objekte zu beurteilen, dann schaltet man zur Vorbereitung der Rangreihung oft eine Kategorienzuordnungsaufgabe vor (vgl. Abschnitt 3.6.2.1).

Die Rangreihungsaufgabe erfordert von den Apn viele Vergleiche zwischen den Objekten und regt so zur Sorgfalt an. Es ist zu vermuten, daß die Reliabilität der Rangreihenantworten hoch ist. Acito (1977) konnte mit einem Test-Retest-Design diese Vermutung bestätigen.

Die Rangreihenbildung als Antwortmodus entspringt dem Grundgedanken des CM, zu dessen Definition die ordinale Skalierung der abhängigen Variablen gehört. Wie bei der Besprechung des CM bereits ausgeführt wurde (Abschnitt 2.3.3), wird den geschätzten Parametern metrische Skaleneigenschaften zugesprochen, obwohl die abhängige Variable nichtmetrisch ist.

Die Nachteile des Verfahrens liegen hauptsächlich darin, daß die Befragungsmaterialien bei mehr als zehn Objekten aus losen Papierkarten bestehen müssen, um die Übersichtlichkeit der Aufgabe für die Apn zu wahren. Für den Untersucher ist der Umgang mit diesen Materialien im Vergleich zu üblichen Fragebögen erschwert (Schneidearbeit, verlorene Karten, ...). Die Datenprotokollierung erfordert praktisch immer einen Interviewer. Ein weiterer Nachteil besteht darin, daß die Aufgabe in der Regel nicht computergestützt vorgegeben werden kann, weil sie einen sehr großen Bildschirm und genügende Fertigkeiten der Apn bei der Bedienung erfordern würde.

Als Parameterschätzverfahren bei rangskalierten Antworten wurden die Verfahren MONANOVA, PREFMAP, LINMAP und JOHNSON konzipiert (vgl. Abschnitt 3.8.1).

In letzter Zeit werden häufig auch Parameterschätzverfahren für metrische Daten, wie z.B. ordinary least square (OLS), bei der Schätzung der Parameter auf der Grundlage von Rangreihen eingesetzt (Wittink & Cattin, 1989). Dies dürfte darauf zurückzuführen sein, daß OLS-Schätzprogramme leicht verfügbar sind und eine Reihe von Simulationsstudien und empirischen Untersuchungen zu dem Ergebnis kommen, daß die geschätzten Parameter und insbesondere deren Validität weitgehend vom Parameterschätzverfahren unabhängig sind

(Carmone, Green & Jain, 1978; Cattin & Bliemel, 1978; Green, 1975; Green & Srinivasan, 1978; Jain, Acito, Malhotra & Mahajan, 1979; Wittink & Cattin, 1981). Obwohl das OLS-Verfahren die Intervallskaliertheit der abhängigen Variablen voraussetzt und es theoretisch bei einer ordinalskalierten Antwortvariablen unangemessen ist, kommt es zum Einsatz, weil die Schätzergebnisse ohnehin den Ergebnissen von z.B. MONANOVA sehr ähnlich sind.

Eine sehr verbreitete Anwendung der Rangreihung findet sich bei der kriterienorientierten Validierung der PSM mittels Holdout-Objekten (vgl. Abschnitt 5.4). Bei dieser Aufgabe tritt die genannte Schwierigkeit mit den Befragungsmaterialien nicht auf, weil meist nur sehr wenige (4-8) Holdout-Objekte ausreichend sind. Die Objekte werden in einem normalen Fragebogen schriftlich fixiert. Die Apn ordnen ihnen einen Rangplatz zu, ohne die Objekte zu verschieben.

3.6.2.3 *Rating*

Eine immer häufiger werdende Alternative zur Rangreihung ist die Einstufung der Objekte auf einer Ratingskala (vgl. Abschnitt 3.6.2.5). Dieser Antwortmodus bietet insbesondere in der Erhebungsphase den Vorteil der einfachen Handhabbarkeit. Bei Interviews mit der Papier-und-Bleistift-Methode, über das Telefon und bei der computergestützten Datenerhebung, ist das Objektrating viel einfacher anzuwenden als die Rangreihenbildung.

Ob die Antworten der Apn tatsächlich Intervallskaleneigenschaften haben, wie dies die Rating-Methode suggeriert, ist oft und zurecht in Frage gestellt worden. Grundsätzlich kann aber auch nur die ordinale Information von Ratingdaten weiterverarbeitet werden. Das größte Problem bei Ratingverfahren dürften Verankerungseffekte und die Änderung der psychologischen Qualität der Skala im Laufe des Interviews darstellen. Das Programm ACA 4.0 verwendet in der letzten Interviewphase Objektratings zur optimalen Verknüpfung der zuvor geschätzten TNW aus verschiedenen Interviewteilen und zu deren Kalibrierung. Um Ankereffekten vorzubeugen, wird das vermutlich am niedrigsten präferierte Objekt zuerst und das vermutlich am höchsten präferierte Objekt direkt danach dargestellt. Auf diese Weise soll sichergestellt werden, daß die Skala sinnvoll

verankert wird. Ein ähnlicher Effekt kann erzielt werden, indem die Apn alle Objekte vor der Ratingaufgabe kennenlernen.

Ein sowohl der Rating- als auch der Paarvergleichsmethode ähnliches Verfahren ist die Magnitudenskalierung. Diese Technik zur Erhebung von Antworten auf Reizgegebenheiten entstammt der Psychophysik. Ihr Prinzip besteht darin, daß die Antwort auf einen Reiz entsprechend der Intensitätswahrnehmung (z.B. Gewicht eines Körpers) so abgegeben wird, daß sie die Relation zur Wahrnehmung eines willkürlich festgelegten Standardreizes (z.B. Gewicht eines Körpers von 1 kg) abbildet. Die Antworten können dabei z.B. numerisch abgegeben werden. Wird beispielsweise der Gewichtswahrnehmung eines Standardkörpers der Wert 100 zugeordnet, dann hat eine Ap die Aufgabe, jedem folgenden Körper eine Zahl zuzuordnen. Wird das Gewicht z.B. als doppelt so hoch wahrgenommen, sollte die Zahl 200 betragen.

Teas (1987) überträgt das Prinzip der Magnitudenskalierung auf die Erhebung der Antwortvariablen bei der CA. Seine Apn geben ihre Antworten dabei numerisch und über die Zuordnung von Linienlängen gegenüber einem Standardobjekt ab. Seine Befunde deuten auf eine hohe Reliabilität der Antworten hin. Er stellt jedoch fest, daß dieser Antwortmodus zu einer erheblichen Heterogenität über die Apn führt und bezweifelt daher die Zweckmäßigkeit der Methode bei der Modellierung von individuellen Präferenzen. Die Einsatzhäufigkeit der Magnitudenskalierung bei der CA dürfte sehr gering sein.

3.6.2.4 Paarvergleichsverfahren

Paarvergleichsverfahren haben eine geringe Effizienz bei der Erhebung der abhängigen Variablen, weil die erhobene Informationsmenge je Zeiteinheit relativ gering ist. Gleichzeitig haben Paarvergleichsverfahren aber auch große Vorteile. Die Antwortaufgabe ist psychologisch angemessen und für die Apn relativ einfach zu erledigen. Aus diesem Grund dürften Paarvergleichsurteile eine hohe Reliabilität besitzen. Der wertvollste Vorteil des Verfahrens liegt jedoch in der Testbarkeit bestimmter Annahmen über das Antwortverhalten einer sinnvoll (konsistent) antwortenden Ap. Insbesondere sind bei der Anwendung von Paarvergleichstests weitverbreitet, ob eine Ap die Transitivitätsannahme verletzt (z.B. Curry, Menasco & van Ark, 1991; David, 1988; Hauser & Shugan, 1980).

Neben diesen psychologischen und psychometrischen Gründen hat die Methode des Paarvergleichs in jüngerer Zeit insbesondere durch die Verwendung von computergestützten Interviews an Attraktivität gewonnen. Paarvergleiche sind bei computergestützten Interviews häufig die Methode der Wahl, weil eine Rangreihung praktisch nicht möglich ist und Ratingverfahren oft deshalb als nachteilig angesehen werden, weil Vergleichsurteile für valider als absolute Urteile angesehen werden (Clancy & Garsen, 1970). Sie sind einfach auf dem Bildschirm darzustellen, und Maße der Modellanpassung (vgl. Abschnitt 5.3) können direkt nach oder sogar schon während des Interviews berechnet werden. Dies kann einen wertvollen Beitrag bei der Einschätzung der Objektbereichskenntnisse der Ap bzw. ihrer Antwortmotivation liefern.

Es gibt verschiedene Unterformen von Paarvergleichsverfahren. Wir wollen in Anlehnung an Schweikl (1985) den einfachen Paarvergleich, den abgestuften Paarvergleich, den Konstant-Summen-Paarvergleich und den Dollar-Metrik-Paarvergleich kurz vorstellen.

Beim *einfachen Paarvergleich* bekommt die Ap zwei Objekte zur Auswahl. Sie hat die Aufgabe anzuzeigen, welches der Objekte sie präferiert. Die Antwortvariable ist damit dichotom. Diese Art des Paarvergleichs ist für die Befragten eine sehr einfache Aufgabe. Es wird jedoch mit jeder Frage nur sehr wenig Information erhoben, was dazu führt, daß bei gegebener Größe des Objektsets relativ viele Urteilsaufgaben erforderlich sind, was wiederum eine lange Interviewdauer bedeutet. Sowohl die Einfachheit der Aufgabe (Unterforderung) als auch die lange Dauer des Interviews dürften zu einem sehr raschen Nachlassen der Auskunftsmotivation führen.

Beim *abgestuften Paarvergleich* ('graded paired comparison') verlangt man von der Ap zusätzlich zum dichotomen Präferenzurteil ein Urteil über die Intensität der angezeigten Präferenz. Dies kann durch zwei Arten der Antwortskala erreicht werden. Entweder wird die Antwort auf zwei Skalen abgegeben (Abbildung 3.3) oder beide Urteile werden durch eine einzige Antwort repräsentiert (Abbildung 3.4). Beide Möglichkeiten sind inhaltlich identisch.

Ein abgestufter Paarvergleich bringt nur dann gegenüber dem einfachen Paarvergleich zusätzliche Information, wenn das Ausmaß der Bevorzugung als Differenz mindestens intervallskalierter Nutzenwerte interpretiert werden kann. Eine notwendige Bedingung hierfür ist, daß die Nutzendifferenz der angebotenen Objekte nicht außerhalb der angebotenen Skala liegt. Auf die Auswahl

der Objekte ist demnach besonderer Wert zu legen (vgl. ACA-Paarvergleichs-differenzendesign in Abschnitt 3.5.3.3).

EIN BEISPIEL: Es könnten die untenstehenden Beschreibungen einer Fahrkarte angegeben werden. Entscheiden Sie bitte, welche Sie davon wählen würden, und geben Sie dabei den Grad Ihrer Bevorzugung mit an.

Großer Gültigkeitsbereich aber teuer	ODER	Kleiner Gültigkeitsbereich aber billig
1) bevorzugte Alternative:	bevorzuge links -1	bevorzuge rechts +1
2) Ausmaß der Bevorzugung:	Indifferenz stark 0 — 1 — 2 — 3 — 4	

Abbildung 3.3: Beispiel für die Beantwortung eines abgestuften Paarvergleichs auf zwei Antwortskalen.

EIN BEISPIEL: Es könnten die untenstehenden Beschreibungen einer Fahrkarte angegeben werden. Entscheiden Sie bitte, welche Sie davon wählen würden, und geben Sie dabei den Grad Ihrer Bevorzugung mit an.

Großer Gültigkeitsbereich aber teuer	ODER	Kleiner Gültigkeitsbereich aber billig
bevorzuge stark links	weiß nicht 0	bevorzuge stark rechts
-4 — -3 — -2 — -1 — +1 — +2 — +3 — +4		

Abbildung 3.4: Beispiel für die Beantwortung eines abgestuften Paarvergleichs auf einer Antwortskala.

Eine andere Möglichkeit, bei der prinzipiell alle (Teil-)Objekte paarweise dargeboten werden können, ist der *Konstant-Summen-Paarvergleich*. Bei dieser Art des Paarvergleichs werden die Apn gebeten, eine konstante Summe (z.B. 100 Punkte) auf die Objekte zu verteilen. Das Verhältnis der Punkteverteilung sollte dabei dem Präferenzenverhältnis entsprechen. Wenn einer Ap das Objekt X_1 zweimal lieber ist, als das Objekt X_2 , so sollte sie Objekt X_1 67 Punkte und X_2 33 Punkte zuweisen. Das Punktverhältnis entspricht damit dem Nutzenverhältnis

$u(X_1) / u(X_2) = 67 / 33 = 2 / 1$. Verhält sich die Ap instruktionskonform, dann sind die Nutzenwerte verhältnisskaliert. Es kann aber mit Recht in Frage gestellt werden, ob die Apn mit dieser Urteilsaufgabe nicht überfordert werden.

Der *Dollar-Metrik-Paarvergleich* kann als Sonderform des abgestuften Paarvergleichs angesehen werden. Die Ap erhält die Aufgabe, zunächst anzugeben, welches der Objekte sie (unter sonst gleichen Bedingungen) bevorzugen würde. Hat die Ap sich für ein Objekt entschieden, wird sie gefragt, wie groß der Preisunterschied sein muß, daß Indifferenz zwischen den beiden Objekten eintritt. Die Dollar-Metrik-Skala hat den Vorteil gegenüber dem abgestuften Paarvergleich, daß die Skalenpunkte eindeutig definiert sind ([100 DM teurer] vs. [2=mäßige Bevorzugung] auf einer Ratingskala), und die Skala nach oben offen ist. Durch diese Offenheit können prinzipiell alle Objekte beim Paarvergleich miteinander kombiniert werden.

Verhält sich die Ap instruktionskonform, dann können die angegebenen Mehrpreise als Differenzen intervallskalierter Nutzenwerte interpretiert werden. Eine Preisdifferenz von 200 DM auf 300 DM bedeutet die gleiche Nutzen-differenz wie die Differenz zwischen 2000 und 2100 DM.

Das Skalenniveau der Präferenzurteile bei abgestuften Paarvergleichen, bei Konstant-Summen-Paarvergleichen und Dollar-Metrik-Paarvergleichen kann auf individueller Ebene durch den Test von Hauser und Shugan (1980) überprüft werden. Die Grundlage für den Test ist dabei die Prüfung der Transitivität von Objekttripeln. Der Test ist am Beispiel der Dollar-Metrik-Skala bei Schweikl (1985) ausführlich dargestellt.

3.6.2.5 *Validität und Anwendungshäufigkeit der Antwortmodi*

Leigh et al. (1984) vergleichen in ihrem Experiment zur Beurteilung von Taschenrechnern verschiedene Antwortmodi bei der Erhebung der abhängigen Variablen (Rangreihung, Rating, einfacher Paarvergleich, abgestufter Paarvergleich, Dollar-Metrik-Paarvergleich). Hinsichtlich der Reliabilität und kriterienorientierten Validität (z.B. Vorhersage von Wahlentscheidungen aus einer Menge von Taschenrechnern) konnten die Autoren keine statistisch signifikanten Unterschiede zwischen den Methoden finden. Es ist jedoch anzumerken, daß bei diesem Experiment die Stichprobengrößen innerhalb der Versuchsbedingungen relativ gering waren. Hinsichtlich der Vorhersage der

Wahlentscheidungen schneiden die Rating-Methode und der Dollar-Metrik-Paarvergleich tendenziell am besten ab.

Tabelle 3.5: Relative Häufigkeit der Verwendung verschiedener Antwortmodi bei der CA in kommerziellen Studien [K] und Studien zu Forschungszwecken [F]. Alle Angaben in %.

Studie	Cattin & Wittink (1982)	Wittink & Cattin (1989)	Schubert (1991)	Wittink et al. (1994)
Zeitraum	1971 - 1980	1981 - 1985	1972 - 1989	1986 - 1991
Zweck der Studien	hauptsächlich K	hauptsächlich K	K & F	hauptsächlich K
Rangordnung	45	36	56	22
Rating	34	49	32	70 ¹
Paarvergleich	11	9	8	5
andere	10	6	3	3

¹ Diese Kategorie enthält den abgestuften Paarvergleich des Programms ACA.

Eine andere Frage ist, wie häufig die Erhebungsmethoden bei CA-Untersuchungen verwendet werden. Tabelle 3.5 gibt die relativen Häufigkeiten der verschiedenen Antwortmodi der Studien von Cattin und Wittink (1982), Wittink und Cattin (1989), Schubert (1991) und Wittink et al. (1994) wieder.

Bei der Studie von Wittink et al. (1994) fällt der abgestufte Paarvergleich des Programms ACA in die Kategorie 'Antwortrating'. Aus diesem Grunde ist eine direkte Vergleichbarkeit der Studien nicht gegeben. Dennoch lassen sich einige grundlegende Aussagen ableiten. Erstens ist festzustellen, daß der einfache Paarvergleich zunehmend an Bedeutung verliert. Die relative Häufigkeit der Rangreihung, der klassische Antwortmodus des CM, nimmt insbesondere bei kommerziellen Untersuchungen in letzter Zeit stark ab. Dies dürfte an der Notwendigkeit zur Berücksichtigung einer größeren Zahl von Attributen je Studie liegen und der Tatsache, daß sich das Programm ACA, zumindest bei kommerziellen Studien, immer mehr durchsetzt. Die letztere Ursache und weitere, in Abschnitt 3.6.2.3 näher bezeichnete Gründe, dürften zu der starken Verbreitung von Ratingurteilen bei der CA geführt haben.

3.6.2.6 Fazit

Bei der CA sind mehrere Antwortmodi einsetzbar. Die Methode der Wahl steht insbesondere auch in Wechselwirkung mit der Definition der Beurteilungsobjekte (Vollprofile, Teilprofile, Trade-off-Ansatz). Die Rangreihung ist beim Trade-off-Ansatz und bei der Verwendung von Vollprofilen bei wenigen Objekten im Objektset bevorzugt einzusetzen. Nimmt die Menge der Objekte im Objektset zu, kann auf Ratingverfahren zurückgegriffen werden. Ratingverfahren und Paarvergleichsverfahren (insbesondere deren Kombinationen in abgestuften Paarvergleichen, Konstant-Summen-Paarvergleichen und Dollar-Metrik-Paarvergleichen) sind insbesondere dann indiziert, wenn die Befragung computergestützt erfolgt.

Die Wahl des Antwortmodus bzw. des angenommenen Skalenniveaus der Antwortvariablen bestimmt genau genommen das nachfolgende Parameterschätzverfahren. Allerdings ist hier ein Trend zur Verwendung metrischer Verfahren zu verzeichnen. Die Gründe liegen in der Verbreitung dieser Verfahren (z.B. OLS). Diese Vorgehensweise wird damit gerechtfertigt, daß einige Simulationen und empirische Untersuchungen nur geringe bzw. keine Unterschiede hinsichtlich der Validität der PSM in Abhängigkeit des Parameterschätzverfahrens erbrachten.

3.7 Verfahren zur Nutzenschätzung

Sofern nicht gesondert erwähnt, gehen wir in diesem Abschnitt davon aus, daß ein additives Teilnutzenwertmodell spezifiziert wurde. Es hat die Form

$$(3.10) \quad u_h(X_i) = \begin{array}{|c|c|} \hline & \\ \hline \end{array} b_{hjk} * x_{ijk}$$

Die Parameter b_{hjk} sind zu schätzen. Da sie approximativ intervallskaliert sind, der Nullpunkt der Skala also nicht bedeutungshaltig ist, kann innerhalb eines jeden Attributs ein Parameter beliebig festgelegt werden. Es hat sich eingebürgert, den Wert des niedrigsten Parameters innerhalb eines Attributs auf Null zu setzen. Bei vier dreistufigen Attributen hätte das additive Teilnutzenwertmodell für die Ap h folgende Form:

$$u_h(X_i) = b_{h12} x_{i12} + b_{h13} x_{i13} + b_{h22} x_{i22} + b_{h23} x_{i23} \\ + b_{h32} x_{i32} + b_{h33} x_{i33} + b_{h42} x_{i42} + b_{h43} x_{i43} \quad (3.19)$$

Es sind in diesem Falle acht Teilnutzenwerte b_{hjk} zu schätzen.

In diesem Abschnitt werden die gebräuchlichsten Verfahren der Nutzen-schätzung dargestellt und diskutiert. In Abschnitt 3.7.1 werden verschiedene kompositorische Ansätze besprochen. Abschnitt 3.7.2 befaßt sich mit der Darstellung dekompositorischer Verfahren der CA.

3.7.1 Kompositorischer Ansatz - explizite Modellbildung

Es sind vielfältige kompositorische Ansätze der Nutzenmessung zu finden. In Abschnitt 3.7.1.1 werden einige frühere Vorschläge vorgestellt, Abschnitt 3.7.1.2 behandelt ein neueres explizites Modellbildungsverfahren, das insbesondere auf die funktional vergleichbare Instruktion der Apn großen Wert legt. Abschnitt 3.7.1.3 behandelt kurz einige Probleme der expliziten Modellbildung, und Abschnitt 3.7.1.4 ist den Befunden zu dieser Verfahrensklasse vorbehalten. Insbesondere wird auf vergleichende Untersuchungen zwischen kompositorischen und dekompositorischen Ansätzen eingegangen.

3.7.1.1 Kompensatorische, explizite Modelle

In Abschnitt 2.3.2 wurde bereits das *SMART-Verfahren der expliziten Modellbildung* (Edwards, 1977) näher beschrieben. Während SMART additiv verknüpfte, lineare Vektorpräferenzmodelle annimmt, gehen die folgenden Verfahren von den allgemeineren additiven Teilnutzenwertmodellen aus.

Ein weitverbreiteter Ansatz der expliziten Modellbildung ist bei Green (1984) beschrieben. Bei diesem Verfahren, hier als *Standardverfahren der expliziten Modellbildung* bezeichnet, wird einer Ap zunächst jeweils ein Attribut mit der Beschreibung seiner Stufen vorgelegt. Die Ap hat die Aufgabe, die

Attributstufen auf einer (z.B.) 10-Punkte Ratingskala einzustufen, wobei der am niedrigsten präferierten Attributstufe der Wert 0 und der am höchsten präferierten Stufe der Wert 10 zugewiesen werden muß. Alle dazwischen liegenden Attributstufen können dann auf der Ratingskala zwischen 0 und 10 frei eingestuft werden. Ist diese Aufgabe für alle Attribute erledigt, wird die Attributwichtigkeit mittels einer Konstant-Summen-Prozedur ermittelt. Die Ap erhält hierzu die Aufgabe, (z.B.) 100 Punkte auf die Attribute entsprechend ihrer Wichtigkeit zu verteilen. Die Teilnutzenwerte der Attributstufen ergeben sich aus der Multiplikation der Präferenzratingwerte mit der jeweiligen Attributwichtigkeit. Dieses Prinzip der kompositorischen Nutzenmessung fand breite Verwendung (z.B. Hoepfl & Huber, 1970; Huber, 1974a; Huber, Daneshgar & Ford, 1971).

Eine andere Form der kompositorischen Nutzenmessung verwenden Leigh et al. (1984). Sie benutzen ein *Dollar-Metrik-Ratingverfahren* bei dichotomen Attributen zur expliziten Modellbildung. Leigh et al. (1984) untersuchen Präferenzen gegenüber Taschenrechnern. Den Apn werden nacheinander die Attribute vorgelegt. Die Attribute beschreiben jeweils, ob ein Objektmerkmal vorhanden ist oder nicht (z.B. Attribut IV, statistische Funktionen vorhanden vs. nicht vorhanden). Die Befragten erhalten die Aufgabe, den zusätzlichen Preis in Dollar anzugeben, den sie in Kauf nehmen würden, wenn ein Taschenrechner das jeweilige Merkmal besitzt.

Man kann diesen Ansatz einfach auf multichotome Attribute übertragen. Man muß lediglich die Beurteilungsaufgabe ein wenig uminterpretieren. Bei zweistufigen Attributen kann die Antwortvariable als der Mehrpreis definiert werden, den die Ap zu bezahlen bereit ist, wenn ein Objekt auf Attribut j von Stufe 1 (nicht vorhanden) auf die Stufe 2 (vorhanden) verändert wird. Die Aufgabe entspricht genau der ursprünglichen von Leigh et al.. Sind die Attribute mehrstufig, kann man analog vorgehen. Es wird, ausgehend von einer Referenzattributstufe innerhalb eines jeden Attributs, nach dem Mehrpreis gefragt, den eine Ap bereit ist zu bezahlen, wenn ein Objekt auf Attribut j von der Referenzstufe auf die Stufe k verändert wird. Die Antworten sind direkt die Teilnutzenwerte. Sie wären bei instruktionskonformem Verhalten der Ap verhältnisskaliert. Ob die Apn in der Lage sind, diese Urteilsaufgabe gemäß der Instruktion zu leisten, kann jedoch angezweifelt werden.

3.7.1.2 Konjunktiv-kompensatorische explizite Modellbildung

Einen allgemeineren Ansatz der *konjunktiv-kompensatorischen expliziten Modellbildung* schlägt Srinivasan (1988) vor. Er nimmt einen zweistufigen Entscheidungsprozeß an. Er geht davon aus, daß in der ersten Phase eine konjunktive Entscheidungsregel zur Eliminierung der Objekte mit nicht akzeptierbaren Attributstufen eingesetzt wird. Beurteiler kennzeichnen in dieser Phase die Attributstufen, bei deren Vorliegen ein Objekt unabhängig von der Ausprägung auf anderen Attributen keiner weiteren Beurteilung unterzogen und als nicht mehr akzeptierbar zurückgewiesen wird. Diese Attributstufen werden eliminiert. Danach folgt die Modellierung der kompensatorischen Präferenzbildung. Srinivasan legt diesen Beurteilungsprozessen ein additives Teilnutzenwertmodell zugrunde (Gleichung 3.10).

Die Beurteiler verständlich zu instruieren, ist bei dieser Aufgabe nicht ganz einfach. Ein Beispielschema der konjunktiv-kompensatorischen Modellbildungsprozedur nach Srinivasan (1988) ist in Anhang B aufgenommen. Die Schätzung der Parameter erfolgt in folgenden Stufen:

1. Zunächst hat ein Beurteiler die Aufgabe, innerhalb eines jeden Attributs die Stufen der höchsten und der niedrigsten Präferenz zu bestimmen.
2. Die relative Wichtigkeit eines Attributs wird als die Differenz des Nutzens zwischen den Konsequenzen definiert, die auftreten, wenn sich die Eigenschaften eines Objekt von der am niedrigsten präferierten Stufe zur am höchsten präferierten Stufe innerhalb des Attributs ändern.
3. Das wichtigste Attribut wird bestimmt. Srinivasan nennt dieses das kritische Attribut. Der am wenigsten präferierten Stufe wird der Wert 0 zugeordnet, der am höchsten präferierten Stufe der Wert 100. Damit ist die relative Wichtigkeit des kritischen Attributs 100.
4. Die Wichtigkeit der anderen Attribute wird eingeschätzt. Jedem Attribut wird ein Wert der Wichtigkeit zwischen 0 und 100 zugeordnet. Diese Werte werden so gewählt, daß sie die Relationen der Wichtigkeiten der Attribute zum kritischen Attribut abbilden.
5. Innerhalb der Attribute wird der am wenigsten präferierten Attributstufe der Wert 0 und der am höchsten präferierten Stufe der Wert 100 zugeordnet. Allen anderen Stufen werden dann Werte zwischen 0 und 100 so zugeordnet, daß die Werte die Relation der Präferenzen der Attributstufen

innerhalb des jeweiligen Attributs widerspiegeln. Die niedrigsten und höchsten Stufen dienen dabei als Anker.

6. Die Teilnutzenwerte der Attributstufen erhält man, indem die Wichtigkeits-einschätzungen (Stufe 4) mit den Präferenzeinschätzungen (Stufe 5) multipliziert werden. Anschließend kann man die Teilnutzenwerte noch normieren.

Wie in Abschnitt 3.4.3 bereits ausführlich dargelegt wurde, sollte mit der Modellierung konjunktiver Prozesse zurückhaltend umgegangen werden. Zum einen bedeuten als 'nicht akzeptierbar' gekennzeichnete Attributstufen nicht unbedingt, daß ein konjunktiver Urteilsprozeß vorliegt. Zum anderen ist zweifelhaft, ob konjunktiv-kompensatorische PSM bessere Vorhersagen erlauben als kompensatorische PSM. Der Ansatz von Srinivasan (1988) ist, zumindest was den kompensatorischen Teil angeht, dennoch positiv zu beurteilen, da die Ausführlichkeit und die Art der Instruktion annähernd intervallskalierte Teilnutzenwerte erwarten läßt.

3.7.1.3 Probleme der kompositorischen Modellbildung

In Anlehnung an Green und Srinivasan (1990) kann man einige potentielle Probleme kompositorischer Modellbildungsansätze aufzeigen.

1. Bei expliziten Modellbildungsansätzen wird das PSM festgelegt. Es gibt keinerlei interne Überprüfungsmöglichkeiten des Modells. Im Gegensatz dazu wird bei dekompositorischen Ansätzen das Modell an die Beurteilungsdaten angepaßt. Maße der Modellanpassung (vgl. Abschnitt 5.3) geben zumindest Hinweise auf die Anpassungsgüte. Weitere Überprüfungsmöglichkeiten der Adäquatheit des Modells sind z.T. möglich.
2. Bei additiven Teilnutzenwertmodellen ist streng darauf zu achten, daß keine Redundanzen zwischen den Attributen auftreten. Vollständig redundante Attribute würden doppelt in die Gesamtnutzenschätzung eingehen. Eine Ap würde bei der expliziten Befragung keine Widersprüche entdecken, weil die Attribute unabhängig voneinander eingeschätzt werden. Bei dekompositorischen Ansätzen dürften solche Redundanzen den Apn deutlich auffallen (wenn z.B. nach einer 'großen Wohnung' mit einer 'geringen Wohnfläche' gefragt würde).

3. Die Nutzenfunktionen qualitativer Attribute dürften durch die spezielle Befragungsmethode zur Linearität neigen. Hat z.B. das Attribut Preis die Stufen 10 DM, 15 DM und 20 DM, und werden die Nutzenwerte der Stufe 10 DM bzw. 20 DM per definitionem auf 0 bzw. 100 festgelegt, dann werden die meisten Apn auf die Frage, den Nutzen der Stufe 15 DM anzugeben, mit dem Wert 50 antworten. Eventuell vorhandene nichtlineare Nutzenverläufe, die sich bei der dekompositorischen Präferenzstrukturmodellierung zeigen könnten, werden durch die direkte Art der Fragen zgedeckt.
4. Die einfache Frage nach Attributwichtigkeiten kann vieldeutig sein. Die Beurteilung hängt vom individuellen Erfahrungshintergrund bzgl. der Spannweite ab, den existierende Objekte auf diesem Attribut haben. Die Definition der Attributwichtigkeit von Srinivasan (1988) durch eine klar beschriebene Attributspannweite kann dieses Problem zumindest mildern.

Trotz dieser Probleme können explizite Modellbildungsansätze insbesondere bei einer großen Anzahl von maßgeblichen Attributen im Attributset wertvolle Dienste leisten.

3.7.1.4 Übereinstimmung von kompositorischen und dekompositorischen Ansätzen und deren kriterienorientierte Validität

Eine große Zahl von empirischen Untersuchungen beschäftigte sich mit der Übereinstimmung der Ergebnisse von kompositorischen und dekompositorischen Verfahren bzw. verglich deren kriterienorientierte Validität.

Bei ihrer Untersuchung zur Beurteilung von Gesundheitserhaltungsprogrammen verglichen Akaah und Korgaonkar (1983) explizite Modellierungsansätze mit mehreren dekompositorischen Verfahren (traditionelle CA, Hybrid-CA; vgl. Abschnitt 3.7.2.3). Fünf Attribute wurden in die Studie aufgenommen. Zwei explizite Verfahren wurden untersucht. Das erste entspricht dem Standardverfahren der expliziten Modellbildung (z.B. Green, 1984), das zweite gewichtet die Attribute gleich. Die explizit gebildeten Modelle sind hinsichtlich der kriterienorientierten Validität zu sechs Holdout-Objekten der traditionellen CA unterlegen. Wie nicht anders erwartet, ist das gleichgewichtete Modell dem Modell unterlegen, das die Attributwichtigkeit explizit erfragt. Die Effektgrößen sind jedoch nicht sehr groß. Das CA-Haupteffektmodell erreicht mit $r_s=0,37$ die

höchste Validität, das gewichtete explizite Modell (Standardverfahren) kann die Rangfolge der Holdout-Objekte mit $r_s=0,25$ vorhersagen.

Zu ähnlichen Ergebnissen kommen die bereits in Abschnitt 3.5.1 erwähnten Untersuchungen von Cattin et al. (1982) und Green et al. (1983). Green et al. (1983) erhalten die niedrigste Validität für explizite Modelle ($r_s=0,34$; Haushaltsprodukte, 7 Attributen). CA-Ansätze erreichen $r_s=0,65$, während die Hybrid-CA (s. Abschnitt 3.7.2.3) mit $r_s=0,77$ die Rangfolge von Holdout-Objekten vorhersagt. Dekompositorische Ansätze zeigen damit eine deutlich höhere Validität im Vergleich zu expliziten Modellbildungsansätzen. Auch Cattin et al. (1982) finden bei ihrer Untersuchung (Bankdienstleistungen, 5 Attribute) eine höhere Validität von CA-Modellen gegenüber expliziten Modellbildungsansätzen.

In einer Überprüfung seines konjunktiv-kompensatorischen Ansatzes verglich Srinivasan (1988) die prognostische Validität des Standardverfahrens expliziter Modellbildungsansätze (z.B. Green, 1984), seines neuen Ansatzes der konjunktiv-kompensatorischen expliziten Modellbildung und einer dekompositorischen CA. Da Srinivasan selbst keine CA durchführte, replizierte er die Untersuchung von Wittink und Montgomery (1979) zur prognostischen Validität der CA auf individueller Ebene bei der Berufswahl von MBA-Studenten mit den zwei kompositorischen Verfahren und verglich seine Validitätskoeffizienten mit denen von Wittink und Montgomery. Die Apn wurden in den Untersuchungen von Wittink und Montgomery (1979) und der Untersuchung von Srinivasan (1988) nach einigen Monaten gebeten, alle bis dahin eingegangenen Arbeitsplatzangebote auf den Untersuchungsattributen einzustufen. Es wurden alle die Apn von der Datenanalyse ausgeschlossen, die nicht mindestens zwei Angebote erhalten hatten bzw. die sich noch nicht für einen Job entschieden hatten. Es wurde geprüft, ob das jeweilige PSM der Apn die tatsächliche Arbeitsplatzentscheidung richtig vorhergesagt hätte. Die Ergebnisse zur Vorhersagefähigkeit des späteren Berufswahlverhaltens der Modelle sind wie folgt: Das konjunktiv-kompensatorische explizite Modell sagte 69% richtig vorher und das explizite Modell des Standardverfahrens 60%. Dieser Unterschied ist auf dem 10%-Niveau signifikant ($p<0,1$; $n=45$). Wittink und Montgomery (1979) konnten per CA 63% richtige Vorhersagen machen. Der Unterschied zum konjunktiv-kompensatorischen expliziten Modell ist nicht statistisch bedeutsam. Dennoch spricht die Untersuchung für die Qualität des Srinivasan-Ansatzes der expliziten

Modellbildung und zeigt insbesondere, daß sorgfältig konstruierte kompositorische Modelle hohe Validitäten erreichen können.

Die Befundlage spricht jedoch nicht einheitlich für die dekompositorischen Verfahren. Eine Prognoseüberlegenheit expliziter Modellbildungsansätze gegenüber CA-Ansätzen erhalten Wright und Kriewall (1980) bei ihrer Untersuchung zur Vorhersage von Collegebewerbungen. Auch in der erwähnten Untersuchung zur Bewertung von Taschenrechnern (5 Attribute) erhalten Leigh et al. (1984) tendenziell leicht bessere Vorhersagen der expliziten Modelle, die mit dem Dollar-Metrik-Verfahren gebildet wurden, als bei verschiedenen CA-Vorgehensweisen bzgl. allen drei verwendeten Validitätskriterien. Diese Unterschiede sind jedoch statistisch nicht signifikant.

Einige Untersuchungen vergleichen die Übereinstimmung zwischen abgeleiteten, dekompositorischen Nutzenschätzungen und expliziten Nutzenaussagen. Green und Wind (1973) finden bei ihrer Discountkartenstudie eine hohe Vorhersage von Vollprofilrankings durch explizite Modelle. Dies gilt sowohl bei 3 als auch bei 5 Attributen. Man sollte jedoch erwähnen, daß in diesem Objektbereich vergleichsweise einfach rationale Urteile gefällt werden können (z.B. eine Karte die zu 10% Rabatt in einem Kaufhaus berechtigt, in dem man monatlich 200 DM umsetzt, kostet 40 DM - würden Sie die Discountkarte kaufen?). Eine ebenfalls hohe Übereinstimmung der beiden Ansätze finden die Autoren bei ihrer Untersuchung zur Bewertung von Menues in Restaurants.

Wiley, MacLachlan und Moinpour (1977) finden dagegen eine nur sehr niedrige Übereinstimmung zwischen expliziten und dekompositorisch ermittelten Modellwerten. Bei ihrer Studie in den Objektbereichen 'Softdrinks' (3 Attribute) und 'Tee' (2 Attribute) erklären die explizit erfragten Attributwichtigkeiten und die explizit erfragte Lage der Idealpunkte (Idealpunktmodell) die entsprechenden dekompositorischen Modellwerte nur gering oder gar nicht. Auch Scott und Wright (1976) finden eine nur niedrige Übereinstimmung. Dorsch und Teas (1992) untersuchten die Übereinstimmung zwischen der expliziten Modellbildung (nach der Methode von Srinivasan, 1988) und der Vollprofil-CA. Die mittlere Korrelation innerhalb der Apn zwischen dekomponierten und explizierten Teilnutzenwerten (Attributwichtigkeiten) liegt bei $r=0,49$ ($r=0,18$). Sie beurteilen damit den Grad der Übereinstimmung als mittel bis niedrig.

3.7.1.5 *Fazit*

Faßt man die Ergebnisse der Validierungsstudien zusammen, dann zeigen kompositorische und dekompositorische Verfahren meist einen nicht besonders hohen Grad an Übereinstimmung, aber beide eine recht zufriedenstellende kriterienorientierte Validität. Dekompositorische Verfahren zur Nutzenschätzung sind gegenüber expliziten Modellbildungsansätzen hinsichtlich der kriterienorientierten Validität meist überlegen. Verwendet man jedoch ausgefeilte Techniken zur expliziten Modellbildung, wie z.B. die Verfahren von Srinivasan (1988) oder von Leigh et al. (1984), dann erreichen explizite Modellbildungsverfahren die Validitätskennwerte von CA-Verfahren oder übertreffen diese sogar noch. Diese Aussage gilt zumindest für einige Objektbereiche. Denkbar sind solche, in denen Beurteilungs- und Entscheidungsprozesse gut verbalisiert werden können.

Die einerseits nicht sehr hohe Übereinstimmung und andererseits die bei beiden Ansätzen vorhandene vergleichsweise hohe kriterienorientierte Validität könnte darauf hindeuten, daß beide Methoden z.T. verschiedene Aspekte des Beurteilungs- und Entscheidungsverhaltens bei der Kriterienaufgabe erfassen. Wenn dies so ist, dann sollte eine kompositorische-dekompositorische Methodenkombination eine höhere kriterienorientierte Validität erbringen, als jeweils einer der Ansätze für sich allein. Erste Hinweise, daß dies so sein könnte, findet man bei Huber et al. (1993). Kontrollierte Untersuchungen stehen hierzu jedoch noch aus.

3.7.2 Dekompositorischer Ansatz - Conjoint Analyse

In diesem Abschnitt werden verbreitete conjointanalytische Verfahren der dekompositorischen Nutzenschätzung dargestellt. Ausgehend von der Schwierigkeit der 'klassischen Ansätze der CA' Präferenzstrukturen erfolgreich bei einer größeren Anzahl von Attributen zu modellieren, wurden die Ansätze der hierarchischen Informationsintegrations-CA (HICA), der Hybrid-CA (HCA) und der adaptiven CA (ACA) entwickelt. Diese Ansätze werden wir vorstellen und gemeinsam auf der Grundlage empirischer Befunde in Abschnitt 3.7.3 kritisch diskutieren.

3.7.2.1 *Klassische Verfahren der CA*

Die klassische CA verwendet Beurteilungen von Objekten aus dem Objektset zur dekompositorischen Parameterschätzung eines zuvor spezifizierten PSM. Die Beurteilungsobjekte sind dabei auf allen Attributen des Attributsets definiert (Vollprofilansatz) oder nur auf einer Teilmenge der Attribute (Teilprofilansatz, Trade-off-Ansatz). Die Meßskala der abhängigen Antwortvariablen kann alle besprochenen Formen annehmen. In aller Regel wird die Rangordnungsmethode oder die Rating-Methode hierfür verwendet.

Die klassische CA ist ein gut untersuchtes und sehr brauchbares Verfahren, sofern die Anzahl der Attribute im Attributset und deren Stufen klein gehalten werden kann. Eine obere Grenze für diese Art der CA liegt bei 5-6 Attributen. Der Umgang mit mehr Attributen stößt beim Vollprofilansatz an kognitive Grenzen, beim Trade-off-Ansatz an motivationale Grenzen (vgl. Abschnitt 3.5.2).

Die drei nachfolgend beschriebenen Verfahren wurden hauptsächlich mit dem Ziel entwickelt, die CA vom Grundgedanken her zu bewahren, aber den klassischen Ansatz so zu verändern, daß die CA auch bei einer höheren Anzahl von Attributen anwendbar bleibt und zuverlässige Ergebnisse erbringt. Die Verfahren HICA, HCA und ACA gehen dabei verschiedene Wege. Allgemein gesagt versuchen sie, komplexe und langwierige Beurteilungsaufgaben in mehrere 'kleine CA' aufzuspalten, Parameter nicht mehr nur auf Individualebene, sondern auch auf aggregierter Ebene zu schätzen und dekompositorische Parameterschätzungen mit kompositorischen Nutzenschätzungen zu verknüpfen.

3.7.2.2 *Hierarchische Informationsintegrations-Conjoint Analyse (HICA)*

Sind bei einer CA eine größere Anzahl von Attributen zu berücksichtigen, besteht grundsätzlich die Möglichkeit, das Attributset in mehrere Subsets aufzuspalten, und die Präferenzstrukturen innerhalb der Subsets mit jeweils einer eigenständigen CA zu modellieren. Da die Parameter der geschätzten PSM (annähernd) intervallskaliert sind, können die Ergebnisse der Sub-CA theoretisch dann verknüpft, d.h. alle Parameterwerte auf eine gemeinsame Skala abgebildet werden, wenn jeweils ein Attribut eines Subsets auch in mindestens einem anderen Subset vorkommt. Solche Attribute nennt man auch 'Brücken-

attribute'. Die statistischen Eigenschaften dieser Vorgehensweise sind jedoch weitgehend unbekannt. Abbildung 3.5 veranschaulicht, a) wie man sich bei diesem Ansatz den Präferenzbildungsprozeß und b) die Modellbildung vorstellen kann.

Die HICA ist ein vergleichsweise neuer Ansatz, der nach einem ähnlichen Prinzip vorgeht. Er wurde von einer Gruppe um Louviere Mitte der achtziger Jahre in die Literatur eingeführt (Louviere, 1984; Louviere & Gaeth, 1987).

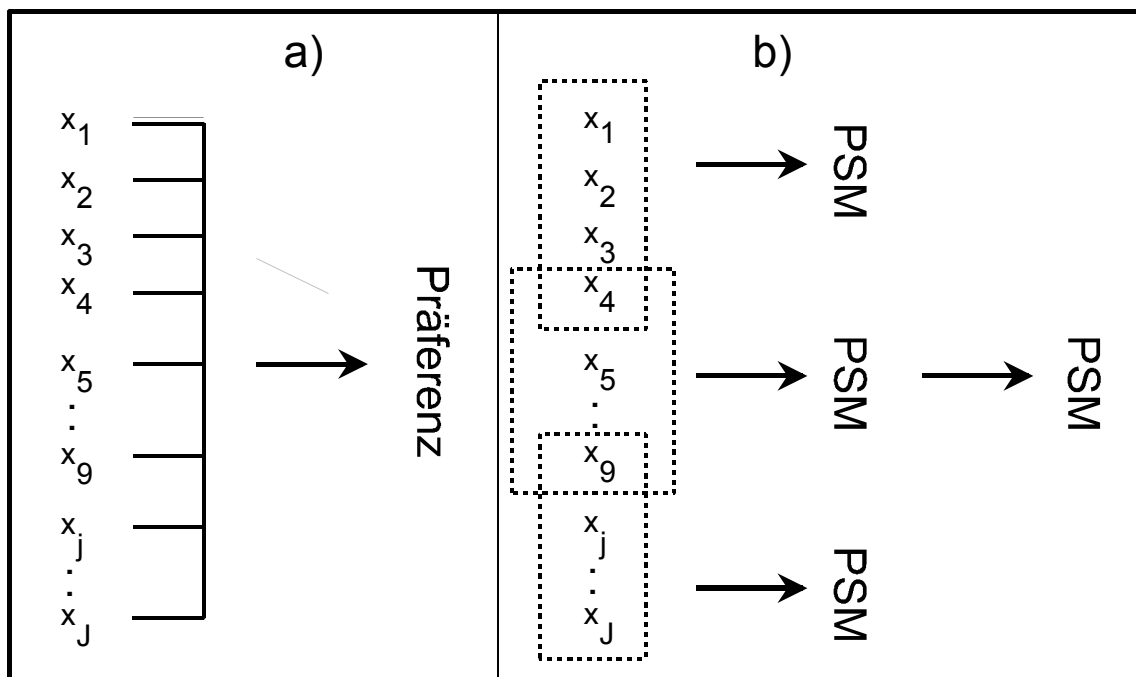


Abbildung 3.5: Der Präferenzbildungsprozeß (a) und die Vorgehensweise bei der Präferenzstrukturmodellierung (b) mit der 'Brücken-CA'. Die gestrichelt gezeichneten Kästen sind conjointanalytische Subuntersuchungen.

Der HICA-Ansatz basiert auf der theoretischen Grundlage der hierarchischen Informationsintegrationstheorie (HII, Louviere, 1984), welche sich als eine Erweiterung der Informationsintegrationstheorie (Anderson, 1981, 1982) versteht. Die HII nimmt an, daß bei komplexen Entscheidungsproblemen Menschen die entscheidungsrelevanten Attribute nach bestimmten Gesichtspunkten gruppieren. Das Ergebnis dieses Kategorisierungsvorganges sind sogenannte (höher geordnete) Entscheidungskonstrukte. Bei der Präferenzenbildung sollten die Objekte (z.B. Supermärkte) entsprechend des 'globalen Eindrucks' auf den Konstrukten (z.B. Preisniveau, Einkaufsbequemlichkeit, Produktqualität) einge-

schätzt werden. Jede Konstrukteinschätzung ist wiederum eine Funktion der zugehörigen Attribute (z.B. das Preisniveau besteht aus den Attributen: Lebensmittelpreise, Kleiderpreise, Parkgebühr, ...). Der Präferenzbildungsprozeß wird somit als hierarchisch geordnet angesehen.

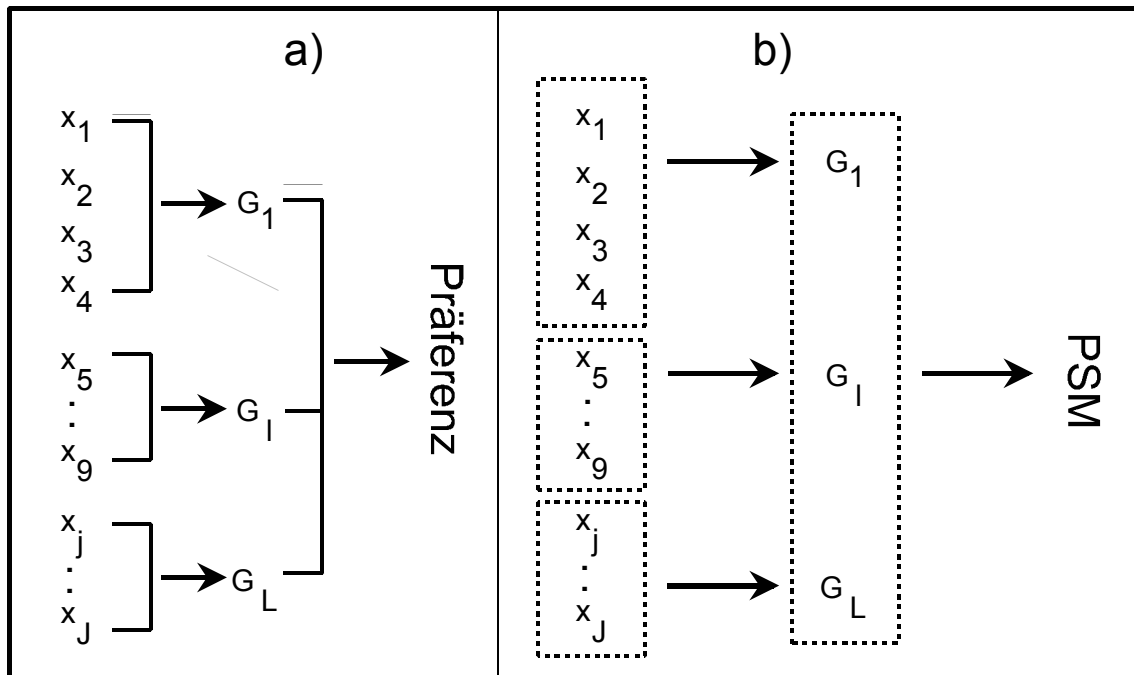


Abbildung 3.6: Der Präferenzbildungsprozeß (a) und die Vorgehensweise bei der Präferenzstrukturmodellierung (b) mit der HICA. Die gestrichelt gezeichneten Kästen sind conjoint-analytische Subuntersuchungen.

Die HICA folgt dieser theoretischen Sichtweise und erfordert zunächst die Gruppierung der Attribute in Konstrukte G_l mit $l=\{1\dots L\}$. Es wird für jedes Konstrukt eine separate CA durchgeführt, so daß die Wirkung der Attribute auf die Konstrukte modelliert werden kann. Danach wird eine CA mit den Konstrukten durchgeführt. In einem letzten Schritt können die Ergebnisse der einzelnen CA verknüpft werden. Eine HICA erfordert damit $(L + 1)$ einzelne CA. Die theoretische Vorstellung des Prozesses a) der Präferenzbildung und b) der Modellbildung ist in Abbildung 3.6 skizziert.

Louviere und Timmermans (1992) untersuchten die kriterienorientierte Validität des Verfahrens. In ihrer Untersuchung zur Vorhersage von Freizeitverhalten erhalten sie erste Hinweise auf die Validität des Verfahrens.

Oppewal, Louviere und Timmermans (1994) kritisieren diese zwei vorgenannten Ansätze und stellen eine erweiterte HICA vor. Die Autoren halten an der Annahme des hierarchischen Präferenzbildungsprozesses der einfachen HICA fest, aber schlagen eine Änderung der Präferenzstrukturmodellierung vor. Nach diesem Ansatz besteht jede CA-Subuntersuchung aus jeweils den Attributen eines übergeordneten Konstrukts und den jeweilig anderen Konstrukten. Abbildung 3.7 skizziert eine mögliche Subuntersuchung. Dadurch wird zwar die Größe der Subuntersuchungen geringfügig erweitert, auf die Brückenuntersuchung kann dafür aber verzichtet werden. Die Modellparameter können bei diesem Ansatz simultan geschätzt werden.

Das Design dieser Untersuchung und die Parameterschätzverfahren sind bei Oppewal et al. (1994) dargestellt. Die Autoren sehen den Hauptvorteil der erweiterten HICA gegenüber der einfachen HICA darin, daß die angenommene hierarchische Struktur der Präferenzbildung getestet werden kann. Sie stellen dazu vier Testverfahren vor.

Durch die ausführliche Darstellung einer empirischen Untersuchung zur Beurteilung von Einzelhandelssupermärkten illustrieren die Autoren ihre Vorgehensweise. Die Tests der hierarchischen Urteilsstruktur bei dieser Untersuchung deuten auf eine teilweise Verletzung der Annahmen hin. Als Validitätskriterium verwenden die Autoren den Marktanteil der Supermärkte (kriterienorientierte Validität auf aggregierter Ebene; vgl. Abschnitt 5.4.2). Die Produkt-Moment-Korrelation zwischen dem vorhergesagten und tatsächlichen Marktanteil liegt bei $r=0,42$. Vergleichende Studien zwischen der (erweiterten) HICA und anderen CA-Ansätzen liegen unseres Wissens bislang nicht vor.

Die HICA ist seit ihrer Einführung außerhalb der Forschergruppe um Louviere kaum genutzt worden. Dies dürfte einerseits daran liegen, daß der Ansatz noch vergleichsweise jung ist. Den Hauptgrund sehen wir aber darin, daß andere Alternativen zur klassischen CA, die ebenfalls mit vielen Attributen umgehen können, einfacher zu handhaben sind. Dies gilt insbesondere für die ACA, die als Softwarelösung vorliegt.

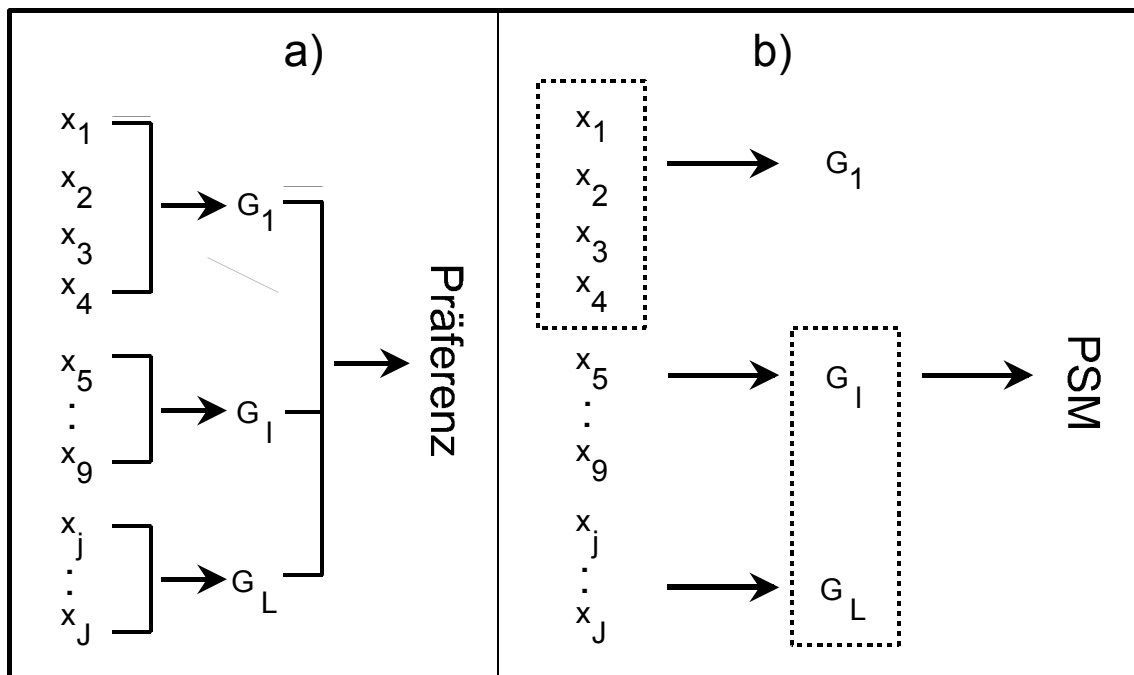


Abbildung 3.7: Der Präferenzbildungsprozeß (a) und die Vorgehensweise bei der Präferenzstrukturmodellierung (b) mit der erweiterten HICA. Die gestrichelt gezeichneten Kästen ist eine der drei conjointanalytischen Subuntersuchungen.

3.7.2.3 Hybrid-Conjoint Analyse (HCA)

Erste Ansätze zur HCA wurden von Green, Carroll und Goldberg (1981) und Green, Goldberg und Montemayor (1981) vorgestellt. Die Ausgangsmotivation für die Entwicklung dieser Alternative zur klassischen CA war, daß insbesondere bei kommerziellen Untersuchungen oft sehr viele Parameter zu schätzen sind, was den Datenerhebungsaufwand bei Anwendung der klassischen CA unverträglich macht. Die hohe Zahl der Parameter resultiert daher, daß kommerzielle CA meist viele Attribute berücksichtigen müssen, oder daß Attributwechselwirkungen modelliert werden sollen. Um diese schätzen zu können, müßten die Apn bei der klassischen CA komplexere bzw. mehr Objekte beurteilen.

Das Ziel der HCA besteht darin, daß trotz einer größeren Attributanzahl oder der zusätzlichen Modellierung von Wechselwirkungen der Datensammlungsaufwand je Ap vertretbar bleibt. Die Vorteile der klassischen CA, insbesondere die Bildung von PSM auf der individuellen Ebene und die Möglichkeit, die Modellanpassung auf individueller Ebene prüfen zu können,

sollen bei der HCA erhalten bleiben. Sie geht dabei nach folgendem Prinzip vor (Green, 1984):

1. Die HCA verwendet die kompositorische Nutzenschätzung als Ausgangspunkt zur Bildung eines individuellen PSM. Diese Art der expliziten Modellbildung ist in Abschnitt 3.7.1 ausführlich dargestellt. Die meisten veröffentlichten HCA setzten das Standardverfahren der expliziten Modellbildung ein (Abschnitt 3.7.1.1). Die Aufgaben bei kompositorischen Nutzenschätzverfahren sind im Vergleich zu den Beurteilungsaufgaben der klassischen CA für die Apn einfach und schnell zu erledigen.
2. Nach der expliziten Modellbildung bekommt jede Ap die Aufgabe, wenige Vollprofil-Objekte zu beurteilen (Rating, Rangreihung etc.). Die Anzahl der Beurteilungsobjekte liegt typischerweise zwischen 3-9 (Green & Srinivasan, 1990). Die Objekte sind einem sog. Masterdesign entnommen. Das Masterdesign ist meist ein reduziertes, orthogonales Haupteffekt-design (OHD) oder, bei der zusätzlichen Modellierung von Wechselwirkungen, ein reduziertes faktorielles Design, das die nicht konfundierte Schätzung der Haupteffekte und ausgesuchter Wechselwirkungen erlaubt. Jede Ap erhält ein speziell ausgesuchtes Unterset aller Objekte aus dem Masterdesign zur Beurteilung.
3. Aufgrund der expliziten Modelle oder auf der Basis von a priori Annahmen können Beurteilersegmente gebildet werden. Ein Segment ist eine homogene Beurteilergruppe. Auf Segmentebene werden aus den Vollprofilbeurteilungen dekompositorisch die Haupteffektparameter und evtl. Wechselwirkungsparameter geschätzt. Das individuelle explizite Modell wird bei jeder Ap durch das auf Segmentebene geschätzte Modell erweitert.

Ein einfaches Hybridmodell kann folgendermaßen dargestellt werden (vgl. Green, 1984; Green, Goldberg & Montemayor, 1981):

$$u_h(X_i) = a * v_h(X_i) + w(X_i) \quad (3.20)$$

mit

- $u_h(X_i)$ Gesamtnutzen des Objekts X_i bei der Ap h
 a Regressionsparameter zur Gewichtung der auf individueller und aggregierter Ebene geschätzten Funktionen
 $v_h(X_i)$ explizit modellierter Nutzen des Objekts X_i bei der Ap h
 $w(X_i)$ auf Gruppenebene dekompositorisch geschätzter Nutzen des Objekts X_i

Die Funktion v_h kann beispielsweise ein auf individueller Ebene explizit ermitteltes Teilnutzenwertmodell sein.

$$v_h(X_i) = \begin{array}{|c|c|} \hline & \\ \hline \end{array} b_{hjk} * x_{ijk} \quad (3.21)$$

Ein Segment bearbeitet gemeinsam ein Masterdesign. Die Funktion $w(X_i)$ wird auf Segmentebene geschätzt. Es kann sich dabei z.B. um ein Modell handeln, das Haupteffekte (lineare Vektormodelle) und Zweifachwechselwirkungen berücksichtigt.

$$w(X_i) = \begin{array}{|c|} \hline \\ \hline \end{array} b_j * x_{ij} + \begin{array}{|c|} \hline \\ \hline \end{array} b_{j_1j_2} * x_{ij_1} * x_{ij_2} \quad (3.22)$$

Das Prinzip der HCA besteht darin, daß zunächst die explizite Modellbildung benutzt wird, um eine grobe Nutzenschätzung auf individueller Ebene vorzunehmen. Anschließend folgt eine dekompositorische Nutzenschätzung auf Segmentebene. Diese hat die Aufgabe, die individuellen expliziten PSM aufzuwerten. Nach Aufnahme des auf Gruppenebene geschätzten Modellteils in das PSM, sollte dies besser an die Objektbeurteilungen angepaßt sein, als das explizite Modell allein. Die HCA verbindet damit die Einfachheit der expliziten Modellbildung mit den Vorteilen der dekompositorischen Präferenzstrukturmodellierung.

Von Anwendungen der HCA wird vereinzelt in der Literatur berichtet (z.B. Tantiwong & Wilton, 1985; Toy, Rager & Guadagnolo, 1989). Es lassen sich nur sehr schwer gültige Aussagen über die Verwendungshäufigkeit der HCA abgeben. Die Untersuchungen zur kommerziellen Anwendung der CA von Cattin

und Wittink (1982) und Wittink und Cattin (1989) erwähnen die HCA zwar kurz im Text, die Autoren spezifizieren aber deren Verwendungshäufigkeit nicht. Wittink et al. (1994) listen unter der Rubrik 'stimulus construction' die Kategorie 'other (hybrid)' auf. 10% der CA-Untersuchungen werden ihr zugeordnet. Wie viele davon auf die HCA entfallen, muß jedoch offen bleiben. Allgemein kann jedoch festgehalten werden, daß die HCA eher selten zum Einsatz kommen dürfte. Obwohl sie vergleichsweise einfach handhabbar ist, dürfte der Grund hauptsächlich in der heterogenen Befundlage vergleichender Studien (vgl. Abschnitt 3.7.3) und in der einfacheren Handhabbarkeit des konkurrierenden Verfahrens ACA liegen. Darüber hinaus sehen wir Mängel in der konzeptuellen Klarheit des Ansatzes, bei dem individuelle Nutzenfunktionen mit solchen auf aggregierter Ebene vermischt werden. Die Begründung für diese Vorgehensweise ist rein pragmatisch und auf die kriterienorientierte Validität gerichtet. Dennoch ist (theoretisch und einem Auftraggeber) nur schwer erklärbar, warum individuelle Nutzenschätzungen durch Gruppenschätzwerte verbessert werden können bzw. wenn dies der Fall ist, warum die Modelle nicht gleich auf Gruppenebene gebildet werden.

3.7.2.4 Adaptive Conjoint Analyse (ACA)

Das Verfahren der adaptiven Conjoint Analyse wurde Mitte der achtziger Jahre entwickelt (Johnson, 1987a; Sawtooth Software, 1985). Es liegt als Softwaresystem vor und wurde seither mehrfach weiterentwickelt. Die derzeit neuste Version, ACA 4.0 (Sawtooth Software, 1994) soll hier eingehend vorgestellt und vor dem Hintergrund empirischer Untersuchungen besprochen werden. Vergleichende Untersuchungen von ACA und anderen Methoden der dekompositorischen Nutzenmessung werden in Abschnitt 3.7.3 diskutiert.

ACA ist insofern unter den CA-Verfahren einzigartig, als daß die gesamte Datenerfassung in Form eines computergestützten Interviews vor sich geht. Der Untersuchungsaufbau wird vom Anwender menügestützt dem Programm vorgegeben. Anschließend erstellt das System ein Interviewprogramm, das die Befragung der Apn steuert. Nach dem Interview werden ein ausführliches Interviewprotokoll und die geschätzten TNW (additives Teilnutzenwertmodell) für jede Apn abgespeichert. Die Untersuchungsdaten der Apn können dann, sofern notwendig, vom Anwender zusammengeführt und dem System zur weiteren

Datenanalyse vorgegeben werden. ACA 4.0 besitzt die Möglichkeit, TNW nach verschiedenen Regeln zu normieren und über die Apn zu mitteln. Es enthält außerdem mehrere Wahlanteilsimulatoren zur Weiterverarbeitung der Untersuchungsergebnisse auf aggregierter Ebene (vgl. Abschnitt 4.2). Die Bedienerfreundlichkeit des Systems ist aus Sicht des Anwenders und der befragten Apn hoch. Aufgrund dieser Merkmale ist das System für CA-Untersuchungen außerordentlich handhabbar und nützlich (Carmone, 1987).

Die Systemgrenzen von ACA erlauben die Berücksichtigung von bis zu 30 neunstufigen Attributen. Diese Maximalzahl kann natürlich nicht sinnvoll in empirischen Untersuchungen verwendet werden. Trotzdem ist ACA in der Lage, mit relativ vielen Attributen/ Stufen umzugehen, indem es bei der Nutzenschätzung zunächst ähnlich vorgeht, wie die HCA. ACA ist in vier Phasen aufgeteilt. Die Phasen I, II, III und IV werden nachfolgend ausführlich dargestellt. In den Phasen I und II wird eine kompositorische Nutzenschätzung vorgenommen. Die Verbesserung dieser initialen Nutzenschätzung findet dann in der dekompositorischen Phase III statt. Anders als bei HCA nimmt ACA jedoch keine Nutzenschätzung auf Gruppenebene, sondern auf individueller Ebene vor. Phase IV dient danach der optimalen Verknüpfung der kompositorischen und dekompositorischen Nutzenschätzung.

Phase I: Diese Phase dient dazu, Präferenzen bzgl. der Stufen innerhalb der Attribute zu erfassen. Zunächst wird nacheinander jedes Attribut mit seinen Stufen dargestellt. Die Apn erhalten die Aufgabe, eventuelle nicht akzeptierbare Attributstufen zu kennzeichnen. Sind solche vorhanden, deutet dies streng genommen auf einen konjunktiven Urteilsprozeß hin. ACA modelliert aber auch diese nicht akzeptierbaren Attributstufen kompensatorisch. Das Programm ordnet ihnen einen sehr niedrigen (benutzerdefinierten) TNW zu und verwendet diese Stufen im weiteren Verlauf des Interviews nicht mehr.

Diese Vorgehensweise verfolgt zwei Ziele. Erstens soll ein eventuell vorhandener konjunktiver Urteilsprozeß durch das PSM approximiert werden und zweitens soll das Interview durch die folgende Nichtberücksichtigung nicht akzeptierbarer Attributstufen verkürzt werden (vgl. Abschnitt 3.4.3 und Dorsch & Teas, 1992; Green, et al., 1988; Olshavsky & Acito, 1980).

Auf die optionale Abfrage nicht akzeptierbarer Attributstufen folgt für jedes Attribut getrennt die Frage nach der Präferenz bzgl. seiner Stufen. Hierzu stehen zwei Antwortmodi zur Verfügung: Die Rangreihung und das Rating.

Wählt der Anwender die Rangreihung, dann bringt die Ap die Stufen eines Attributs gemäß ihrer Präferenz in eine Rangfolge. Beim Rating gibt die Ap das Ausmaß ihrer Präferenz für jede Attributstufe gesondert auf einer maximal neunstufigen Skala ab.

Um die initialen TNW zu schätzen, werden die numerischen Präferenzrangwerte (z.B. 2; 1; 3) zunächst umgedreht, so daß die am höchsten präferierte Attributstufe den höchsten Wert besitzt (2; 3; 1) und anschließend auf den Range 1 normiert und den Mittelwert 0 zentriert (0; 0,5; -0,5). Bei Ratingwerten wird analog vorgegangen. In diesem Fall können die initialen Werte auch nicht äquidistant sein. Auch die Abfrage der Präferenz innerhalb der Attribute ist optional und kann weggelassen werden, wenn diese für alle Apn a priori feststeht (z.B. bei Preis). Es ist zu vermuten, daß die realitätsgerechte Annahme einer a priori-Ordnung der Intraattributpräferenz zu einer Validitätssteigerung des geschätzten PSM führt (van der Lans, Wittink, Huber & Vriens, 1992; Srinivasan, Jain & Malhotra, 1983).

Besitzt ein Attribut mehr als fünf Stufen, dann verwendet das Programm ab Phase III nur die fünf Stufen mit der höchsten Präferenz. Vermutet der Anwender, daß die fünf höchstpräferierten Attributstufen nicht die fünf am höchsten beurteilungs- und entscheidungsrelevanten Attribute sind, kann als nächstes abgefragt werden, in welchem Attributstufenbereich die Ap am wahrscheinlichsten Entscheidungen fällt⁵.

Phase II: Diese Phase dient dazu, die Attributwichtigkeiten zu erfassen. Die relative Wichtigkeit eines Attributs wird dabei als die Differenz zwischen der am niedrigsten und am höchsten präferierten Attributstufe definiert. Die Ap erhält z.B. die Schilderung, daß sich zwei sonst identische Autos nur dadurch unterscheiden, daß das eine 10000 DM und das andere 30000 DM kostet. Sie wird befragt, wie wichtig ihr dieser Unterschied ist. Die relativen Wichtigkeiten der Attribute werden nacheinander so erfaßt. Die Antwort wird auf einer verankerten Ratingskala [1; 4] abgegeben.

⁵ Dieser Fall kann z.B. eintreten, wenn bei Autos das Attribut 'Preis' folgende sechs Stufen besitzt: '10000 DM', '14000 DM', '18000 DM', '22000 DM', '26000 DM' und '30000 DM'. Die Präferenzrangreihe einer jeden Apn wird vernünftigerweise so aussehen, daß sie das billigste Auto (unter sonst gleichen Bedingungen) am höchsten präferiert und das teuerste Auto am niedrigsten. ACA würde die dekompositorische Parameterfeinschätzung in diesem Fall nicht für die Attributstufe '30000 DM' vornehmen. Dies könnte für eine solvente Apn jedoch ein Fehler sein, weil sie insbesondere im oberen Preisbereich Entscheidungen treffen wird. Eine solche Apn würde auf die Frage, die fünf Stufen anzugeben, in deren Bereich sie am wahrscheinlichsten Entscheidungen fällt, die Stufen '14000 DM'

Die initialen TNW aus Phase I werden mit dem jeweiligen Ratingwert der relativen Wichtigkeit multipliziert. Daraus ergibt sich innerhalb eines Attributs eine Streubreite der TNW in Höhe der relativen Wichtigkeit (bei einer relativen Wichtigkeit von 3 ergeben sich für das obige Beispiel die TNW 0; 1,5; -1,5).

Um die Interviewdauer zu begrenzen, kann der Anwender dem Programm vorgeben, wie viele Attribute in Phase III und IV verwendet werden sollen. Sind mehr Attribute in der Untersuchung vorhanden, als hier spezifiziert, wählt ACA die individuell wichtigsten Attribute aus Phase II für nachfolgende Interviewteile aus.

Phase I+II zusammen bilden ein Verfahren der kompositorischen Nutzenmessung. Die geschätzten TNW aus diesen Phasen nennen wir *explizite TNW* (im Begriffssystem des Programmherstellers: prior utilities). Die expliziten TNW b_{phjk} der $A_p h$ berechnen sich bei Rangreihung der Präferenzen innerhalb der Attribute durch

$$b_{hjk} = w_{hj} [(K_j - r(x_{hjk})) / (K_j - 1) - 0,5] \quad (3.23)$$

mit

- K_j Anzahl der Stufen des Attributs j
- w_{hj} Wichtigkeitsrating des Attributs j bei der $A_p h$
- $r(x_{hjk})$ Rangplatz der Attributstufe k des Attributs j bei der $A_p h$

Die expliziten TNW der $A_p h$ bilden den Vektor \mathbf{b}_{ph} mit den Elementen b_{phjk} und über alle $A_p n$ einer Untersuchung die Matrix \mathbf{P} .

Das kompositorische ACA-Verfahren ist der konjunktiv-kompensatorischen expliziten Modellbildung (Srinivasan, 1988) insofern ähnlich, als es nicht akzeptierbare Attributstufen zuläßt und die relative Attributwichtigkeit über die Differenz zwischen der am niedrigsten und am höchsten präferierten Attributstufe definiert. Es gibt jedoch auch deutliche Unterschiede. Diese liegen insbesondere in der besonderen Art der Aufgabeninstruktion des Srinivasan-Verfahrens und darin, daß die Attribute bei der Abfrage der Attributwichtigkeit beim ACA-Verfahren sequentiell dargeboten werden. Der überwiegende Teil der anderen kompositorischen Verfahren bietet die Attribute bei der Wichtigkeits-einschätzung simultan dar und fördert eine simultane Betrachtung meist noch durch die Verwendung von Konstant-Summen-Antwortmodi (vgl. Abschnitt

bis '30000 DM' angeben. Dies führt dazu, daß die TNW genau dieser Stufen in Phase III

3.7.1). Es ist zu vermuten, daß die sequentielle Darbietungsform sehr anfällig für Verzerrungen durch Ankereffekte ist. Die Ziele der Phasen I+II lassen sich folgendermaßen zusammenfassen:

1. Eventuelle Approximation eines konjunktiven Urteilsprozesses und Reduktion der Komplexität des Interviews durch die Identifikation nicht akzeptierbarer Attributstufen.
2. Identifikation der am höchsten determinanten Attribute im Attributset und der am höchsten urteilsrelevanten Attributstufen. Diese Information kann bei der dekompositorischen Nutzenschätzung zu einer eventuellen Verminderung der Attribute im Attributsets und einer Verringerung der Attributstufen herangezogen werden (vgl. auch ähnliche 'tailored-CA-Ansätze' durch die Verwendung individuell wichtiger Attribute, z.B. Nickerson et al., 1990; Schweikl, 1985).
3. Parametervorschätzung des PSM durch die ACA-explizite Modellbildung (explizite TNW).

Da die kompositorische Nutzenschätzung von ACA sich konzeptionell von den anderen Verfahren der expliziten Modellbildung unterscheidet, ist es auch sinnvoll, ihr einen eigenen Namen zu geben. Wir nennen dieses Verfahren die *ACA-explizite Modellbildung*.

Von einigen systematischen Vergleichen der ACA-expliziten Modellbildung mit dem Standardverfahren der expliziten Modellbildung (Abschnitt 3.7.1.1) wurde bisher berichtet. Green, Krieger und Agarwal (1991) untersuchten in ihrer Studie zur Beurteilung von Studentenappartments u.a. die kriterienorientierte Validität dieser beiden Verfahren der expliziten Modellbildung. Als Kriterium dienten Holdout-Objekte, als Maßzahlen der Validität wurden die Rangkorrelation zwischen vorhergesagter und tatsächlicher Rangfolge, der Prozentsatz der Erste-Wahl-Hits und der Prozentsatz der vorhergesagten richtigen Rangplätze verwendet (diese Konzepte der Validierung sind ausführlich in Abschnitt 5.4 erklärt). Das Standardverfahren zeigte eine höhere Validität als das ACA-explizite Modellbildungsverfahren auf allen Kriterienmaßzahlen. Die Validität war sogar höher, als die des gesamten ACA-Verfahrens und das, obwohl das Standardverfahren der expliziten Modellbildung eine geringere Interviewzeit in Anspruch nahm (zur Kritik der Untersuchung von Green, Krieger

& Agarwal, 1991 siehe Johnson, 1991; z.B. Konfundierung des Nutzenmeßverfahrens mit der Reihenfolge der Verfahren). Green, Krieger und Agarwal (1991) attribuieren diesen Befund u.a. auf eine zu grobe Abstufung der Antwortskala bei der Abfrage der Attributwichtigkeit (4-Punkte Ratingskala) und schlagen aus diesem Grunde vor, die Skalen der ACA-expliziten Modellbildung feiner abzustufen (was bei ACA 4.0 auch als Option umgesetzt wurde). McLauchlan (1991) überprüft die Hypothese von Green, Krieger und Agarwal (1991), daß eine feinere Skalenabstufung bei der Attributwichtigkeiterfassung zu einer höheren Validität der ACA-expliziten Modellbildung führt. Er manipuliert experimentell die Art der expliziten Modellbildung (4-Punkte Wichtigkeitsrating vs. 9-Punkte Wichtigkeitsrating + jeweils wenige Paarvergleiche) in seinem Experiment zur Beurteilung von Lebensmittelgeschäften (10 Attribute). Er findet keinen Unterschied hinsichtlich der kriterienorientierten Validität gegenüber Holdout-Konzepten.

Damit ist sehr fraglich, ob Green, Krieger und Agarwal (1991) das Problem der ACA-expliziten Modellbildung im Kern getroffen haben. Wir vermuten, dieses liegt weniger in der Skalenabstufung, als vielmehr in der sequentiellen Darbietung der Attribute bei der Wichtigkeitseinschätzung.

Phase III: In Phase III werden den Apn ausgewählte Teilprofilpaarvergleiche zur Beurteilung vorgegeben. Aufgrund der Antworten auf einer abgestuften Paarvergleichsskala (Abbildung 3.4) werden die initialen Parameterschätzungen nach jeder Antwort angepaßt.

Welche Paare den Apn zur Beurteilung vorgelegt werden, hängt von den zuvor gegebenen Antworten ab. Das adaptive Paarvergleichsdesign wurde in Abschnitt 3.5.3.3 vorgestellt und diskutiert. Abschnitt 3.8.2 beschreibt den Algorithmus zur Schätzung der Modellparameter. Die TNW, die nur in Phase III geschätzt werden, nennen wir *Paarvergleichs-TNW*. Die Paarvergleichs-TNW bilden den Vektor \mathbf{b}_{qh} und über alle Apn die Matrix \mathbf{Q} . Nach Phase III liegen TNW vor, die auf der Basis der expliziten TNW im Laufe der Paarvergleichsphase III verfeinert geschätzt wurden. Die Gesamtschätzung nach dem letzten Paarvergleich nennen wir die *nichtreskalierten TNW*. Diese TNW bilden den Vektor \mathbf{b}_{sh} und über alle Apn die Matrix \mathbf{S} .

Phase IV: In Phase IV werden den Apn ausgewählte Objektbeschreibungen vorgegeben (*'calibrating concepts'*). Das Programm erlaubt eine Spezifikation dieser Objekte auf maximal acht Attributen. Werden in einer Untersu-

chung nicht mehr als acht Attribute verwendet, beurteilen die Apn hier in aller Regel Vollprofile. Die Beurteilung wird auf einer Prozentrangskaala (0-100) vorgenommen (meist als Wahl-/ Kaufwahrscheinlichkeitseinschätzung: Wie wahrscheinlich ist es, daß Sie dieses Produkt kaufen würden? Antwort von 0% bis 100%). Man beabsichtigt damit eine Antwort auf Verhältnisskalenniveau, nachdem eine geeignete Transformation Decken- und Bodeneffekten ausgeglichen hat. ACA verwendet hierzu die Logit-Transformation. Die calibrating concepts dienen in erster Linie dazu, TNW so zu reskalieren, daß die Gesamtnutzenwerte der Objekte Logits sind. Damit sollte für jedes beliebige Objekt die Wahl-/ Kaufwahrscheinlichkeit vorhergesagt werden können. Diese TNW nennen wir *reskalierte TNW*. Sie bilden den Vektor \mathbf{b}_{uh} und über alle Apn die Matrix \mathbf{U} . ACA 4.0 bietet zwei Möglichkeiten, diese Reskalierung vorzunehmen:

1. Die expliziten TNW und die Paarvergleichs-TNW werden *gleichgewichtet* aufsummiert. Die Regressionsparameter a_{1h} und a_{2h} werden per OLS-Regression auf individueller Ebene so geschätzt, daß die reskalierten TNW die Logits der Wahlwahrscheinlichkeiten bestmöglich vorhersagen.

$$\ln [P_{ih} / (1 - P_{ih})] \approx a_{1h} + a_{2h} (\boxed{} \boxed{} b_{phjk} x_{ijk} + \boxed{} \boxed{} b_{qhjk} x_{ijk}) \quad (3.24)$$

mit

P_{ih} angegebene Wahl-/ Kaufwahrscheinlichkeit des Objekts X_i durch Ap h

a_{1h}, a_{2h} zu schätzende Regressionsparameter zur TNW-Reskalierung

b_{phjk} explizite TNW der Ap h für Attributstufe k auf Attribut x_j

b_{qhjk} Paarvergleichs-TNW der Ap h für Attributstufe k auf Attribut x_j

$x_{ijk} = \begin{cases} 1 & \text{falls Objekt } X_i \text{ die Attributausprägung k auf Attribut } x_j \text{ besitzt} \\ 0 & \text{in anderen Fällen.} \end{cases}$

Um die zwei Regressionsparameter zu schätzen und eine Maßzahl für den Modellfit zu erhalten, ist bei dieser Vorgehensweise die Beurteilung von mindestens drei calibrating concepts notwendig.

2. Die expliziten TNW und die Paarvergleichs-TNW werden *verschieden gewichtet* aufsummiert. Die Regressionsparameter a_{1h} , a_{2h} und a_{3h} werden per OLS auf individueller Ebene so geschätzt, daß die reskalierten TNW die Logits der Wahlwahrscheinlichkeiten bestmöglich vorhersagen.

$$\ln [P_{ih} / (1-P_{ih})] \approx a_{1h} + a_{2h} \left(\boxed{\phantom{b_{phjk}}} \boxed{\phantom{b_{phjk}}} b_{phjk} x_{ijk} \right) + a_{3h} \left(\boxed{\phantom{b_{qhjk}}} \boxed{\phantom{b_{qhjk}}} b_{qhjk} x_{ijk} \right) \quad (3.25)$$

mit

P_{ih} angegebene Wahl-/ Kaufwahrscheinlichkeit des Objekts X_i durch $A_p h$

a_{1h} , a_{2h} , a_{3h} zu schätzende Regressionsparameter zur Reskalierung der TNW. a_{2h} gewichtet die expliziten TNW, a_{3h} die Paarvergleichs-TNW.

b_{phjk} explizite TNW der $A_p h$ für Attributstufe k auf Attribut x_j

b_{qhjk} Paarvergleichs-TNW der $A_p h$ für Attributstufe k auf Attribut x_j

$x_{ijk} = \begin{cases} 1 & \text{falls Objekt } X_i \text{ die Attributausprägung } k \text{ auf Attribut } x_j \text{ besitzt} \\ 0 & \text{in anderen Fällen.} \end{cases}$

Um die drei Regressionsparameter zu schätzen und eine Maßzahl für den Modellfit zu erhalten, ist bei dieser Vorgehensweise die Beurteilung von mindestens vier calibrating concepts notwendig.

In Phase IV wird zuerst die Wahlwahrscheinlichkeit des Objekts mit dem niedrigsten Nutzen abgefragt. Dann folgt das Objekt mit dem höchsten Nutzen. Der Nutzen der weiteren Objekte liegt zwischen diesen Extrema, und ihre expliziten TNW und Paarvergleichs-TNW sind sehr unterschiedlich. Durch die Darbietung der ersten beiden Extremobjekte kann die A_p Anker für die Beurteilung bilden. Darüber hinaus wird die Parameterschätzung der logistischen Regression durch die hohe Nutzenvarianz der Beurteilungsobjekte effektiver. Die Phase IV verfolgt folgende Ziele:

1. Die Gesamtnutzenwerte jedes beliebigen Objekts sind Logits. Dadurch kann jeder Objektnutzen in einen Wahl-/ Kaufwahrscheinlichkeitsprozentwert umgerechnet werden.

$$P(X_i) = e^{u(X_i)} / (1 + e^{u(X_i)}) \quad (3.26)$$

2. Bei Anwendung der 2. Methode der Reskalierung erhält man für jede Apn durch die verschiedene Gewichtung der expliziten TNW und der Paarvergleichs-TNW das Präferenzstrukturmodell, das die Beurteilungen der calibrating concepts am besten vorhersagt. Unter der Annahme, daß die Wahl-/ Kaufwahrscheinlichkeitsurteile valide sind, sollte die Validität des reskalieren PSM zunehmen.
3. Die Apn reagieren auf die Objektklasse verschieden sensibel. Beispielsweise könnte Ap 1 dem Objekt mit dem niedrigsten Nutzen eine Kaufwahrscheinlichkeit von 1% und dem Objekt mit dem höchsten Nutzen eine Kaufwahrscheinlichkeit von 20% zuweisen. Ap 2, die identische Vektoren b_{p2} und b_{q2} hat, gibt dagegen die Werte 1% bzw. 100% an. Durch die logistische Regression werden sich die reskalierten TNW b_{u1} und b_{u2} unterscheiden. Bei späteren Wahlanteilsimulationen auf aggregierter Ebene (vgl. Abschnitt 4.2) wird der wenig sensitiven Ap 1 ein niedrigeres Gewicht bei bestimmten Entscheidungsmodellen zugewiesen.
4. Für jede Ap kann der Anteil der erklärten Varianz des Logit-Regressionsmodells (R^2) berechnet werden. Diese Maßzahl gibt an, wie gut die Prozentwerturteile der Phase IV durch das reskalierte PSM (mit den TNW b_{uh}) vorhergesagt werden können. Auf diese Weise hat man ein Maß der Modellanpassung. Apn mit einer niedrigen Modellanpassung können z.B. bei der Datenanalyse gesondert behandelt oder von ihr ausgeschlossen werden.

Interessant ist, daß ACA zur Reskalierung der TNW in Phase IV nicht die letzte Parameterschätzung aus Phase III, also die nichtreskalierten TNW b_{sh} verwendet, sondern die (ungewichtete oder gewichtete) Summe aus den expliziten TNW b_{ph} und den Paarvergleichs-TNW b_{qh} . Die endgültigen PSM und alle weiteren Auswertungen innerhalb des Programms ACA basieren auf diesen reskalierten TNW.

Insbesondere die gewichtete Reskalierung der TNW aus den Phasen I-III durch die Ergebnisse der Beurteilungsaufgabe der ACA-Phase IV ist äußerst

problematisch, weil derzeit nicht klar ist, inwieweit das Wahlwahrscheinlichkeitsrating von Vollprofilen valide ist. Diese Frage wurde bisher noch nicht systematisch untersucht.

Die theoretische Begründung der Programmautoren (Sawtooth Software, 1994) für diese Vorgehensweise ist wenig überzeugend und argumentiert weitgehend mit der Vermeidung schlimmerer Übel. Sie basiert auf der Tatsache, daß die expliziten TNW und Paarvergleichs-TNW immer in irgendeiner Form verrechnet werden müssen. Green, Krieger und Agarwal (1991) kritisieren, daß bei keiner Reskalierung oder bei einer gleichgewichteten Reskalierung die relativen Beiträge der ACA-Teile von den numerischen Eigenschaften der Antwortskalen (z.B. Attributwichtigkeit von 1 - 4 und abgestufter Paarvergleich von -4 bis 4) abhängen. Die weitere Überlegung ist die, daß eine Verrechnung der ACA-Teile zwar nicht nach allgemeingültigen Regeln möglich ist (Johnson, 1991), aber trotzdem nicht willkürlich erfolgen sollte. Eine nichtwillkürliche Verrechnungsregel ist die individuell gewichtete Reskalierung der TNW auf der Grundlage der Konzeptbeurteilungen in Phase IV.

Eine interessante Frage ist, *inwieweit das reskalierte PSM nach den ACA-Phasen III+IV höher valide ist, als das explizite PSM nach Phase I+II*. Falls die Validität durch die Paarvergleiche zunimmt, stellt sich die Frage, *wie viele Paarvergleiche* in Phase III vorgenommen werden sollen. Johnson (1987b, zit. in Agarwal, 1989) führte eine Monte-Carlo-Simulation (8 Attribute, 2-5 Stufen) durch, um zu prüfen, in welchem Ausmaß ACA die Struktur eines vorgegebenen 'wahren PSM' wiedererkennen kann. Als Maßzahl der Strukturwiedererkennung verwendete er R^2 zwischen den wahren und den geschätzten TNW. In einem Simulationslauf variierte R^2 von 0,884 (= 11,4% nicht erklärte Fehlervarianz) für das ACA-explizite Modell (0 Paare), 0,931 (6,9%) bei 10 Paaren, 0,958 (4,2%) bei 20 Paaren bis zu 0,988 (1,2%) nach 50 Paaren. Jeweils 10 Paarvergleiche halbieren damit die nicht erklärte Fehlervarianz um etwas weniger als die Hälfte. Bei dieser Simulation wurde dem simulierenden Antworter kein Antwortfehler überlagert. In dieser fehlerfreien Simulation liefern die Paarvergleiche damit zusätzliche Information zum expliziten PSM. Die Frage ist, ob dies auch bei fehlerbehafteten Simulationen und bei empirischen Untersuchungen der Fall ist.

In einer empirischen Untersuchung fanden Wittink et al. (1992, gleichgewichtete Reskalierung), daß sich die expliziten TNW (Phase I+II) von den reskalierten TNW (nach Phase IV) signifikant unterscheiden. In der Studie von

Huber et al. (1993, gleichgewichtete Reskalierung) zeigen reskalierte PSM eine höhere kriterienorientierte Validität als explizite PSM. Dies bedeutet, daß der Aufwand der Phasen III+IV, aus der Perspektive der Validität betrachtet, lohnend ist.

Andererseits scheint die Validitätssteigerung des reskalierten PSM nach den Paarvergleichen gegenüber dem expliziten PSM nur gering zu sein (Green, Krieger & Agarwal, 1991; Green, Schaffer & Patterson, 1991; gleichgewichtete Reskalierung). Die Befunde zur kriterienorientierten Validität dieser beiden Untersuchung sind in Tabelle 3.6 dargestellt.

In der Studie von Finkbeiner und Platz (1986, zit. in Agarwal, 1989) erbringt ACA mit 8 Paarvergleichen eine geringfügig niedrigere kriterienorientierte Validität auf aggregierter Ebene, als ACA mit 16 Paarvergleichen.

Tabelle 3.6: Die kriterienorientierte Validität der Studien von Green, Krieger und Agarwal (1991) und Green, Schaffer und Patterson (1991) im Überblick. Angegeben ist die Rangkorrelation, der Prozentsatz von Erste-Wahl-Hits und der Prozentsatz richtig vorhergesagter Rangplätze zu einem Kreuzvalidierungsobjektset (jeweils Mittelwert und Standardfehler).

	Green, Krieger & Agarwal (1991)			Green, Schaffer & Patterson (1991)			
	16 Holdout-Objekte; n=170			12 Holdout-Objekte; n=96			
	0 Paare	8 Paare	15 Paare	0 Paare	7 Paare	14 Paare	20 Paare
Rangkorrelation							
Mittelwert	0,543	0,600	0,618	0,229	0,236	0,245	0,255
SE	0,029	0,019	0,014	0,019	0,018	0,018	0,017
Erste-Wahl-Hits (%)							
Mittelwert	58,4	60,4	62,0	25,5	27,4	32,4	32,3
SE	3,780	3,751	3,723	4,448	4,552	4,777	4,769
Rangpositionshits (%)							
Mittelwert	16,0	16,7	17,1	14,3	15,1	15,5	15,9
SE	2,812	2,861	2,905	3,573	3,654	3,694	3,732

Differenzierter beleuchtet Agarwal (1988b) diese Thematik. Er kommt zu dem Schluß, daß die kriterienorientierte Validität (individuelle Ebene) schwach aber signifikant mit der Anzahl der Paarvergleiche zunimmt, wenn es sich um Paarvergleichsobjekte der Profilstärke zwei handelt. Dieser Zusammenhang ist nicht bei Dreierprofilen vorhanden (vgl. Abschnitt 3.5.3.3).

Wie bereits mehrfach angedeutet, ist die optimale Anzahl der Paarvergleiche nicht unabhängig von der *Profilstärke der Paare* in Phase III. Diese

Frage wurde bereits auf dem Hintergrund verfügbarer Forschungsbefunde in Abschnitt 3.5.3.3 im Zusammenhang mit der Besprechung der Paarvergleichs-differenzendesigns diskutiert.

Weitere Forschungsfragen treten in Zusammenhang mit der Reskalierung der TNW in Phase IV auf. Insbesondere wäre interessant zu wissen, *wie sich die Validität eines nicht reskalierten PSM (mit Teilnutzenwerten, die nach dem letzten Paarvergleich in Phase III geschätzt wurden) gegenüber der Validität eines gleichgewichtet reskalierten PSM und eines nicht gleichgewichtet reskalierten PSM (Phase IV) verhält*. Der Unterschied zwischen den nicht reskalierten TNW \mathbf{b}_{sh} und den reskalierten TNW \mathbf{b}_{uh} besteht darin, daß \mathbf{b}_{sh} die Parameterschätzung nach dem letzten Paarvergleich ist, während die expliziten TNW \mathbf{b}_{ph} und die Paarvergleichs-TNW \mathbf{b}_{qh} getrennt in die reskalierten TNW \mathbf{b}_{uh} eingehen. Werden die \mathbf{b}_{ph} und \mathbf{b}_{qh} gewichtet verrechnet (gewichtete Reskalierung), dann üben die wenigen Konzeptbeurteilungen der Phase IV einen sehr großen Einfluß auf die Endergebnisse der Präferenzstrukturmodellierung aus. Dies ist insbesondere deswegen problematisch, da die Phase IV am Ende des Interviews durchgeführt wird, und die Auskunftsmotivation der Apn bereits auf einem niedrigen Niveau angelangt sein könnte. Wir schließen aus der Bemerkung (Sawtooth Software, 1994),

If the two regression coefficients [gemeint sind die Regressionskoeffizienten der logistischen Regression, die die Gewichte der expliziten TNW und der Paarvergleichs-TNW bestimmen; Anm. des Autors] differ too widely we assume the estimation is faulty and use more conservative values. (H-2)

daß die Programmautoren dieses Problem durchaus sehen und deshalb Vorsichtsmaßnahmen in ACA 4.0 eingebaut haben. Eine logisch stringente Vorgehensweise ist dies nach unserem Dafürhalten jedoch nicht. Veröffentlichte empirische Untersuchungen, die sich mit dieser Forschungsfrage beschäftigen, sind uns nicht bekannt.

Die *Verwendungshäufigkeit von ACA* war in den letzten Jahren sehr hoch. Nachdem in der ersten Hälfte der achtziger Jahre in den USA schwerpunktmäßig der Vollprofilansatz der CA zum Einsatz kam (Wittink & Cattin, 1989), war ACA in den Jahren 1986-1991 bei kommerziellen conjointanalytischen Studien in Europa mit 42% relativer Anwendungshäufigkeit die dominierende Methode (Wittink et al., 1994). Trotz fehlender kontrollierter Untersuchung ist zu vermuten, daß diese Entwicklung auch in den USA stattgefunden hat.

3.7.3 Vergleichende Diskussion der Verfahren hinsichtlich ihrer Validität

Es gibt eine Vielzahl von Simulationen und empirischen Untersuchungen, die die Validität der klassischen CA bestätigen. Wir sehen diese als gut bestätigt an und widmen diesen Abschnitt vergleichenden Untersuchungen zwischen klassischen Ansätzen der CA und HCA bzw. ACA.

3.7.3.1 HCA gegenüber der expliziten Modellbildung und klassischen CA

Einige Untersuchungen vergleichen die HCA mit der klassischen CA und expliziten Modellen. In der bereits erwähnten Untersuchung (Abschnitt 3.7.1.4) von Akaah und Korgaonkar (1983) finden die Autoren eine ähnliche kriterienorientierte Validität für die CA und die HCA. Die CA schneidet dabei geringfügig aber statistisch nicht signifikant besser ab. Beide Arten der Nutzenschätzung sind kompositorischen Verfahren überlegen. Cattin et al. (1982; vgl. 3.7.1.4) berichten von einer höheren kriterienorientierten Validität der CA gegenüber der HCA. In dieser Studie zeigte die HCA keine Überlegenheit gegenüber expliziten Modellen, sie schnitt tendenziell sogar schlechter ab. Green et al. (1983; vgl. 3.7.1.4) erhalten eine vergleichbare Validität für die CA und die HCA, wobei die Validitätskennwerte der HCA tendenziell sogar höher sind. Doch auch wenn die Vorhersagefähigkeit als gleich interpretiert wird, kann man dies als Erfolg für die HCA werten, da die Apn bei der HCA relativ einfache Urteile über uniattributive Objekte abgeben und nur wenige multiattributive Objekte zu beurteilen haben. Die Beurteilungsaufgabe der HCA kann deshalb für die Apn als einfacher angesehen werden als die der CA, bei welcher Präferenzurteile über eine größere Zahl von multiattributiven Objekten notwendig sind. In der Studie von Timmermans (1987) zum Freizeitverhalten zeigt die klassische Vollprofil-CA eine höhere kriterienorientierte und prognostische Validität als die HCA und explizite Modelle.

Von einer vergleichsweise ausführlichen experimentellen Untersuchung zum Vergleich der klassischen CA, der HCA und der expliziten Modellbildung (Standardverfahren, gewichtet und ungewichtet) berichten Moore und Semenik (1988). In dieser Studie zur Beurteilung von Autos manipulierten die Autoren die Komplexität der Beurteilungsaufgabe (5 bis 12 Attribute, die Anzahl der Vollprofilbeurteilungsobjekte bei der HCA und die Anzahl der Apn, über welche beim dekompositorischen Teil der HCA die Parameterschätzung vorgenommen

wurde. Die kriterienorientierte Validität der geschätzten bzw. expliziten Modelle gegenüber Kaufwahrscheinlichkeitsratings von 16 Kreuzvalidierungsobjekten ist bei allen Modellen recht hoch. Die klassische CA zeigt eine leichte Überlegenheit gegenüber den anderen Verfahren, die HCA besitzt eine geringfügig höhere Validität als explizite Modelle. Wird bei der HCA nur ein Teil der Stichprobe zur dekompositorischen Modellanpassung verwendet, verringert sich die Validität geringfügig. Eine Segmentierung der Stichprobe vor der dekompositorischen Parameterschätzung der HCA bringt bei dieser Untersuchung keine Validitätsvorteile. Dies liegt wahrscheinlich aber daran, daß die PSM der Stichprobe sehr homogen sind.

Ein sehr interessantes Ergebnis von Moore und Semenik (1988) ist, daß die Steigerung der Anzahl der Attribute (auf 12!) keine Vorteile für die HCA gegenüber der CA bringt. Im Gegenteil, die Vorhersagefähigkeit der HCA nahm relativ zur klassischen CA sogar ab. Dieser Befund ist nicht erwartungskonform (Cattin et al., 1982; Green, 1984), wird aber von den Autoren nicht weiter erklärt.

Es läßt sich zusammenfassend festhalten, daß die HCA höher valide Ergebnisse erbringt, als die explizite Modellbildung. Andererseits scheint sie hinsichtlich der kriterienorientierten Validität vergleichbare Ergebnisse zu erbringen wie klassische Ansätze der CA (Green & Srinivasan, 1990). Der Einsatz der HCA sollte gegenüber klassischen Verfahren dann lohnend sein (Cattin et al., 1982; Green, 1984), wenn

- ❑ die Beurteilungsaufgabe bei einer klassischen CA sehr schwierig durch viele oder komplexe Beurteilungsobjekte im Objektset ist. Diese Schwierigkeit entsteht bei einer klassischen CA-Untersuchung durch die Aufnahme vieler Attribute/ Attributstufen oder durch die Modellierung von Wechselwirkungen.
- ❑ eine Untersuchung viele Apn hat oder wenn die Apn in ihrem Antwortverhalten sehr homogen sind. Dadurch stehen viele Datenpunkte zur Schätzung der PSM auf Segmentebene zur Verfügung.
- ❑ in einem Objektbereich die Validität expliziter Modelle hoch ist.

3.7.3.2 ACA gegenüber der klassischen CA

In einer Monte-Carlo-Simulation vergleicht Agarwal (1988a) die Strukturwiedererkennung eines 'wahren PSM' durch die Vollprofil-CA mit den Parameter-

schätzprozeduren MONANOVA und LINMAP III (unconstrained und constrained) und durch ACA unter verschiedenen Fehlerbedingungen. Bei der klassischen CA dienten Modellantworten auf 16, 18, 27, 54 und 81 Vollprofile als Input. ACA wurden 27 Paarvergleiche vorgegeben. Als Maßzahl der Strukturwiedererkennung verwendete Agarwal die Korrelation zwischen den geschätzten und den wahren TNW. In der Nullfehlerbedingung entdeckte ACA die wahren TNW besser als die anderen Verfahren unter allen Objektanzahlbedingungen. Enthalten die 'wahren Antworten' Fehler, dann ist ACA den klassischen Verfahren immer überlegen, sofern diese die TNW auf der Grundlage von 18 oder weniger Stimuli schätzen. Haben die klassischen Verfahren 27 Stimuli zur Verfügung, leistet ACA z.T. bessere oder ungefähr gleiche Ergebnisse, bei mehr Stimuli (was bei realen Untersuchungen kaum umsetzbar ist) gibt sich ein gemischtes Bild. Bei dieser Simulation schneidet ACA allgemein sehr gut ab.

Die Frage ist, ob dies auch bei empirischen Untersuchungen der Fall ist. Finkbeiner und Platz (1986, zit. in Finkbeiner, 1988; zit. in Agarwal, 1989) vergleichen die kriterienorientierte Validität von PSM, die mittels einer klassischen Vollprofil-CA und ACA geschätzt wurden (Objektbereich Girokonten, 6 Attribute). Sie kommen zu dem Schluß, daß die durch ACA geschätzten PSM und deren kriterienorientierte Validität (auf individueller und aggregierter Ebene) gegenüber einer Kreuzvalidierungsstichprobe von Beurteilungsobjekten vergleichbar mit den PSM sind, die durch einen CA-Vollprofilansatz geschätzt wurden. Die Interviewzeit dauert mit ACA länger.

Agarwal (1988c, zit. in Agarwal, 1989) findet eine leichte, aber nicht signifikante Unterlegenheit von ACA gegenüber der Vollprofil-CA hinsichtlich der kriterienorientierten Validität. ACA dauerte jedoch etwas länger, als die Vollprofil-CA. Dafür wurde jedoch ACA als interessanter und weniger schwierig wahrgenommen, was wahrscheinlich dazu führte, daß sich die wahrgenommene Interviewdauer zwischen den beiden Verfahren nicht unterschied.

Huber et al. (1993) berichten von einer Studie von MacBride und Johnson (1979), in der sich eine geringe und statistisch nicht signifikante Überlegenheit eines Vorläufers von ACA bei der Vorhersage von Erste-Wahl-Hits gegenüber dem Trade-off-Ansatz der CA zeigte.

Huber, Wittink, Fiedler und Miller (1991) untersuchten Präferenzen gegenüber Kühlschränken. Ihr ausführliches Experiment hat zum Ziel, die Vollprofil-

methode der klassischen CA und ACA zu vergleichen. Bei jeder Ap wurde eine Nutzenmessung sowohl mit CA als auch mit ACA durchgeführt (Reihenfolge balanciert). Die klassische CA wurde ebenfalls computergestützt durchgeführt, was eine sequentielle Darbietung der Beurteilungsobjekte und die Antwort auf einer Ratingskala notwendig machte. Huber et al. nahmen bei der Hälfte der Apn fünf Attribute, bei der anderen Hälfte neun Attribute in die Untersuchung auf. Nach der CA und ACA wurde jeweils eine identische Kriteriumsaufgabe durchgeführt. Die Apn hatten dabei die Aufgabe, von zwei Vollprofilpaaren jeweils das Objekt anzugeben, das sie bei einem entsprechenden Angebot wählen würden. Ebenso sollten sie sich bei zwei Vollprofiltripeln entscheiden, welches Objekt sie am liebsten und welches sie am wenigsten wählen würden. Die Kriteriumsaufgabe enthält damit acht Paarvergleiche (2 explizite Paarvergleiche, je 3 implizite Paarvergleiche bei den Tripeln). Als Maßzahl der kriterienorientierten Validität der PSM verwendeten die Autoren den Anteil der richtig vorhergesagten Paarvergleiche. Durch die zweimalige Durchführung der Kriteriumsaufgabe kann deren Reliabilität geschätzt werden. Zu den Ergebnissen: ACA erbrachte insgesamt eine höhere Validität als die traditionelle CA (72,6% vs. 68,5% richtige Vorhersage der impliziten Paarvergleiche). Die Überlegenheit von ACA ist sowohl bei fünf als auch bei neun Attributen vorhanden. Sie ist besonders ausgeprägt, wenn nur die jeweils als erstes durchgeführten Nutzenmessungen verglichen werden (73% vs. 64%) und verschwindet beim Vergleich der zweiten Messung (73% vs. 74%). Dies deutet darauf hin, daß der Vollprofilansatz deutlich von einer Aufwärmphase (in diesem Fall durch ACA) profitiert (vgl. Abschnitt 3.5.1; Huber et al., 1993). In einem Fragebogen nach den jeweiligen Nutzenmessungen beurteilen die Apn ACA als etwas angenehmer und stimmten dem Item 'took too long to do' weniger zu. Insgesamt spricht die Untersuchung von Huber et al. (1991) für die Überlegenheit von ACA gegenüber der klassischen CA, und zwar sowohl bei neun Attributen als auch, wenn auch unerwartet, bei nur fünf Attributen. Es ist jedoch kritisch anzumerken, daß den Apn die Beurteilungsobjekte bei der Vollprofilmethode sequentiell dargeboten wurden. Dadurch sind Verzerrungen, z.B. durch Ankerbildung, zu erwarten. Diese Vermutung könnte auch erklären, warum die klassische CA dem Programm ACA nur unterlegen war, wenn diese als erste Interviewaufgabe vorgegeben wurde.

In einer differenzierten Analyse des bereits beschriebenen Experiments berichten Huber et al. (1993) von einer Überlegenheit des reskalierten PSM (nicht gewichtet) von ACA gegenüber der ACA-expliziten Modellbildung und einer klassischen Vollprofil-CA. Eine Kombination der Nutzenmeßmethoden (wie dies bei ACA ohnehin der Fall ist) ergab durchweg eine höhere Vorhersagesteigerung auf aggregierter Ebene. Vor dem Hintergrund der Vermutung, daß die verschiedenen Nutzenmeßverfahren unterschiedliche Aspekte des menschlichen Entscheidungsprozesses abbilden, empfehlen die Autoren eine multimodale Präferenzstrukturmodellierung.

3.7.4 Fazit

In den Abschnitten 3.7.2 und 3.7.3 haben wir die Verfahren der Nutzenmessung HICA, HCA und ACA als Alternativen zur klassischen CA vorgestellt und diskutiert. Obwohl es sich bei der HICA um eine sehr interessante Alternative zur Nutzenmessung handelt, kann ihre Leistungsfähigkeit und Handhabbarkeit derzeit noch nicht beurteilt werden. Die HCA konnte sich bisher nicht durchsetzen. Obwohl die Validität des Verfahrens positiv beurteilt wird, leidet die HCA an ihrer konzeptionellen Unklarheit und daran, daß keine spezielle Software vorliegt.

Ganz anders dagegen stellt sich der derzeitige Stand zur ACA dar. Eine umfassende Softwarelösung, die oft eingesetzt und kritisch überprüft wurde, erlaubt dem Anwender die vergleichsweise einfache Durchführung von Untersuchungen. ACA-Interviews werden von den Befragten als vergleichsweise angenehm und interessant beurteilt. Die PSM werden von ACA auf individueller Ebene geschätzt. Die Vorgehensweise verbindet die Vorteile der kompositorischen mit der dekompositorischen Nutzenschätzung. Im Gegensatz zu HCA bleibt die Konzeption von ACA klar, da individuelle explizite Modelle durch individuelle dekompositorische Modelle aufgewertet werden.

Vergleichende Studien belegen eine mindestens vergleichbare Validität des Verfahrens zur klassischen CA. Übersteigt die Anzahl der Attribute die derzeit geltenden Forschungsnormen, dann verhält sich ACA vermutlich wesentlich besser als die klassische CA bzw. das letztgenannte Verfahren ist dann überhaupt nicht mehr einsetzbar. Dasselbe ist in Objektbereichen zu erwarten, in denen sehr inhomogene Präferenzstrukturen herrschen.

Fiedler (1988) vergleicht die praktische Handhabbarkeit der CA durch die Trade-off-Methode und durch ACA. Er war an zwei CA-Untersuchungen zur Beurteilung von Eigentumswohnungen im Jahre 1972 (Trade-off) und 1988 (ACA) beteiligt. Er berichtet anekdotisch (Fiedler, 1988):

One of the most striking findings of the current study is the ease with which a conjoint measurement study can be fielded and the data analyzed. A process which took month fifteen years ago can be accomplished in weeks; a research procedure that was extremely costly then is extremely cost effective today. The respondent task is far easier and the computational procedures more effective. (S. 31)

Trotz des Erfolgs, den ACA zweifelsohne verbuchen kann, bleiben einige Forschungsfragen und Probleme offen:

- ❑ Die Vorgehensweise bei der Bildung des expliziten Modells kann keinesfalls als optimal eingestuft werden. Es wurden mehrere gute Verfahren der expliziten Modellbildung vorgeschlagen (vgl. Abschnitt 3.7.1). Es wäre zu überprüfen, ob diese auch rechnergestützt einsetzbar sind und ähnlich valide Ergebnisse erbringen.
- ❑ Trotz des Vorliegens einiger Untersuchungen zur optimalen Anzahl der Attribute je Beurteilungsobjekt bei der dekompositorischen Nutzenmessung (Profilstärke) und der damit verbundenen Frage nach der optimalen Anzahl der Beurteilungsobjekte (Anzahl der Paarvergleiche), kann noch keine abschließende Antwort auf diese Fragen gegeben werden. Erste Vorschläge empfehlen die Verwendung von Zweier-, höchstens aber Dreierprofilen bei der Hälfte der maximal möglichen Paarvergleiche (Gleichung 3.18). Weitere Untersuchungen zu diesem Fragenkomplex sind notwendig.
- ❑ Da das Verfahren aus mehreren Methoden (explizite Modellbildung und dekompositorische Modellbildung) besteht, stellt sich die Frage, wie Ergebnisse dieser Untersuchungsteile optimal kombiniert werden können. ACA 4.0 hat ein sehr eigenwilliges Verfahren hierzu implementiert. Dieses bedarf der eingehenden Prüfung. Auch alternative Vorschläge zur Lösung dieses Problems und deren (vergleichende) Prüfung sind wünschenswert.
- ❑ ACA erlaubt nur die Verwendung von additiven Teilnutzenwertmodellen. Eine Erweiterung der Präferenzmodelle auf lineare Vektormodelle und Idealpunktmodelle würde bei manchen Untersuchungen eine Einsparung von zu schätzenden Parametern erlauben. Bei den Integrationsmodellen

würde die Möglichkeit der Modellierung von ausgesuchten Interaktionen die Flexibilität des Systems erhöhen.

- ACA ist sehr sparsam, was die Überprüfungsmöglichkeiten der Angemessenheit der PSM angeht. Die Entwicklung und/ oder Implementation von Modellfit-Statistiken wäre sicherlich ein Gewinn für das Verfahren. Eine solche Maßzahl könnte z.B. auf der Diskrepanz zwischen empirischen Paarvergleichsantworten und denen, die aufgrund der letzten Modellschätzung zu erwarteten wäre, beruhen.

3.8 Parameterschätzverfahren

In diesem Abschnitt möchte ich tabellarisch die wichtigsten Parameterschätzverfahren aufführen. Das Parameterschätzverfahren von ACA 4.0 soll ausführlicher dargestellt werden, weil es bei unserer empirischen Untersuchung zum Einsatz kommt.

3.8.1 Überblick über die Verfahren

Das passende Parameterschätzverfahren hängt in seiner Logik von der Art der Antwortvariablen ab. Aus diesem Grund unterscheidet man zwischen der nominalskalierten Antwortvariable (z.B. Kategorienzuordnung), der ordinalskalierten Antwortvariable (z.B. Rangreihung), der intervallskalierten Antwortvariable (z.B. Rating) und der Wahlantwort (z.B. einfacher Paarvergleich), die als Ergebnis eines probabilistischen Wahlprozesses aufgefaßt wird.

Teilweise werden die Schätzprozeduren auch dann eingesetzt, wenn sie eigentlich unpassende Annahmen machen. Dies ist zum Beispiel der Fall, wenn OLS zur Parameterschätzung eingesetzt wird, obwohl die Antwortvariable eine Rangreihe ist. Diese Vorgehensweise wird z.T. aufgrund der einfacheren Verfügbarkeit von bestimmten Schätzalgorithmen (z.B. OLS) gewählt und meist mit der Robustheit des Verfahrens begründet, die z.B. in Monte-Carlo-Simulationen gefunden wurde. Eine Reihe von Simulationsstudien (z.B. Carmone, Green & Jain, 1978; Cattin & Bliemel, 1978; Green, 1975; Wittink & Cattin, 1981) und empirischen Untersuchungen (z.B. Jain, Acito, Malhotra & Mahajan, 1979) kommen zu dem Schluß, daß die geschätzten Parameter und insbesondere deren Validität weitgehend vom eingesetzten Parameterschätzverfahren unabhängig sind. So berichtet z.B. Green und Srinivasan (1978) von einer Simulationsstudie (Cattin & Wittink, 1976), in der die Ergebnisse zwischen OLS und MONANOVA hinsichtlich ihrer kriterienorientierten Validität als kaum unterscheidbar bezeichnet werden.

Tabelle 3.7 zählt die wichtigsten Parameterschätzverfahren auf, ordnet sie dem Skalenniveau der abhängigen Variablen zu und verweist auf Quellen, in denen weitere Informationen zum Verfahren zu finden sind.

Tabelle 3.8 veranschaulicht die Verwendungshäufigkeit der wichtigsten Parameterschätzmethoden (Cattin & Wittink, 1982; Wittink & Cattin, 1989; Schubert, 1991; Wittink et al., 1994). Daraus kann entnommen werden, daß den Prozeduren OLS und MONANOVA die größte Bedeutung zukommt, wobei OLS eine zunehmende Tendenz zeigt.

Tabelle 3.7: Die wichtigsten Parameterschätzverfahren.

Antwortvariable	Verfahren	Bemerkungen
Nominalskala	CCM	<i>Categorical Conjoint Measurement</i> (Carroll, 1969; 1973)
Ordinalskala	MONANOVA	<i>Monotonic Analysis of Variance</i> Kruskal, 1965; Kruskal & Carmone, 1969)
	LINMAP	<i>Linear Programming Techniques for Multi-dimensional Analysis of Preference</i> (Srinivasan & Shocker, 1973a, 1973b, 1981)
	PREFMAP	<i>Preference Mapping</i> (Carroll, 1972, 1973)
Intervallskala	JOHNSON	Parameterschätzung von ordinalen Urteilen in Trade-off Matrizen (Johnson, 1973, 1975)
	OLS	<i>Ordinary Least Square</i> (z.B. Judd & McClelland, 1989)
	ANOVA	<i>Analysis of Variance</i> (z.B. Green & Wind, 1973)
Wahlantworten	MSAE	<i>Minimizing Sum of Absolute Errors</i>
	LOGIT	(McFadden, 1974, 1976)

Tabelle 3.8: Relative Häufigkeit der Verwendung verschiedener Parameterschätzprozeduren bei der CA in kommerziellen Studien [K] und Studien zu Forschungszwecken [F]. Alle Angaben in %.

Studie	Cattin & Wittink (1982)	Wittink & Cattin (1989)	Schubert (1991)	Wittink et al. (1994)
Zeitraum	1971 - 1980	1981 - 1985	1972 - 1989	1986 - 1991
Zweck der Studien	hauptsächlich K	hauptsächlich K	K & F	hauptsächlich K
MONANOVA	24	11	53	15
LINMAP	0	6	8	7
OLS	16	54	23	59
LOGIT	10	11	1	7
andere	51	18	15	12

3.8.1 Das Parameterschätzverfahren bei ACA

In Phase I+II schätzt ACA initiale Werte der TNW durch explizite Modellbildung. Das Prinzip dieser Schätzung wurde in Abschnitt 3.7.2.4 dargestellt. Den Apn werden in Phase III Paare zum abgestuften Vergleich dargeboten (adaptives Paarvergleichsdifferenzendesign, Abschnitt 3.5.3.3). Die Apn geben ihre Antwort auf einer (in der Regel) neunstufigen Skala ab. 5 bedeutet dabei Präferenzindifferenz zwischen den dargebotenen Paaren, eine Zahl zwischen 1 und 4 (6 und 9) eine entsprechend ausgeprägte Präferenz für das auf der linken (rechten) Bildschirmseite dargebotene Paar (Abbildung 3.4). Diese Antworten werden intern auf den Wertebereich $\{-4, \dots, +4\}$ transformiert. Die Null entspricht damit einer Indifferenzantwort.

In Phase III schätzt ACA nach jeder Paarvergleichsantwort die Parameter durch eine OLS-Regression. Die Zielfunktion der Regression ist dabei:

$$\boxed{} \mathbf{b}_{uh} \approx \boxed{} \quad (3.27)$$

mit

- I** Identitätsmatrix. Die Spalten der Matrix repräsentieren die Attributstufen aller Attribute.
- X_h** Designmatrix der Ap h. Jede Zeile der Matrix steht für einen Paarvergleich. Den Attributstufen, die links (rechts) auf dem Bildschirm dargeboten wurden, wird eine -1 (+1) zugewiesen. Die Matrix **X_h** hat so viele Spalten wie Attributstufen und so viele Zeilen wie Paarvergleiche.
- b_{ph}** Vektor der expliziten TNW für Ap h. Sie wurden in Phase I+II bestimmt.
- y_h** Vektor der Antworten der Ap h bei den Paarvergleichen.
- b_{uh}** Vektor der zu schätzenden TNW für die Ap h.

Die TNW **b_{uh}** werden durch die Kleinst-Quadrate Minimierung geschätzt:

$$\mathbf{b}_{uh} = (\mathbf{I} + \mathbf{X}_h' \mathbf{X}_h)^{-1} (\mathbf{b}_{ph} + \mathbf{X}_h' \mathbf{y}_h) \quad (3.28)$$

Damit wäre das Problem der Parameterschätzung prinzipiell gelöst. Da das adaptive Design von ACA von der jeweils aktuellen Schätzung der TNW abhängt, muß eine Regressionsschätzung nach jedem Paarvergleich erfolgen.

Die jeweilige Neuschätzung der Parameter nach dem Prinzip der Gleichungen (3.27) und (3.28) ist auf einem Personalcomputer aufgrund der notwendigen Matrixinversion recht zeitintensiv, weil bei jeder Schätzung eine Matrixinversion durchgeführt werden muß. Aus diesem Grunde wird ein Schätz-'updating' durchgeführt, welches ohne die Berechnung der Inversen auskommt und die oben genannte Lösung approximiert. Das Verfahren ist in Sawtooth Software (1994) dargestellt.

4 Weiterverarbeitung der Präferenzstrukturmodelle

Eine CA kann nach dem in Abschnitt 3 bzw. Anhang A skizzierten Ablaufschema durchgeführt werden. Nach der Durchführung liegt als Ergebnis für jede Ap ein individuelles PSM mit den geschätzten Parametern vor. Damit ist der Kern der CA beendet. Die individuellen PSM können interpretiert und in Abhängigkeit der Ziele einer CA-Untersuchung umgesetzt werden (z.B. Unterstützung von Managemententscheidungen).

Tabelle 4.1: Überblick über die Weiterverarbeitung von Präferenzstrukturmodellen auf aggregierter Ebene.

Problembereich	Vorgehensweise/ Verfahren
Normierung	
Niveaueausgleich	niedrigster TNW auf Null setzen Attributzentrierung
Variationsausgleich	Attributwichtigkeitsnormierung Summennormierung Standardabweichungsnormierung
Aggregation auf der Ebene der Nutzenfunktion	
	Aggregation durch Mittelung der PSM-Parameter
Segmentierung	
a-priori-Segmentierung	a-priori-Segmentierung
Nutzensegmentierung	Segmentierung auf Antwortvektorbasis Segmentierung auf Präferenzstrukturmodellbasis Segmentierung nach der Modellanpassung
Wahlanteilsimulation	
	Maximum-utility-Modell BTL-Modell Logit-Modell Adjusted ACA-Logit-Modell Logistisches Modell

In aller Regel möchte man jedoch nicht nur Aussagen über Nutzenstrukturen auf individueller Ebene machen. Das Hauptziel einer CA-Untersuchung besteht meist darin, generalisierte Aussagen über eine Stichprobe (bzw. Population), oder über homogene Personengruppen in der Stichprobe (bzw. Population) abzuleiten. Um dies leisten zu können, müssen Daten aggregiert bzw. segmentiert werden. Diesen Schritt bezeichnen wir als Weiterverarbeitung der individuellen PSM. Abschnitt 4 widmet sich dieser Weiterverarbeitung.

Green (1984) unterscheidet die Aggregation auf der Ebene der Nutzenfunktionen und auf der Ebene der Wahlanteile. Die Aggregation auf der Nutzenfunktionsebene wird in Abschnitt 4.2, die Aggregation auf der Wahlanteilesebene durch Wahlsimulationen in Abschnitt 4.4 behandelt. Bevor Daten aggregiert werden oder in eine Simulation eingehen, müssen Überlegungen zur Normierung der PSM und zur Segmentierung der Apn-Stichprobe angestellt werden. Diese werden in Abschnitt 4.1 bzw. 4.3 zusammenfassend dargestellt. Um die Darstellung einfach halten zu können, beschränken wir uns in Abschnitt 4 auf das additive Teilnutzenwertmodell (Gleichung 3.10). Tabelle 4.1 skizziert den Aufbau von Abschnitt 4.

4.1 Normierung

Die Annahme, daß die TNW eines PSM intervallskaliert sind, und damit der Nullpunkt beliebig wählbar ist, bedingt die Notwendigkeit, daß die individuellen PSM vor einer Weiterverarbeitung vergleichbar gemacht werden. Diesen Vorgang nennt man *Normierung*. Man kann die Normierung von intervallskalierten Teilnutzenwertmodellen gedanklich in zwei Schritte unterteilen.

Der erste Schritt dient dem Ausgleich von *Niveauunterschieden* zwischen den individuellen PSM. Hierzu wird in aller Regel der niedrigste TNW eines jeden Attributs auf Null gesetzt (Backhaus et al. 1994; Sawtooth Software, 1994) oder die Teilnutzenwerte innerhalb der Attribute zentriert. In beiden Fällen addiert man zu den TNW des individuellen PSM innerhalb eines jeden Attributs einen Summanden. Welches der Verfahren verwendet wird, ist von untergeordneter Bedeutung, da der Nullpunkt bei intervallskalierten Daten ohnehin willkürlich ist.

Der zweite Schritt der Normierung dient dem *Ausgleich von Variationsunterschieden* zwischen den individuellen PSM. Eine ganze Reihe von Trans-

formationsvorschriften wurde vorgeschlagen, um die Variation der TNW zwischen den Apn vergleichbar zu machen. Eine dieser Vorschriften macht sich den Gedanken der relativen Attributwichtigkeiten (vgl. Abschnitt 2.3.2) zu eigen. Die TNW eines jeden PSM werden mit einem individuellen Faktor so multipliziert, daß die Summe der relativen Attributwichtigkeiten (höchster minus niedrigster TNW innerhalb eines Attributs) für jedes PSM einen festgelegten Wert ergeben (z.B. Summe der Attributwichtigkeiten = 1; SPSS-Prozedur CONJOINT, Backhaus et al., 1994; SPSS, 1994; Summe der Attributwichtigkeiten = $100 * \text{Anzahl der Attribute}$; ACA 4.0, Sawtooth Software, 1994). Diese Art der Normierung nennen wir *Attributwichtigkeitsnormierung*.

Eine andere Form der Normierung kommt ebenfalls bei ACA 4.0 zum Einsatz. Hier werden nach dem Ausgleich der Niveauunterschiede die TNW mit einem individuellen Faktor so multipliziert, daß sich die TNW auf einen festgelegten Wert summieren (*Summennormierung*; z.B. Summe der TNW = $100 * \text{Anzahl der Attribute}$; ACA 4.0, Sawtooth Software, 1994).

Ein drittes Verfahren, welches die Vergleichbarkeit verschiedener PSM herstellen soll, bezeichnen wir als *Standardabweichungsnormierung*. Bei diesem Verfahren werden nach der Niveauekorrektur die Standardabweichungen der TNW innerhalb eines PSM gleichgesetzt. Hierzu werden die TNW ebenfalls mit einem PSM-spezifischen Faktor multipliziert (z.B. Standardabweichung der TNW = $10 * \text{Anzahl der Attributstufen}$).

4.2. Aggregation

Sind die PSM der Einzelindividuen normiert, können sie über mehrere Individuen aggregiert werden. Dies erfolgt meist durch Mittelwertbildung. Da die Parameter der PSM intervallskaliert sind, ist ihr Mittelwert eine sinnvolle Maßzahl.

Bei der Mittelung der TNW stellt sich die Frage, über welche Individuen die Aggregation vorgenommen werden soll. Da bei jeder Datenaggregation Information verloren geht, muß sehr genau überlegt werden, wie man vorgeht. Wird über die gesamte Stichprobe aggregiert, dann besteht die Gefahr, einen sog. 'Mehrheitstrugschluß' ('majority fallacy'; vgl. Kuehn & Day, 1962) zu begehen. Dieses Phänomen kann bei jeder Mittelwertbildung auftreten und soll anhand eines CA-spezifischen Beispiels verdeutlicht werden. Im Rahmen einer Unter-

suchung der Präferenzstrukturen von Studierenden der Psychologie gegenüber hypothetischen Prüfungsordnungen wurde festgestellt, daß ein Teil der Befragten eine klinisch-psychologische Schwerpunktausbildung gegenüber einer organisationspsychologischen stark bevorzugte. Ein anderer Teil dagegen zeigte die gegenteiligen Präferenzen. Technisch ausgedrückt heißt dies, daß bei Attribut 'Schwerpunktausbildung' Segment A einen hohen Teilnutzen durch die Attributstufe 'Schwerpunkt Klinische Psychologie' und einen niedrigen Teilnutzen durch die Attributstufe 'Schwerpunkt Organisationspsychologie' erfährt. Für Segment B gilt diese Aussage umgekehrt. Mittelt man die Teilnutzenwerte über alle Individuen der Auskunftsstichprobe, stellt man fest, daß die Studierenden im Durchschnitt aus beiden Schwerpunktausbildungsmöglichkeiten einen ähnlichen, aber nur mittleren Teilnutzen erfahren. Vergleicht man die relative Wichtigkeit dieses Attributs mit einem anderen, dann kommt man zu dem Schluß, daß die relative Wichtigkeit des Attributs 'Schwerpunktausbildung' eine untergeordnete Bedeutung einnimmt. Beide Schlüsse sind falsch. Es handelt sich um einen Artefakt, der bei der Aggregation durch Mittelwertbildung entsteht.

Um bei der Datenaggregation möglichst wenig Information zu verlieren und einen Mehrheitstrugschluß zu vermeiden, sollte darauf geachtet werden, daß nur möglichst homogene PSM zusammengefaßt werden. Sind die Präferenzstrukturen einer Teilmenge von Apn im Vergleich zur Gesamtstichprobe homogen, spricht man auch von einem Segment. Vor der Aggregation von PSM sollte zunächst überprüft werden, ob sich die Stichprobe in homogene Segmente einteilen läßt. Im nächsten Abschnitt werden verschiedene Verfahren der Segmentbildung kurz vorgestellt.

4.3 Segmentierung

Den Grundgedanken der Segmentierung soll am Beispiel der Marktsegmentierung aufgezeigt werden. Marktsegmentierungsansätze nehmen an, daß Käuferpräferenzen gegenüber Produkten/ Dienstleistungen heterogen sind. Das Ziel der Segmentierung besteht darin, homogene Untergruppen von Käufern zu finden und diese anhand beobachtbarer Merkmale zu beschreiben. Marketingaktivitäten können dann differenziert für die identifizierbaren Segmente statt-

finden. Die CA als individuelles Verfahren der Modellierung von Präferenzen ist eine gute Ausgangsbasis für Segmentierungsbemühungen.

Wind (1978) nennt zwei grundlegende Ansätze der Segmentierung. Sie unterscheiden sich nach der Art, wie die Stichprobe in homogene Untermengen unterteilt wird. Eine Möglichkeit ist die *a-priori-Segmentierung*. Bei dieser Vorgehensweise werden neben den Präferenzdaten sogenannte Hintergrundvariablen der Apn erhoben. Dabei kann es sich um personenbezogene und situative Variablen handeln (z.B. Currim, 1981; Green & Krieger, 1989; Green & Krieger, 1991). Personenvariablen sind z.B. demographische Variablen (Alter, Geschlecht, ...), sozioökonomische Variablen (Familienstand, Einkommen, ...), psychologische Variablen (Persönlichkeitseigenschaften, ...) oder Variablen des Produktverwendungsverhaltens (Produktbekanntheit, Markenbenutzung, Nutzungshäufigkeit, ...). Situationsvariablen unterscheiden bestimmte Situationen, in denen Verhalten stattfindet (z.B. Zweck des Verhaltens: Etwas für sich selbst oder als Geschenk kaufen; Begleiterscheinungen des Verhaltens: Getränkekonsum zu bestimmten Anlässen). Hinter dem Gedanken der a-priori-Segmentierung steht die Annahme, daß die Hintergrundvariablen einen Zusammenhang mit den individuellen Präferenzstrukturen aufweisen. Sie verlangt theoretische Vorstellungen darüber, welche Segmente zu erwarten sind und wie sich diese unterscheiden lassen.

In vielen Fällen liegen die differenzierten Vorstellungen, die eine a-priori-Segmentierung ermöglichen, nicht vor. In diesem Fall kann man sich verschiedener *Nutzensegmentierungsverfahren (a-posteriori oder post hoc Segmentierungsverfahren)* bedienen. Dabei werden die Parameter der PSM meist einem clusteranalytischen Verfahren (z.B. Backhaus et al. 1994) unterzogen. Als Inputdaten für die Clusteranalyse sind z.B. die Rohdaten des Antwortvektors, die TNW des PSM oder Maßzahlen der Modellanpassung (vgl. Abschnitt 5.3) denkbar. Welche der Daten verwendet werden, hängt von theoretischen Überlegungen und dem Zweck der Segmentierung ab. Das am häufigsten verwendete Verfahren der Segmentierung dürfte auf den PSM basieren. Nach einer geeigneten Normierung der Modellparameter (bei ACA 4.0 wird z.B. die Summennormierung der TNW empfohlen; Sawtooth Software, 1994) wird eine Clusteranalyse (z.B. Ward-Verfahren auf der Basis der euklidischen Distanz der normierten TNW) gerechnet. Für welche Cluster-Lösung man sich entscheidet,

hängt dabei sicherlich nicht unerheblich von der theoretischen Interpretierbarkeit der Segmente und deren Beschreibbarkeit durch Hintergrundvariablen ab.

Hagerty (1985) bezweifelt grundsätzlich die Angemessenheit einer PSM-Segmentierung durch Methoden, die eine Stichprobe in nichtüberlappende Cluster aufteilt. Er begründet seine Haltung damit, daß die Annahme nichtüberlappender Cluster im Marketing nur selten haltbar ist. Werden die Apn in einem multidimensionalen Raum abgebildet, zeigen sich meist keine dichten Gruppierungen von Apn. Vielmehr sind die Apn ziemlich gleichmäßig über den Raum verteilt. Er schlägt für diesen Fall eine faktorenanalytische Methode ('optimal respondent weighting') zur Segmentierung vor.

Hagertys Haltung sollte nach unserem Dafürhalten kritisch gesehen werden. Hierfür möchten wir zwei Gründe anführen. Zum einen beziehen sich seine empirischen Argumente gegen die Clusteranalyse und für die Faktorenanalyse im wesentlichen auf die Steigerung der kriterienorientierten Validität aggregierter PSM durch die Segmentierung. Segmentierung kann jedoch auch andere Ziele verfolgen (z.B. Bildung eines einfachen Ordnungsschemas von Abnehmertypen). Zum anderen kann es keineswegs als geklärt angesehen werden, ob die aggregierten PSM nach der Hagerty-Segmentierungsmethode höher valide sind als nach einer clusteranalytischen Segmentierung (zu dieser Problematik siehe z.B. Green, Krieger & Schaffer, 1993; Hagerty, 1993; Wedel & Kistemaker, 1989; Wedel & Steenkamp, 1989, 1991).

4.4 Wahlanteilsimulation

Wahlanteilsimulationen stellen, neben den beschriebenen Verfahren der Mittelwertbildung, eine wertvolle Ergänzung zu den Aggregationsmethoden von CA-Ergebnissen dar. Das Grundprinzip der Wahlanteilsimulation ist denkbar einfach: Ist das PSM einer Ap bekannt, kann für jedes beliebige Objekt der Objektklasse der Gesamtnutzen berechnet werden. Besitzt bei einer Ap h das Objekt X_1 einen höheren Nutzen als X_2 , dann bedeutet dies, daß sie X_1 gegenüber X_2 präferieren sollte. Handelt es sich bei der Objektklasse z.B. um Konsumgüter und sind beide Produkte gleichermaßen verfügbar, dann sollte sich die Ap in einer Kaufentscheidungssituation entsprechend dem Nutzenmaximierungsprinzip für X_1 entscheiden.

Besteht jedoch bei der Wahlwahrscheinlichkeit ein Unterschied, ob die Nutzendifferenz der Objekte hoch oder gering ist? Welche Wahlanteile können verschiedene Objekte bei einer Stichprobe von Personen mit jeweils individuell verschiedenen Präferenzstrukturen auf sich vereinen? Diese und ähnliche Fragen zu beantworten, ist das Ziel von Wahlanteilsimulationen. Sie werden mit Hilfe sogenannter 'Wahlanteilsimulatoren' durchgeführt.

Ein *Wahlanteilsimulator* modelliert den Zusammenhang zwischen den Präferenzstrukturen einer Stichprobe von A_{pn} und deren Wahlentscheidungen für Objekte einer definierten Objektmenge. Bei der Stichprobe kann es sich, je nach Fragestellung, um die (repräsentative) Gesamtstichprobe oder auch um die A_{pn} eines Beurteilersegments handeln. Die Menge der Objekte, die in eine Simulation eingehen, ist eine Teilmenge aus der Objektklasse. Diese Objektmenge wird auch als 'Szenario' bezeichnet. Es besteht typischerweise aus 3-8 Objekten.

Im folgenden möchten wir gängige Wahlanteilsimulatormodelle vorstellen und diskutieren (z.B. Sawtooth Software, 1994; Finkbeiner, 1988; Wiley & Low, 1983). Alle Modelle nehmen an, daß Wahlentscheidungen nach dem Nutzenmaximierungsprinzip erfolgen (z.B. Thurstone, 1945; Huber & Moore, 1979). Dies bedeutet, daß bei Wahlentscheidungen Objekte mit höherem Nutzen solchen prinzipiell vorgezogen werden, die einen niedrigeren Nutzen besitzen. Wie in Tabelle 4.1 skizziert, möchten wir das Maximum-utility-Modell, das BTL-Modell, das Logit-Modell, das adjusted ACA-Logit-Modell und das logistische Modell vorstellen. Weitere Modelle werden z.B. bei Currim (1982) besprochen.

4.4.1 Maximum-utility-Modell

Das Maximum-utility-Modell nimmt an, daß jedes Individuum aus mehreren Alternativen das Objekt auswählt, welches den höchsten Nutzen besitzt. Die Wahlanteilschätzung eines Objekts ist der proportionale Anteil der Stichprobe, bei der das Objekt den höchsten Gesamtnutzen in der Objektmenge innehat. Haben mehrere Objekte bei einer A_p den gleichen Nutzen, wird angenommen, daß sie zufällig zwischen ihnen wählt. Das Modell kann folgendermaßen spezifiziert werden:

$$s(X_i) = \frac{p_{ni}}{H} \quad (4.1)$$

mit

$s(X_i)$ Wahlanteil ('share') für Objekt X_i .

$$p_{hi} = \begin{cases} 1/n, & \text{wenn } u_h(X_i) = \max [u_h(X_1), u_h(X_2), \dots, u_h(X_i), \dots, u_h(X_l)] \\ 0 & \text{in anderen Fällen.} \end{cases}$$

l Anzahl der Objekte im Simulationsszenario.

n Anzahl der Objekte, die den gleichen Gesamtnutzen haben.

$u_h(X_i)$ Gesamtnutzen des Objekts X_i bei Ap h .

H Stichprobengröße.

Das Maximum-utility-Modell macht außer der Nutzenmaximierung keine weiteren Annahmen. Dies ist insofern eine Besonderheit, als daß alle weiteren Modelle nicht ohne Verteilungsannahmen der Gesamtnutzenwerte der konkurrierenden Objekte im Szenario auskommen.

Neben der Annahmensparsamkeit hat das Maximum-utility-Modell zwei wesentliche Vorteile gegenüber den anderen Modellen. Erstens sind die geschätzten Wahlanteile invariant gegenüber monotonen und linearen Transformationen der Nutzenwerte. Diese Eigenschaft ist konsistent mit der Annahme, daß die Gesamtnutzenwerte bestenfalls intervallskaliert sind. Zweitens ist die Logik des Maximum-utility-Modells einfach, was die Kommunizierbarkeit von Wahlsimulationsergebnissen gewährleistet.

Diesen Vorteilen stehen auch Nachteile gegenüber. Der deterministische Charakter des Modells berücksichtigt nicht, wie groß der 'Vorsprung' des Objekts mit dem höchsten Nutzen vor seinen Konkurrenten ist. Ihm wird immer die Wahlwahrscheinlichkeit 1 zugeordnet, während alle übrigen Objekte die Wahlwahrscheinlichkeit 0 erhalten. Dies ist auch der Fall, wenn ein Objekt des Szenarios einen nahezu gleich hohen Nutzen wie das nutzenhöchste Objekt besitzt. Intuitiv scheint dies jedoch unangemessen, da beide Objekte eine ähnliche Wahlwahrscheinlichkeit haben sollten.

Das Maximum-utility-Modell dürfte neben dem BTL-Modell und dem Logit-Modell das meistgenutzte Wahlsimulationsmodell bei kommerziellen CA-Untersuchungen sein (vgl. auch Green & Krieger, 1988).

4.4.2 BTL-Modell

Beim BTL-Modell wird zur Schätzung des Wahlanteils der Gesamtnutzenwert eines Objekts zur Summe der Nutzenwerte aller Objekte im Szenario in Bezie-

hung gesetzt und über die Apn gemittelt (Bradley & Terry, 1952; Luce, 1959; Luce, Bush & Galanter, 1965). Das Modell hat folgende Form:

$$s(X_i) = \frac{1}{H} \left[\frac{u_h(X_i)}{\sum_{j=1}^I u_h(X_j)} \right] \quad (4.2)$$

mit

$s(X_i)$ Wahlanteil für Objekt X_i .

I Anzahl der Objekte im Szenario.

$u_h(X_i)$ Gesamtnutzen des Objekts X_i bei Ap h.

H Stichprobengröße.

Der grundlegende Unterschied dieses und der folgenden Modelle gegenüber dem Maximum-utility-Modell besteht darin, daß die Beeinflußbarkeit von Wahlentscheidungen durch die Nutzenähnlichkeit der Objekte im Szenario berücksichtigt wird.

Finkbeiner (1988) weist darauf hin, daß das BTL-Modell für die Wahlanteilsimulation unangemessen ist, weil die Nutzenwerte bestenfalls intervallskaliert sind, d.h. daß sie einer willkürlichen linearen Transformation unterzogen werden können, ohne daß dadurch ihre Repräsentation der ursprünglichen Antworten der Apn beeinflusst wird. Die Simulationsergebnisse des BTL-Modells sind dagegen nicht invariant gegenüber dieser Transformation.

Ein Beispiel soll dies verdeutlichen. Nehmen wir an, ein Szenario enthält drei Objekte mit den Nutzen $u_h(X_1)=2$, $u_h(X_2)=3$ und $u_h(X_3)=5$ für die Ap h. Die Wahlwahrscheinlichkeit der Objekte für Ap h nach dem BTL-Modell wäre

$$(4.2) \quad s(X_1) = 2 / 10 = 0,2$$

$$(4.2) \quad s(X_2) = 3 / 10 = 0,3$$

$$(4.2) \quad s(X_3) = 5 / 10 = 0,5$$

Da die Nutzenwerte intervallskaliert sind, kann man ihnen jeweils die Zahl 5 zuaddieren, ohne ihren Informationsgehalt zu ändern. Die Wahlwahrscheinlichkeit der Objekte für Ap h sind dann

$$(4.2) \quad s(X_1) = 7 / 25 = 0,28$$

$$(4.2) \quad s(X_2) = 8 / 25 = 0,32$$

$$(4.2) \quad s(X_3) = 10 / 25 = 0,4.$$

Obwohl die Nutzenwerte einer zulässigen Transformation unterzogen wurden, ändert sich das Ergebnis der Wahlanteilsimulation. Trotz dieser berechtigten

Kritik kommt das Modell häufig zum Einsatz und ist in gängigen Softwaresystemen implementiert (z.B. in der SPSS-Prozedur CONJOINT).

Ein weiterer Kritikpunkt am BTL-Modell hängt mit der IIA-Annahme (*Independence of Irrelevant Alternatives*) zusammen. Wir möchten diesen am Beispiel des Logit-Modells im nächsten Abschnitt näher beschreiben.

4.4.3 Logit-Modell und ACA-Logit-Modell

Das Logit-Modell ist im Aufbau dem BTL-Modell sehr ähnlich. Es modelliert jedoch explizit eine Fehlerkomponente bei der Wahlentscheidung. Unter Annahme von identischen und unabhängigen Verteilungen der Fehler von Wahlentscheidungen (McFadden, 1974, 1976) gilt:

$$s(X_i) = \frac{e^{U_h(X_i)}}{\sum_{h=1}^H e^{U_h(X_i)}} \quad (4.3)$$

mit

$s(X_i)$	Wahlanteil für Objekt X_i .
I	Anzahl der Objekte im Simulationsszenario.
$u_h(X_i)$	Gesamtnutzen des Objekts X_i bei $A_p h$.
H	Stichprobengröße.

Wie für das BTL-Modell gilt auch für das Logit-Modell die Kritik von Finkbeiner (1988), sofern die Nutzenwerte intervallskaliert sind.

Beim Softwaresystem ACA 4.0 werden die Nutzenwerte auf der Basis der Kaufwahrscheinlichkeiten in Phase IV reskaliert (vgl. Abschnitt 3.7.4.2). Die Transformationsvorschrift ist dabei nicht willkürlich und es wird angenommen, daß die reskalierten Gesamtnutzenwerte ('logits') eines Objekts $u_h(X_i)$ eine Abbildung auf einer Verhältnisskala darstellen. Die Kritik von Finkbeiner (1988) am BTL-Modell und Logit-Modell trifft damit nicht auf das ACA-Logit-Modell zu.

Das Logit-Modell macht jedoch eine andere problematische Annahme. Es unterstellt, daß das Verhältnis der Wahlwahrscheinlichkeiten zweier beliebiger Objekte im Szenario vom Vorhandensein weiterer Objektalternativen unabhängig ist (*Independence of Irrelevant Alternatives*, IIA-Annahme). IIA impliziert damit, daß die Wahlwahrscheinlichkeit für ein Objekt ausschließlich durch seinen relativen Nutzen determiniert ist. Es wird damit angenommen, daß Wahlentscheidungen unabhängig von den Ähnlichkeiten zwischen den Objekten getroffen werden. Dies steht im Widerspruch zu vielen Gegenbeispielen und empirischen Befunden (vgl. Currim, 1982). Im Marketing ist z.B. der Effekt der Substituierbarkeit von Gütern bekannt. Wird ein neues Produkt auf den Markt gebracht, so vereinnahmt dieses die Marktanteile alternativer Produkte nicht gleichmäßig. Vielmehr entfallen auf das neue Produkt überproportional viele Kaufentscheidungen, die sonst für sehr ähnliche Produkte gefällt würden.

Daß die IIA-Annahme unter bestimmten Bedingungen unangemessen sein kann, verdeutlicht ein einfaches Beispiel. Besitzt in einem Szenario Objekt X_1 den Nutzen $u_h(X_1)=0,8$ und Objekt X_2 den Nutzen $u_h(X_2)=0,4$, dann berechnen sich die Wahlwahrscheinlichkeiten für $A_p h$ mit

$$(4.3) \quad s(X_1) = e^{0,8} / (e^{0,8} + e^{0,4}) = 0,60$$

$$(4.3) \quad s(X_2) = e^{0,4} / (e^{0,8} + e^{0,4}) = 0,40.$$

Wird dem Szenario ein Objekt X_3 hinzugefügt, das mit Objekt X_1 identisch ist, dann sollte man intuitiv erwarten, daß sich die beiden Objekte die ursprüngliche

Wahlwahrscheinlichkeit von Objekt X_1 teilen, also auf beide Objekte 30% der Wahlentscheidungen entfallen. Das Logit-Modell ergibt aber folgendes Ergebnis:

$$(4.3) \quad s(X_1) = e^{0,8} / (e^{0,8} + e^{0,8} + e^{0,4}) = 0,37$$

$$(4.3) \quad s(X_2) = e^{0,4} / (e^{0,8} + e^{0,8} + e^{0,4}) = 0,25$$

$$(4.3) \quad s(X_3) = e^{0,8} / (e^{0,8} + e^{0,8} + e^{0,4}) = 0,37.$$

Auf Objekt X_1 und X_3 entfallen zusammen fast 75% der Wahlentscheidungen, d.h. das neu hinzugefügte Objekt X_3 raubt nicht nur seinem identischen Objekt X_1 Wahlanteile, sondern auch von Objekt X_2 . Das Beispiel zeigt, daß das Logit-Modell insbesondere dann unangemessen sein kann, wenn im Szenario sehr ähnliche Objekte vorhanden sind.

4.4.4 Adjusted ACA-Logit-Modell

Dieses Modell, das bei ACA 4.0 'share of preference model, with correction for similarity' genannt wird, ist eine Veränderung des Logit-Modells. Das Grundprinzip des Modells besteht darin, die Ergebnisse des Logit-Modells so zu korrigieren, daß auf identische Objekte jeweils die Hälfte des Wahlanteils von nur einem der Objekte entfallen würde. Hierzu wird zunächst eine $I \times I$ Objekt-ähnlichkeitsmatrix aus den I Objekten des Szenarios erstellt. Eine 1 in dieser Matrix bedeutet eine vollkommene Ähnlichkeit zwischen zwei Objekten, eine 0 deutet auf eine vollkommene Unähnlichkeit hin.

Um zu dieser Ähnlichkeitsmatrix zu gelangen, werden die Stufen eines jeden Attributs mit 10, 20, 30, ... codiert. Die Unähnlichkeit zwischen zwei Produkten auf einem Attribut ist die Differenz zwischen diesen Codierungen, wobei das Maximum der Differenz innerhalb eines Attributs auf 10 begrenzt. Die Programmautoren begründen dies schwer nachvollziehbar damit, daß auf diese Weise qualitative und quantitative Attribute gleich behandelt werden können. Die gesamte Unähnlichkeit zweier Objekte ist die Summe ihrer Unähnlichkeit über alle Attribute.

Danach werden die (Gesamt-)Unähnlichkeitswerte so normiert, daß der maximale Wert 3 (statt $10 \cdot \text{Attributanzahl}$) ist. Diese Unähnlichkeitswerte werden in Ähnlichkeitswerte umgewandelt, indem sie als negatives Argument in die Exponentialfunktion eingesetzt werden. Damit ist die minimale Ähnlichkeit $\exp(-3) \approx 0,05$. Dieser Wert wird erreicht, wenn zwei Szenarioobjekte auf jedem Attribut sehr verschieden sind. Der maximale Ähnlichkeitswert ist $\exp(0) = 1$. Er

wird erreicht, wenn zwei Produkte auf allen Attributen identisch sind. Eine weitere Reskalierung setzt das Minimum des Ähnlichkeitswertebereichs auf 0 und das Maximum auf 1.

Die Spaltensummen der Objektähnlichkeitsmatrix ('Gesamtähnlichkeiten') werden berechnet. Den kleinstmöglichen Wert 1 würde man erhalten, wenn ein Objekt gegenüber allen anderen maximal unähnlich ist (aber natürlich sich selbst gegenüber maximale Ähnlichkeit zeigt). Zwei identische Objekte, die gegenüber allen anderen Objekte maximal unähnlich sind, erreichen eine Spaltensumme von 2.

Für jede Apn werden die Wahlwahrscheinlichkeiten der Objekte, wie sie durch das ACA-Logit-Modell berechnet werden, durch die Gesamtähnlichkeiten dividiert und danach renormalisiert, so daß sie sich auf 1 summieren. Dies hat den Effekt, daß die Wahlanteile von Objekten reduziert werden, die eine relativ hohe Ähnlichkeit zu anderen Objekten aufweisen.

Finkbeiner (1988) kritisiert diese Vorgehensweise als theoretisch unangemessenen. Sie löst keinesfalls das IIA-Problem. Trotz dieser theoretischen Defizite wird das adjusted ACA-Logit-Modell in der Praxis oft eingesetzt.

4.4.5 Logistisches Modell

Das multivariate logistische Modell (Bock & Jones, 1968) nimmt an, daß die Gesamtnutzenwerte der Objekte im Szenario normal verteilt sind. Sei \mathbf{U} die Matrix der Gesamtnutzenwerte mit den Zellen u_{hi} . In den Spalten der Matrix sind die Nutzenwerte der Szenarioobjekte X_i und in den Zeilen die Ap h repräsentiert. u_i sind die Spaltensummen der Matrix. Die Differenz der Spaltensummen zweier Objekte ist $u_{ij^*} = u_i - u_{i^*}$. Dann berechnet sich die Wahlwahrscheinlichkeit für das Objekt X_{i^*} durch

$$s(X_{i^*}) = 1 / [1 + \sum_{j \neq i^*} e^{-c \mu_{ij^*} / \sigma_{ij^*}}] \quad (4.4)$$

mit

$s(X_{i^*})$ Wahlanteil für Objekt X_{i^*} .

I Anzahl der Objekte im Simulationsszenario.

c $\pi \sqrt{3}$.

- μ_{ij}^* Mittlere Differenz der Nutzenwerte über die Ap h.
 σ_{ij}^* Standardabweichung der Nutzendifferenzen über die Ap h.

Das multivariate logistische Modell macht nicht die IIA-Annahme. Wiley und Low (1983) fanden in ihrer Monte-Carlo-Simulation, daß das logistische Modell effizientere Schätzungen erlaubt als das Maximum-utility Modell. Allerdings unterschieden sich die Ergebnisse beider Modelle nur unwesentlich.

4.4.6 Wahlanteilsimulation - Marktanteilsimulation

Sind in einer CA-Untersuchung die PSM und der Wahlanteilsimulator valide, dann sollte das Wahlverhalten in Holdout-Objektsets richtig vorhergesagt werden können (siehe Abschnitt 5.4.2 'kriterienorientierte Validität auf aggregierter Ebene'). In einer marketingorientierten Studie interessieren jedoch letztendlich nicht diese Wahlanteile, sondern Marktanteile, die ein Objekt unter anderen Objekten am Markt erringen kann. Werden die Wahlanteilsimulationsergebnisse mit tatsächlichen Marktanteilen verglichen, wäre nur in vollkommenen Märkten eine Übereinstimmung zu erwarten. Reale Märkte sind jedoch in aller Regel unvollkommen. Das Entscheidungsverhalten wird von weiteren Faktoren als der Objektpräferenz beeinflusst, weil neben den Objektpräferenzen andere Präferenzen wirksam sind, die sich z.B. auf räumliche und zeitliche Gegebenheiten beziehen, weil Märkte nicht vollkommen transparent sind und weil Anpassungsprozesse auf Marktveränderungen nur langsam erfolgen. Neben den reinen Objektpräferenzen stellt sich z.B. die Frage, ob ein Objekt an einem bestimmten Ort und zu einem bestimmten Zeitpunkt verfügbar ist, ob beim Abnehmer das Bewußtsein darüber besteht, daß es verfügbar ist und über welche Eigenschaften es verfügt, wie hoch der Beschaffungsaufwand für das Objekt liegt und ob die momentanen Beschaffungsressourcen vorhanden sind.

Die Methode der CA modelliert den Zusammenhang zwischen wenigen Objektattributen und der Objektpräferenz. Simulationen, die auf CA-Ergebnissen beruhen, können in den meisten Fällen reales Marktverhalten nur sehr vage vorhersagen, da dieses von weiteren Faktoren als den Objektpräferenzen beeinflusst wird. Soll dies trotzdem versucht werden, bietet z.B. das Softwaresystem ACA 4.0 die Möglichkeit, sog. 'externe Faktoren' in eine Marktanteilsimulation aufzunehmen. Die Gesamtnutzenwerte der Szenarioobjekte werden

mit diesen Faktoren multipliziert. Die Faktoren müssen von Experten geschätzt werden. Ist z.B. ein Szenarioprodukt X_1 doppelt so häufig am Markt verfügbar als Produkt X_2 , sollte Produkt X_1 ein höherer externer Faktor als Produkt X_2 zugeordnet werden. Es braucht wohl kaum erwähnt zu werden, daß diese Schätzungen mit einer enormen Unsicherheit behaftet sind und bei dieser Vorgehensweise extreme Grenzbereiche der Vorhersage durch wissenschaftliche Methoden erreicht werden.

Zusammenfassend kann man festhalten, daß Marktanteile durch Untersuchungen nach dem Prinzip der CA und darauf beruhender Simulationen nicht in ihrer absoluten Höhe vorhergesagt werden können. Dies bedeutet aber nicht, daß Simulationen wertlos sind. So dürfte es doch möglich sein, relative Marktanteilsänderungen aus 'Was-wäre-wenn?-Simulationen' abzuleiten. Bei dieser Art der Simulation werden die Attributausprägungen der Objekte des Szenarios in aufeinanderfolgenden Simulationen systematisch verändert. So kann z.B. von Interesse sein, wie der Markt tendenziell reagiert, wenn Konkurrent A sein Produkt in einer bestimmten Weise verändert und Konkurrent B in einer bestimmten anderen Weise gegensteuert. Die Simulatoren lassen zu, solche Szenarien durchzuspielen.

5 Meßgütekriterien der Conjoint Analyse

In diesem Abschnitt wollen wir ausführlicher die Meßgütekriterien besprechen, die zur Beurteilung von CA-Ergebnissen verwendet werden. Die erste Kriterienklasse, die subjektiven und objektiven Reaktionen auf die Erhebungsprozedur (Abschnitt 5.1), enthält keine Meßgütekriterien im eigentlichen Sinne. Dennoch sind solche Reaktionen wichtige Beurteilungskriterien einer CA. Die meßtheoretischen Eigenschaften der CA können über Maße der Reliabilität, Modellanpassung, Validität und Strukturwiedererkennung von 'wahren PSM' eingeschätzt werden. In den Abschnitten 5.2 bis 5.5 wollen wir typische Vorgehensweisen und Maßzahlen vorstellen und jeweils beispielhaft Untersuchungen anführen, die diese Maße verwenden. Die Darstellung in Abschnitt 5 soll knapp gehalten werden. Aus diesem Grunde werden wir Aspekte der Meßgütekriterien, die in der Evaluationsforschung allgemeingültig sind, nicht weiter thematisieren und uns nur vertieft den Besonderheiten zuwenden, die im Zusammenhang mit der CA auftreten. Im allgemeinen wird die Vorgehensweise anhand eines additiven Teilnutzenwertmodells (Gleichung 3.10) besprochen.

5.1 Subjektive und objektive Reaktion

Durchläuft eine Ap das Interview, können objektive Reaktionen, wie die benötigte Interviewzeit etc., beobachtet werden. Daneben können von ihr subjektive Reaktionen, wie die Einstellung gegenüber dem Interview, die eingeschätzte Aufgabenschwierigkeit, die eingeschätzte Auskunftsmotivation, die erlebte Interviewdauer etc. erhoben werden (z.B. Huber et al., 1991). Der Grundgedanke dabei ist, daß objektive und subjektive Reaktionen durch die Eigenschaften und Einstellungen der Apn, durch die Interviewsituation und den Aufbau des Interviews beeinflußt werden. Bei der Erhebung der subjektiven Reaktionen sind bei der CA keine Besonderheiten gegenüber anderen Evaluationsbemühungen zu erwarten.

5.2 Reliabilität

Bei der multiattributiven Präferenzstrukturmodellierung durch die CA werden mehrere Nutzenwerte simultan geschätzt. Diese Werte sind voneinander nicht unabhängig. Aus diesem Grunde kann das Konzept der Reliabilität, wie es in der klassischen Testtheorie definiert ist (Lord & Novick, 1974), nicht direkt übertragen werden. In der Literatur zur CA wird dennoch zur Bewertung der Meßgüte von PSM der Grundgedanke des Konzepts der Test-Retest-Reliabilität herangezogen. Er besteht darin, daß man die Präferenzstruktur einer Ap zweimal unter vergleichbaren Bedingungen durch eine CA modelliert. Nimmt man an, daß die Präferenzen der Ap zwischen den beiden Messungen stabil geblieben ist, dann besteht eine hohe Reliabilität dann, wenn die beiden PSM möglichst ähnlich sind. Zur Bestimmung von Indikatoren der Reliabilität einer CA sind in der Literatur mehrere Vorgehensweisen zu finden.

Grundsätzlich kann die CA-Reliabilität auf mehreren Ebenen bestimmt werden. Zur Reliabilitätsbestimmung auf der *Ebene des Antwortvektors* (z.B. Acito, 1977) sind zwei Messungen mit denselben Beurteilungsobjekten notwendig. Die Antwortvektoren der beiden Messungen werden verglichen. Unter der Annahme, daß die Präferenzstruktur einer Ap über die beiden Meßzeitpunkte stabil geblieben ist, gibt ein Zusammenhangsmaß Hinweise auf die Reliabilität des Erhebungsverfahrens. Welche Maßzahl hierzu berechnet wird, hängt von der Erhebungsform der abhängigen Variablen ab. Bei einer Vollprofil-Rangreihung ist beispielsweise die Spearman-Rangkorrelation der Objekt-rangstufen ein angemessener Indikator.

Die Bestimmung dieses Reliabilitätsindikators auf der Ebene des Antwortvektors ist einfach durchzuführen. Sie hat jedoch einige Unzulänglichkeiten. Da bei diesem Vorgehen mehrmals die gleichen Objekte beurteilt werden müssen, können Gedächtniseffekte auftreten. Findet man zwei hoch übereinstimmende Rangreihen, könnte die Interpretation, daß diese durch stabile und reliabel meßbare Urteilsstrukturen zustande gekommen ist, falsch sein. Der Befund könnte auch durch Erinnerung der ursprünglichen Rangreihe entstehen. Ein weiterer Nachteil der Reliabilitätsbestimmung auf der Ebene des Antwortvektors ist, daß nur die Reliabilität des Inputs der CA bestimmt wird, nicht aber ein Reliabilitätsindikator der Präferenzstrukturmodellierung selbst. Diese wird jedoch

nicht nur durch Fehler in den Inputdaten gefährdet. Sie könnte auch z.B. vom verwendeten Objektset oder vom Design abhängig sein.

Um diese Nachteile zu vermeiden, kann die Reliabilität auf der *Ebene des PSM* abgeschätzt werden (z.B. Acito, 1979; McCulloch & Best, 1979). Bei einem Test-Retest-Ansatz werden die PSM der beiden Messungen miteinander verglichen. Als Maßzahlen sind alle diejenigen einsetzbar, die auch bei der Strukturwiedererkennung (Abschnitt 5.5) verwendet werden können (vgl. auch Bateson et al., 1987). Diese Vorgehensweise hat den Vorteil, daß man den Apn bei der Retestmessung andere Objekte zur Beurteilung vorgeben kann, als bei der Erstmessung (z.B. Parker & Srinivasan, 1976). Beispielsweise könnte das zweite Objektset ein anderes reduziertes Design sein. Beide Designs stellen aber eine Teilmenge des vollfaktoriellen Designs dar, das durch das Attributset bestimmt wird.

Diese Art der Reliabilitätsbestimmung ist sehr viel strenger als die erstgenannte, weil mehrere Fehlerquellen berücksichtigt werden, die die Reliabilität des Verfahrens bedrohen. Sie reagiert sensibel auf die fehlende Stabilität der Präferenzstruktur, Fehler durch Variationen des Objektsets, Fehler in der Schätzprozedur und Ungenauigkeiten der Beurteilungsantworten.

Eine weitere Ebene der Test-Retest-Reliabilitätsbestimmung schlagen Leigh et al. (1984) vor. Sie bestimmen die Gesamtnutzenwerte aller Objekte durch beide PSM und schätzen die Reliabilität durch deren Übereinstimmung. Genau nach demselben Muster kann auch jeweils ein Maß der Modellanpassung (vgl. Abschnitt 5.3), d.h. ein Kennwert der Übereinstimmung der geschätzten Nutzenwerte der Objekte mit den empirischen Beurteilungswerten (z.B. Rangfolge, Rating) und Kennwerte der Kreuzvalidität der beiden PSM berechnet werden. Bei der Kreuzvalidität wird die Übereinstimmung zwischen den geschätzten Nutzenwerten der Objekte eines PSM und den empirischen Beurteilungswerten der jeweilig anderen CA berechnet. Die verschiedenen Indizes (Reliabilität, Modellanpassung, Kreuzvalidität; z.B. durch Rangkorrelation oder R^2) können miteinander verglichen werden, um den Einfluß verschiedener Fehlerquellen abzuschätzen.

Das Prinzip der verschiedenen Ansätze zur Abschätzung der Reliabilität und die damit in Zusammenhang stehenden Konzepte der Modellanpassung und der Kreuzvalidität (vgl. auch Abschnitte 5.3 und 5.4) ist in Abbildung 5.1 skizziert.

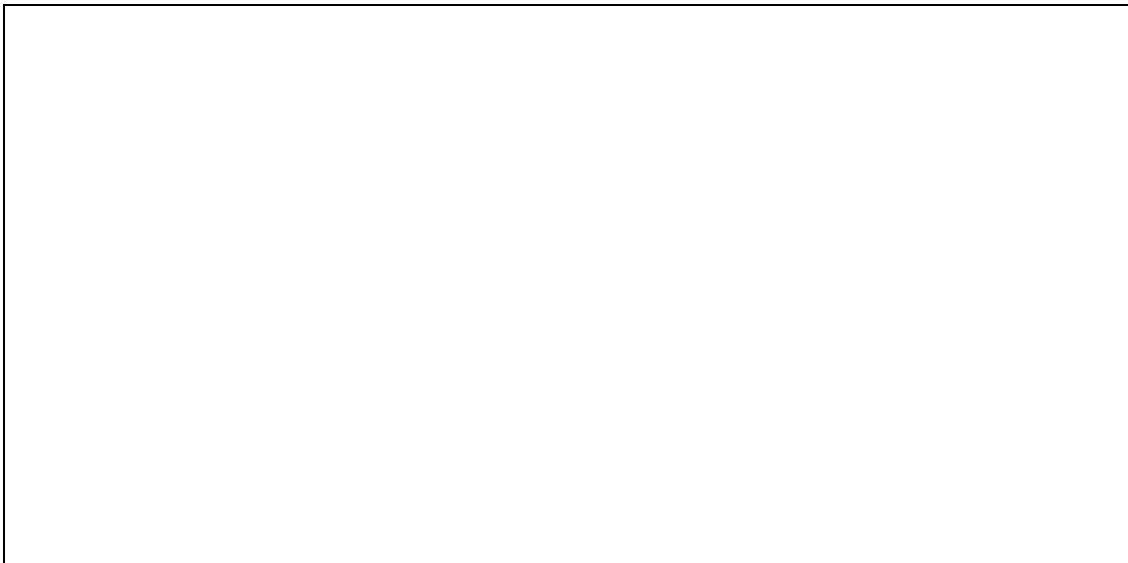


Abbildung 5.1: Die Konzepte Reliabilität, Modellanpassung und Kreuzvalidität bei ähnlichen Conjoint Analysen I und II (Pfeil: Vorhersage, Doppelpfeil: Zusammenhang).

5.3 Modellanpassung

Das Konzept der Modellanpassung berührt die Frage nach der Angepaßtheit (fit) und der Angemessenheit des geschätzten PSM (vgl. Abschnitte 3.4.2.2 und 3.4.2.3). Die *Modellanpassung* wird vergleichsweise häufig bestimmt. Sie stellt eine notwendige Bedingung für die Modellangemessenheit dar. Die passende Maßzahl der Modellanpassung hängt vom angenommenen Skalenniveau der Antwortvariablen und dem Parameterschätzverfahren ab. Wird die Antwort ordinalskaliert angenommen und die Parameter mit MONANOVA geschätzt, kann der Stress-Wert (Backhaus et al., 1994) nach der letzten Iteration als Maßzahl des Modellfits herangezogen werden. Er kann als Indikator für die Güte der Reproduktion der ordinalen Ausgangsdaten durch die metrisch skalierten Teilnutzenwerte des PSM interpretiert werden. Wird die Antwort intervallskaliert angenommen und die Parameter des PSM per OLS-Regression geschätzt, kann in analoger Weise die Maßzahl R^2 zur Beurteilung der Modellanpassung dienen. Die SPSS-Prozedur CONJOINT berechnet die Produkt-Moment-Korrelation zwischen den vorhergesagten (metrischen) Nutzenwerten und den

(normalerweise ordinalskalierten) empirischen Beurteilungswerten der Objekte des Objektsets.

Green, Krieger und Agarwal (1991) bestimmen die Modellanpassung von abgestuften Paarvergleichsurteilen bei ACA. Sie berechnen die Übereinstimmung von vorhergesagten und tatsächlichen Antworten beim Paarvergleich. Als Maßzahlen verwenden sie die Produkt-Moment-Korrelation und den mittleren absoluten Fehler (MAE) zwischen den Modellvorhersagen und den empirischen Daten.

Der Begriff der Modellanpassung kann sich auch auf die *Modellangemessenheit* beziehen. Bei der statistischen Modellanpassung fragt man sich, ob ein PSM die empirischen Daten gut reproduzieren kann. Im Gegensatz dazu beschäftigt sich die Modellangemessenheit mit der Frage, ob die Struktur des PSM (Präferenzmodell und Integrationsmodell) der realen Urteils-/ Entscheidungsstruktur entspricht. Dies ist weit schwieriger zu beurteilen. Die Prüfung der Modellangemessenheit erfolgt meist durch die Überprüfung der Axiome des CM (vgl. Abschnitte 2.3.3, 3.4.2.3 und 3.6.2.4).

5.4 Kriterienorientierte Validität

Im Zusammenhang mit der kriterienorientierten Validität sind in der Literatur zur CA einige begriffliche Verwirrungen vorhanden. Daher möchten wir zunächst einige Vorbemerkungen zu unserem Verständnis der Begriffe und deren Abgrenzung vorausschicken.

Die Unterscheidung zwischen den Begriffen Modellanpassung und kriterienorientierte Validität ist nicht immer ganz eindeutig. Beide Begriffe beziehen sich auf die Vorhersagegüte des Modells auf empirische Daten. Im Regelfalle benutzt man den Begriff Modellanpassung, wenn es um die Vorhersage von Daten geht, aus denen die Parameter des Modells geschätzt wurden. Handelt es sich um Daten, die keinen Einfluß auf die Parameterschätzung haben, spricht man von der kriterienorientierten Validität.

Die manchmal nicht eindeutige Unterscheidbarkeit der beiden Begriffe werden an folgendem Beispiel deutlich: ACA 4.0 verwendet die Konzeptratings in Phase IV zur Integration und Reskalierung der bisherigen Parameterschätzungen (logistische Regression der expliziten und Paarvergleichs-TNW auf die Kaufwahrscheinlichkeitsurteile) und berechnet die Modellanpassung R^2 .

Insofern handelt es sich um eine Maßzahl der Modellanpassung. Andererseits könnte auch die Vorhersagegüte der Konzeptratings durch das nichtreskalierte PSM als Indikator der kriterienorientierten Validität herangezogen werden.

Das Validitätskriterium kann entweder direkt mit dem CA-Interview erhoben werden oder es wird außerhalb dieses Kontextes ermittelt. Im ersteren Falle handelt es sich meist um die Beurteilung zusätzlicher Objekte ('holdout concepts', Holdout-Objekte, Holdout-Konzepte, Kreuzvalidierungsobjekte). Die Vorhersagegüte dieser Holdout-Beurteilungen durch das PSM bezeichnen viele Autoren als 'interne Validität'. Dieser Begriff ist etwas unglücklich gewählt, weil er im Zusammenhang mit der Beurteilung von experimentellen Ansätzen geprägt wurde (Cook & Campbell, 1979) und hierfür reserviert bleiben sollte. Der Begriff der 'kriterienorientierten Validität' bezieht sich dagegen auf die Vorhersage eines Kriteriums durch Prädiktoren, in diesem Fall auf die Vorhersage des Kriteriums durch das PSM. Wir schließen uns damit der Begriffswahl und Begründung von Leigh et al. (1984) an und sprechen, wie dies im Zusammenhang mit der Beurteilung von Meßverfahren üblich ist, von der *kriterienorientierten Validität* der PSM, wenn Daten durch das PSM vorhergesagt werden, die keinen Einfluß auf die Schätzung der Parameter haben. Wird das Kriterium erst einige Zeit nach der CA-Messung realisiert, verwenden wir den Begriff *prognostische Validität*. Sie ist ein Sonderfall der kriterienorientierten Validität (vgl. Abschnitt 5.4.4).

Häufig wird auch die begriffliche Unterscheidung 'interne vs. externe Validität' getroffen. Die betreffenden Autoren versuchen auf diese Weise dem Umstand gerecht zu werden, daß das Validierungskriterium im Kontext der CA-Aufgabe erhoben wird und dieser damit sehr ähnlich sein kann ('interne Validität'), oder daß das Kriterium in einem externen Kontext gemessen wird und der CA-Aufgabe meist sehr unähnlich ist ('externe Validität'). Ein Beispiel für ein typisches 'internes' Kriterium wäre die Rangreihung einiger Holdout-Vollprofile direkt im Anschluß an eine Vollprofil-CA. Ein sehr unähnliches Kriterium dagegen wäre die Beobachtung von tatsächlichen Kaufentscheidungen im Objektbereich außerhalb der CA-Situation (individuelle Ebene) bzw. tatsächliche Marktanteile von Produkten des Objektbereichs (aggregierte Ebene). Wir möchten uns aus bereits genannten Gründen (Prägung des Begriffs in einem anderen Zusammenhang) dieser begrifflichen Unterscheidung ebenfalls nicht anschließen. Statt dessen sollten Validierungsbemühungen stets kritisch und in

der üblichen Form auf das Zustandekommen der Ergebnisse durch Artefakte geprüft werden. Ist die Kriteriumsaufgabe den Aufgaben der CA sehr ähnlich, kann ein hoher Zusammenhang zwischen Prädiktor und Kriterium auch der Methodenähnlichkeit entspringen. In diesem Fall bedroht Kriterienkontamination die richtige Interpretation der Maßzahlen der kriterienorientierten Validität.

Eine letzte wichtige Unterscheidung ist die der Validierungsebene. Die kriterienorientierte Validität von PSM kann auf *individueller Ebene* bestimmt werden. Man geht dabei so vor, daß das PSM und das Kriterium für jede Ap bestimmt wird. Die Güte der Vorhersage eines Kriteriums durch das PSM einer Ap ist ein Indikator für dessen kriterienorientierte Validität. Übliche Kriteriumsaufgaben und Maßzahlen der kriterienorientierten Validität auf individueller Ebene werden in Abschnitt 5.4.1 beschrieben. Ein anderer Weg der Validitätsbestimmung besteht darin, daß ein Kriterium auf *aggregierter Ebene* erhoben wird. Wahl- oder Marktanteile auf Stichproben- oder Segmentebene sind typische Kriterien bei dieser Vorgehensweise. Abschnitt 5.4.2 beschäftigt sich mit dieser Art der kriterienorientierten Validitätsbestimmung.

5.4.1 Kriterienorientierte Validität auf individueller Ebene

Eine weitverbreitete Vorgehensweise bei der Ermittlung der kriterienorientierten Validität auf individueller Ebene besteht darin, daß man den Apn direkt nach der Erhebung der CA-Daten mehrere Holdout-Vollprofilobjekte zur Beurteilung vorgibt. Die Apn antworten z.B. durch Rangreihung dieser Objekte gemäß ihrer Präferenz. Als Validitätsmaßzahl kann ein *Zusammenhangsmaß* zwischen der vorhergesagten und der empirischen Rangreihe dienen. Als Maßzahl des Zusammenhangs zweier Rangreihen wird z.B. die Spearman-Rangkorrelation, (z.B. Agarwal, 1989; Jain, Acito, Malhotra & Mahajan, 1979; Srinivasan et al., 1983) oder Kendalls tau (z.B. Johnson, 1974; Safizadeh, 1989) verwendet.

Antworten die Apn auf die Holdout-Objekte durch ein Präferenzrating, besteht eine angemessene Maßzahl der kriterienorientierten Validität in der Produkt-Moment-Korrelation zwischen den Gesamtnutzenwerten und den empirischen Ratingwerten (z.B. Agarwal, 1989; Moore & Semenik, 1988; van der Lans & Heiser, 1992). Als weitere Maßzahlen sind auch der *Prozentsatz von Rangpositionstreffern* (z.B. Akaah & Korgaonkar, 1983; Anderson & Donthu, 1988; Green, Schaffer & Patterson, 1991) und die *Summe der absoluten Rangplatz-*

abweichungen (z.B. Acito & Jain, 1980) zu finden. Ein Rangpositionstreffer ist dann gegeben, wenn der vorhergesagte Rangplatz eines Holdout-Konzepts mit dem empirischen Rangplatz übereinstimmt. Die Summe der absoluten Rangplatzabweichung berechnet sich aus der Summe der Abweichungen der vorhergesagten Rangplätze von den empirischen Rangplätzen.

Einige Untersuchungen geben den Apn als Validierungsaufgabe mehrere Sets von Holdout-Konzepten zur Beurteilung vor. Typischerweise erhalten die Apn die Aufgabe, aus jedem Holdout-Set jenes Konzept auszuwählen, das sie unter den jeweiligen Alternativen am meisten präferieren. Als Maßzahl kann der *Prozentsatz der Erste-Wahl-Treffer auf individueller Ebene* verwendet werden (z.B. Elrod et al., 1992; Huber et al., 1993). Erste-Wahl-Treffer bezeichnen die Übereinstimmung des Holdout-Objekts mit dem höchsten Gesamtnutzen mit dem höchstpräferierten Objekt des Holdout-Sets. Eine ähnliche Vorgehensweise findet sich bei Huber und Hansen (1986). Sie gaben ihren Apn drei Holdout-Sets mit je vier Vollprofilkonzepten vor. Die Apn hatten die Aufgabe, das Konzept der ersten und der zweiten Wahl anzugeben. In dieser Aufgabe sind je Set fünf implizite Paarvergleiche enthalten (1>2, 1>3, 1>4, 2>3, 2>4). Als Validitätsmaßzahl berechnen die Autoren den *Anteil der korrekt vorhergesagten impliziten Paarvergleiche*.

Die Verwendung solcher Wahlaufgaben zur Validierung von CA-Untersuchungen haben gegenüber der Rangreihung oder dem Rating von Holdout-Objekten insbesondere den Vorteil, daß die Kriteriumsaufgabe gegenüber der Aufgabe der CA-Datenerhebung unähnlicher wird. Dadurch werden alternative Interpretationen der gefundenen Zusammenhänge zwischen PSM-Vorhersage und Kriteriumsaufgabe als Methodenartefakte erschwert. Daß manche Forscher großen Wert auf die Verschiedenheit der CA-Interviewaufgabe und der Kriteriumsaufgabe legen, zeigt die Untersuchung von Huber et al. (1993). Sie verwenden zur Validierung einer Vollprofil-CA und einer ACA-Untersuchung mehrere Wahlsets, aus welchen jeweils das meistpräferierte Objekt auszuwählen ist. Die Autoren betonen die Verschiedenheit der Aufgaben zur Nutzenschätzung und der Kriterienaufgabe (z.B. Antwortmodus: Beurteilung durch Rating vs. Entscheidung durch Kategorienzuordnung; Profilstärke der Objekte: bei ACA Zweier- und Dreierprofil vs. Fünferprofile bei der Kriteriumsaufgabe).

Es gibt auch Ansätze, die Validität der PSM auf der Grundlage von theoretischen Überlegungen abzuschätzen. Eine Vorgehensweise besteht in der

Überprüfung, ob Teilnutzenwertmodelle gut fundierte a priori Erwartungen über Nutzenverläufe innerhalb bestimmter Attribute bestätigen (z.B. Acito & Jain, 1980). Gibt es beispielsweise gute Gründe für die Annahme, daß in einem Objektbereich der Nutzen eines Objekts mit der Zunahme seines Preises bei allen Apn abnehmen sollte, dann ist die Validität eines TNW-Modells, das diesen Nutzenverlauf nicht zeigt, in Frage zu stellen. Als Maßzahl könnte der *Anteil der Verletzung von a priori Annahmen* dienen. Bei dieser Art der Validitätsbestimmung ist besonders darauf zu achten, daß das Kriterium der a priori Annahmen valide ist. Man ist deshalb gut beraten, wenn diese Annahmen sehr konservativ getroffen werden.

Die ermittelte kriterienorientierte Validität eines PSM wird von der Reliabilität der Kriteriumsaufgabe nach oben begrenzt. Viele Autoren vernachlässigen diesen Sachverhalt. Sie verhalten sich so, als ob das Kriterium mit einer perfekten Reliabilität erhoben werden könnte, oder sie nehmen implizit an, daß die Kriterienreliabilität über alle Apn konstant ist. Beide Annahmen sind nicht einfach zu rechtfertigen und können Untersuchungsergebnisse verzerren.

Die Haltbarkeit beider Annahmen ist nicht nur anzuzweifeln. Es gibt sogar einen guten theoretischen Grund, die Konstanz der Kriterienreliabilität zwischen verschiedenen Apn in Frage zu stellen: Die Präferenzstrukturmodellierung durch die CA wird deshalb auf individueller Ebene vorgenommen, weil angenommen wird, daß sich die Präferenzstrukturen der Apn unterscheiden. Ist diese Annahme sinnvoll, dann bedeutet dies aber auch, daß die Nutzendifferenzen der (in aller Regel über die Apn konstanten) Kriteriumsobjekte ebenfalls verschieden sind. Es könnte der Fall eintreten, daß eine Ap Kriterienobjekte zur Beurteilung erhält, deren Nutzen sich sehr stark unterscheiden. Sie hätte kaum Probleme, diese gemäß ihrer Präferenz z.B. in eine Rangreihe zu bringen. Dies bedeutet aber wiederum, daß die Reliabilität der Kriteriumsaufgabe hoch ist. Eine andere Ap erhält die gleichen Kriterienobjekte. Da sie jedoch eine andere Präferenzstruktur besitzt, könnten die Kriterienobjekte für sie sehr nutzenähnlich sein. Sie hat große Schwierigkeiten bei der Rangreihenbildung, was zu einer niedrigeren Kriterienreliabilität führen sollte. Mit anderen Worten: Aus der subjektiven Sicht der Apn ist die Schwierigkeit der objektiv gleichen Kriteriumsaufgabe unterschiedlich. Nur sehr wenige Untersuchungen berücksichtigen diesen Sachverhalt. Abbildung 5.2 veranschaulicht diese Denkweise schematisch.

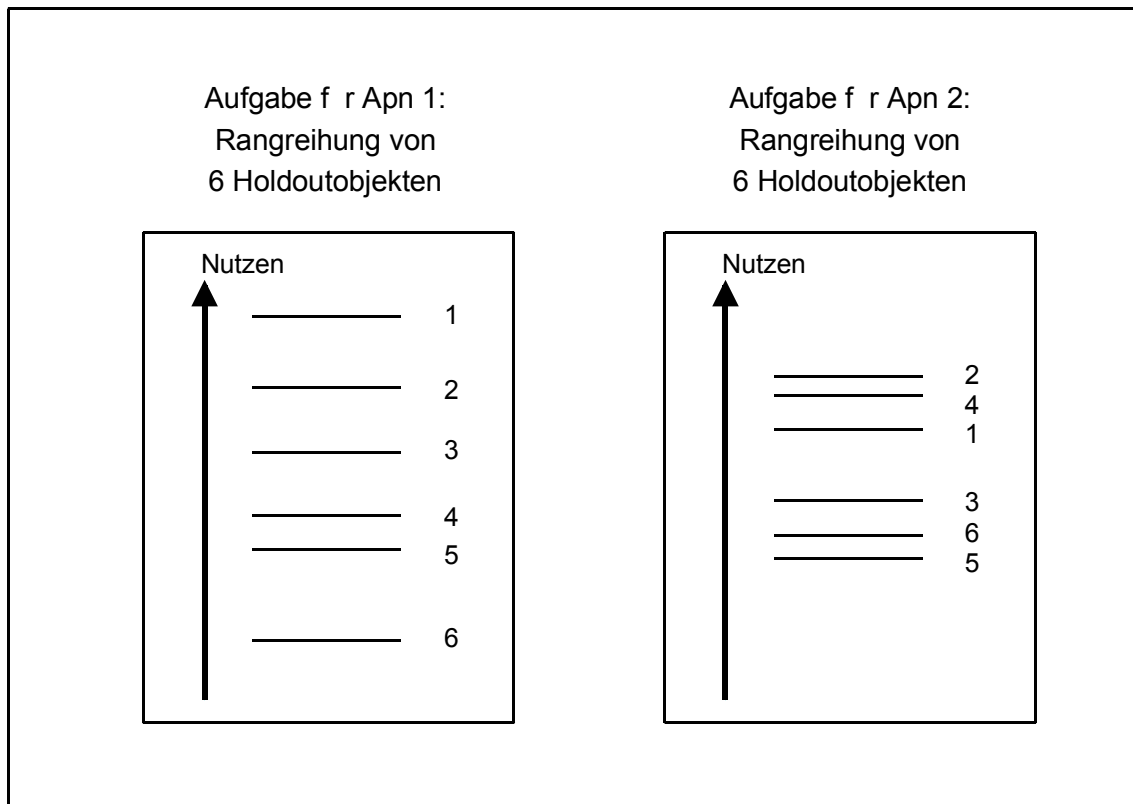


Abbildung 5.2: Schematische Darstellung des Reliabilitätsproblems der Kriteriumsaufgabe. Individuell verschiedene Nutzenstrukturen haben zur Folge, daß die Schwierigkeit der Kriteriumsaufgabe - trotz objektiv gleicher Beurteilungsobjekte - individuell unterschiedlich ist.

Prinzipiell sind zwei Möglichkeiten denkbar, mit dem *Reliabilitätsproblem der Kriteriumsaufgabe* umzugehen. Entweder man hält die *Schwierigkeit der Kriteriumsaufgabe über die Untersuchungsgruppen konstant* (durch Verwendung von Meßwiederholungsversuchsplänen; bei between subject designs durch Konstanthaltung der Schwierigkeit der Urteilsaufgabe, z.B. indem die Nutzendifferenzen der Kriterienobjekte zwischen den Apn vergleichbar sind), oder man schätzt die Kriterienreliabilität auf individueller Ebene und korrigiert die kriterienorientierte Validität entsprechend der individuellen Reliabilität statistisch (*Minderungskorrektur*).

McLauchlan (1991) erhebt die Test-Retest-Reliabilität seiner Kriteriumsaufgabe (Rangreihung von fünf Holdout-Konzepten, identische Konzepte jeweils vor und nach der Nutzenmessung). Er stellt einen Unterschied hinsichtlich der Kriterienreliabilität zwischen seinen experimentellen Gruppen fest. Diese Unterschiede beeinflussen natürlich auch die ermittelten Validitätsmaßzahlen, was

ohne Berücksichtigung der verschiedenen Kriterienreliabilitäten zu einer falschen Interpretation seiner Untersuchungsergebnisse führen würde. McLauchlan (1991) begegnet diesem Problem durch Minderungskorrekturen und der gesonderten Auswertung von 'konsistenten Apn'. Dabei handelt es sich um eine Teilmenge der Apn, deren Kriterium besonders reliabel erhoben wurde.

5.4.2 Kriterienorientierte Validität auf aggregierter Ebene

Die Mehrzahl der Untersuchungen zur Beurteilung von CA-Verfahren verwenden als Hauptbeurteilungskriterium die kriterienorientierte Validität der geschätzten PSM hinsichtlich Holdout-Konzepten auf individueller Ebene. Einige Autoren haben jedoch in jüngerer Zeit darauf hingewiesen, daß die Beurteilung der Leistungsfähigkeit eines CA-Verfahrens auf dieser Grundlage inadäquat sein kann (Hagerty, 1993; Johnson et al., 1989; Wedel & Steenkamp, 1991). Dies ist insbesondere deswegen der Fall, weil die Untersuchungsergebnisse sehr sensibel auf das spezielle Set der Holdout-Objekte zu reagieren scheinen (Kamakura, 1988; Wedel & Steenkamp, 1991).

Aus diesem Grunde empfiehlt es sich, mehrere Meßgütekriterien zu verwenden. Zusätzlich zur kriterienorientierten Validität auf individueller Ebene kann z.B. auch die kriterienorientierte Validität auf aggregierter Ebene bestimmt werden. Dies erfordert kaum mehr Erhebungsaufwand und ist deswegen besonders interessant, weil Hagerty (1986) einen schwierig erklärbaren empirischen Befund erhalten hat. Er verglich zwei CA-Verfahren. Das erste erbrachte höhere Validitätsindikatoren auf der individuellen Ebene als das zweite Verfahren. Gleichzeitig war jedoch das zweite CA-Verfahren dem ersten hinsichtlich der kriterienorientierten Validität auf aggregierter Ebene überlegen. Die Ergebnisse dieser Untersuchung zeigen, wie sensibel die Validitätsmaße auf spezielle Bedingungen der CA reagieren können.

Die einfachste Form der Ermittlung der kriterienorientierten Validität auf aggregierter Ebene besteht in der anteilmäßigen Bestimmung des höchsten Gesamtnutzenwerts, den jedes der Holdout-Objekte in der Apn-Stichprobe erringen kann. Dieser Prozentsatz wird mit der relativen Häufigkeit verglichen, mit der jedes Holdout-Objekt den ersten Platz im Holdout-Set einnimmt. Eine geeignete Maßzahl wäre der *Prozentsatz von Erste-Wahl-Treffern auf aggregierter Ebene* (first-choice-hits; z.B. Green, Krieger & Agarwal, 1991; Green,

Schaffer & Patterson, 1991; van der Lans & Heiser, 1992). Ein Erste-Wahl-Treffer ist damit die Übereinstimmung des höchstpräferierten Holdout-Objekts mit dem Holdout-Objekt, das den höchsten Gesamtnutzen besitzt. Analog kann mit dem zweiten, dritten, ... und letzten Rangplatz verfahren werden.

Die kriterienorientierte Validität der PSM auf aggregierter Ebene kann auch mit Hilfe von Wahlsimulationen (Abschnitt 4.4) bestimmt werden. Das Grundprinzip dieser Vorgehensweise besteht darin, die empirischen Erste-Wahl-Anteile im Holdout-Set mit Wahlanteilsimulationsergebnissen zu vergleichen. Eine gewisse Ähnlichkeit mit der Bestimmung des Prozentsatzes von Erste-Wahl-Treffern auf aggregierter Ebene hat der Einsatz des Maximum-utility Wahlsimulators. Die vorhergesagten Wahlanteile werden mit den empirischen Erste-Wahl-Anteilen bei der Kriteriumsaufgabe verglichen. Als Maßzahl kann der *mittlere absolute Fehler (MAE) zwischen vorhergesagten und empirischen Wahlanteilen* dienen (z.B. Elrod et al., 1992; Oppewal et al., 1994; van der Lans & Heiser, 1992). Elrod et al. (1992) schlagen als geeignete Maßzahl des Wahlanteilvergleichs außerdem *PRMSEP (proportional reduction in mean square error of prediction)* vor. Die Maßzahl ist identisch mit dem mittleren quadrierten Fehler (MSE, mean square error), nachdem die vorhergesagten und empirischen Wahlanteile einer Transformation unterzogen wurden. Die Wahlanteile werden deswegen transformiert, weil die Varianz eines beobachteten Wahlanteils von seinem Erwartungswert abhängt. Darüber hinaus ist die Verteilung eines Wahlanteils - außer beim Erwartungswert 0,5 - nicht symmetrisch. Die Transformationsvorschrift ist bei Elrod et al. im Anhang angegeben.

Statt des Maximum-utility-Simulators sind prinzipiell natürlich auch andere Wahlanteilsimulatoren einsetzbar. Im Gegensatz zum deterministischen Maximum-utility-Modell transformieren diese die intervallskalierten Gesamtnutzenwerte in Wahlwahrscheinlichkeitswerte, bevor diese über die Apn aggregiert werden (vgl. Abschnitt 4.4). Huber et al. (1993) verwenden z.B. das Logit-Modell zur Wahlanteilsimulation und MAE (mean absolute error) als Maßzahl.

In der Literatur finden sich weitere Maßzahlen zum Vergleich der vorhergesagten und empirischen Wahlanteile. Jain, Acito, Malhotra & Mahajan (1979) berechnen den *Prozentsatz der richtig vorhersagten Wahlanteilrangplätze* und die *Summe der absoluten Wahlanteilrangplatzabweichungen* (mit dem Wertebereich jeweils von 0 bis Anzahl der Ränge - 1). Eine andere Möglichkeit besteht

in der Bestimmung der *Produkt-Moment-Korrelation zwischen tatsächlichem und vorhergesagtem Marktanteil* (z.B. Oppewal et al., 1994; zur allgemeinen Kritik der Marktanteilsvorhersage, siehe Abschnitt 4.4.6).

5.4.3 Welche Art der Validierung und welche Maßzahl?

Es gibt eine Vielzahl von Möglichkeiten, die kriterienorientierte Validität der PSM auf individueller und aggregierter Ebene abzuschätzen. Tabelle 5.1 unterscheidet die verschiedenen Arten der Gewinnung eines Validierungskriteriums und die Validierungsebene. Die Tabelle 5.1 veranschaulicht die im Text genannten Maßzahlen und ordnet diese der Validierungsebene und der Art der Kriteriengewinnung zu.

Welche der Vorgehensweisen sinnvoll ist, hängt, wie auch bei Validierungsstudien in anderen Forschungsbereichen, von der intendierten Verwendung der CA-Ergebnisse ab. Wird eine CA beispielsweise vornehmlich zu Zwecken der Segmentierung einer Käuferpopulation durchgeführt, steht die kriterienorientierte Validität der PSM eines jeden Individuums der (repräsentativen) Stichprobe im Vordergrund. Geht es dagegen um die Vorhersage des Marktanteils eines neuen Produkts, sollte die kriterienorientierte Validität der PSM auf aggregierter Ebene wichtiger sein.

Tabelle 5.1: Systematisierung der Validierungsebene und der Art der Kriteriengewinnung bei der Beurteilung der kriterienorientierten Validität von PSM. Die Validitätsmaßzahlen beziehen sich auf den Vergleich zwischen vorhergesagten und tatsächlichen Kriteriumsbeurteilungen. Weitere Erläuterungen im Text.

Validierungsebene	Art der Kriteriengewinnung		
	theoretische Überlegungen	Urteile in einem Holdout-Objekte-Set	Urteile in mehreren Holdout-Objektsets
individuelle Ebene	Verletzungen von a priori Annahmen	Zusammenhangsmaß Rangpositionstreffer implizite Paarvergleichstreffer Rangplatzabweichungen	Erste-Wahl-Treffer implizite Paarvergleichstreffer
aggregierte Ebene		Erste-Wahl-Treffer Wahlanteilsimulation und - Zusammenhangsmaß - Rangplatztreffer - Rangplatzabweichungen - MAE/ PRMSEP	

Welche Maßzahl die kriterienorientierte Validität conjointanalytischer Ergebnisse am besten repräsentiert, kann ebenfalls nicht losgelöst vom Untersuchungskontext beantwortet werden. Im allgemeinen liefern verschiedene Maßzahlen sehr ähnliche Ergebnisse und legen damit auch ähnliche Interpretationen nahe. Bestehen jedoch Unterschiede zwischen verschiedenen Maßzahlen, dann sollte wieder der Untersuchungszweck zur Auswahl des geeignetsten Maßes dienen. Ein fiktives Beispiel kann dies verdeutlichen: Ein Unternehmen führt im Rahmen einer Produktinnovation einen Produktkonzepttest durch. Man möchte entscheiden, welches von mehreren Produktkonzepten in Serie hergestellt werden soll. Eine CA soll über die Bestimmung der relevanten Präferenzstrukturen als Entscheidungshilfe dienen. Es besteht jedoch noch Unsicherheit darüber, welches von zwei möglichen CA-Verfahren (A und B) zum Einsatz kommen soll. Eine Vorstudie zur Bestimmung der kriterienorientierten Validität auf individueller Ebene zeigt, daß die Vorhersagen der PSM des Verfahrens A eine mittlere Produkt-Moment-Korrelation zur Beurteilung mehrerer Holdout-Objekte-Sets von $\rho=0,8$ erbringen. Das Verfahren B leistet dagegen nur $\rho=0,7$. Betrachtet man jedoch den mittleren Prozentsatz der Erste-Wahl-Treffer auf individueller Ebene, dann ist Verfahren B mit 70% gegenüber Verfahren A mit 60% überlegen.

Nehmen wir an, daß beide Unterschiede mit derselben Irrtumswahrscheinlichkeit signifikant sind, also inferenzstatistische Überlegungen zu keiner Entscheidung führen können. Welches Verfahren besitzt für die intendierten Untersuchungszwecke eine höhere Validität und sollte damit (unter sonst gleichen Bedingungen) bei der Hauptuntersuchung eingesetzt werden? Wir würden uns für Verfahren B entscheiden, weil dieses besser in der Lage ist, die Beurteilungen und Entscheidungen in den Bereichen hoher Präferenz vorherzusagen (Rangplatz 1). Verfahren A zeigt zwar eine Überlegenheit, was die Vorhersage der gesamten Holdout-Rangreihen betrifft. Es kann demnach Beurteilungen und Entscheidungen in den Bereichen mittlerer und niedriger Präferenz besser als B vorhersagen. Dies ist jedoch von untergeordnetem Interesse, da Kaufentscheidungen ausschließlich in den Bereichen hoher Präferenz zu erwarten sind.

5.4.4 Prognostische Validität

Kann die kriterienorientierte Validität eines Prädiktors nicht sofort bestimmt werden, weil das Kriterium erst in einigem zeitlichen Abstand meßbar ist, spricht man auch von der *prognostischen Validität*. Positive Befunde zur prognostischen Validität sind meist sehr überzeugend. Dies liegt daran, daß Alternativinterpretationen, die die Befunde durch Methodenartefakte zu erklären versuchen, meist leichter widerlegt werden können als bei der konkurrenten, kriterienorientierten Validierung. Wir wollen beispielhaft einige Untersuchungen anführen, die diesen Weg der Überprüfung von PSM gehen. In diesem Abschnitt möchten wir auch Untersuchungen besprechen, die zeitlich sehr stabile Kriterien verwenden. Zwar erheben diese Untersuchungen oft das Kriterium zeitgleich mit den Präferenzstrukturen. Es handelt sich demnach nicht um die 'prognostische Validität' im wörtlichen Sinn. Durch die Stabilität des Kriteriums spielt jedoch der Zeitpunkt der Kriterienmessung keine große Rolle. Dies wäre z.B. bei stabilen Marktanteilen von Produkten der Fall. Die Vorhersagegüte von PSM wird in diesem Fall nicht davon beeinflusst, ob das Kriterium vor, nach oder während der Nutzenmessung erfolgt.

Die prognostische Validität kann, wie dies allgemein bei der kriterienorientierten Validierung möglich ist, auf individueller Ebene und auf aggregierter Ebene bestimmt werden. Eine Reihe von Studien belegt die Qualität conjoint-

analytischer Ergebnisse durch prognostische Validierung auf individueller Ebene (z.B. Krishnamurthi, 1988; Mohn, 1990; Montgomery & Wittink, 1980; Srinivasan, 1988).

Wird versucht, die prognostische Validität auf aggregierter Ebene zu bestimmen, werden in aller Regel spätere Marktanteile als Kriterium verwendet (zur Problematik der Verwendung von Marktanteilen als Kriterium, siehe Abschnitt 4.4.6). Es liegen einige Untersuchungen vor, die diesen Weg der Validierung gehen (z.B. Benbenisty, 1983; Louviere & Woodworth, 1983; Oppewal et al., 1994; Page & Rosenbaum, 1987).

Fiedler (1988) führte eine CA durch, um Präferenzen für Eigentumswohnungen zu untersuchen (mit ACA). Er vergleicht die gemittelten PSM-Parameter mit den Regressionsgewichten der Wohnungsattribute auf tatsächlich erzielte Verkaufspreise. Das gemittelte PSM der CA stimmt mit den Parametern des Regressionsmodells mehr oder weniger gut überein. Die größten Vorhersagefehler treten beim Wohnungsattribut 'Aussicht auf den nahegelegenen Fluß' auf. Die relative Wichtigkeit dieses Attributs wird von der CA weit überschätzt. Fiedler (1988) interpretiert diesen Befund dahingehend, daß in der Interviewsituation emotional beladene Attribute die Beurteilungen stärker beeinflussen als in der tatsächlichen Kaufsituation. Dennoch kann festgehalten werden, daß das gemittelte PSM eine erhebliche Validität hinsichtlich des harten Kriteriums 'erzielter Verkaufspreis' besitzt.

5.5 Strukturwiedererkennung

Will man die Meßgüte eines Verfahrens zur Nutzenschätzung überprüfen, sind die Maße der Reliabilität, der Modellanpassung an die Antworten der Apn und der kriterienorientierten Validität der geschätzten PSM prinzipiell bei jeder CA bestimmbar. Anders verhält es sich mit dem Meßgütekriterium der *Wiedererkennung einer wahren Nutzenstruktur*. Sie verlangt ein besonderes Untersuchungsparadigma. Da eine notwendige Bedingung für die Bestimmung der Strukturwiedererkennung in der Kenntnis der 'wahren Nutzenstruktur' liegt, ist dieses Gütemaß bisher nur bei Computersimulationsstudien zum Einsatz gelangt (z.B. Agarwal, 1988a; Cattin & Bliemel, 1978, Colberg, 1978; Teichert, 1994). Dabei geht man bei diesen Simulationsstudien prinzipiell so vor, daß ein Computer gemäß einer festgelegten Nutzenstruktur ('wahres PSM') die Ant-

worten im Interviewteil einer CA simuliert. Die Antworten des Antwortsimulators können dabei fehlerfrei sein ('idealer Beurteiler') oder von einem festgelegten (und meist normal verteilten) Fehler überlagert werden. Als Kriterium der Meßgüte wird bestimmt, wie gut die CA das wahre PSM wiedererkennt, das dem Antwortsimulator zugrunde liegt und die CA-Inputdaten generierte.

5.5.1 Beurteilung der Strukturwiedererkennungsleistung

Um die Strukturwiedererkennungsleistung eines Nutzenschätzverfahrens beurteilen zu können, muß das geschätzte PSM und das wahre PSM miteinander verglichen werden. Damit man diesen Vergleich sinnvoll durchführen kann, sind zwei Probleme zu bewältigen. Erstens muß das wahre PSM und die geschätzten PSM miteinander vergleichbar gemacht werden. Zweitens muß man eine geeignete Maßzahl finden, die die Ähnlichkeit zwischen zwei PSM möglichst unverzerrt abbildet.

Das erste Problem wird durch Normierung gelöst. Da die TNW eines PSM eine Abbildung auf einer Intervallskala sind, bleiben ihre Eigenschaften gegenüber linearen Transformationen invariant. Vor der Bildung einer Ähnlichkeitsmaßzahl kann man die PSM mit dem Ziel transformieren, alle PSM miteinander vergleichbar zu machen. In Abschnitt 4.1 sind mehrere Möglichkeiten der Normierung beschrieben. Eine geeignete Methode der Normierung besteht z.B. in der Zentrierung der TNW innerhalb der Attribute zum Ausgleich von Niveauunterschieden und der nachfolgenden Standardabweichungsnormierung (StaNorm).

Nachdem alle PSM vergleichbar sind, kann man einen geeigneten Indikator bilden, der die Ähnlichkeit zwischen den empirischen PSM und dem wahren Modell beschreibt. Mehrere Maßzahlen sind hierfür denkbar. Agarwal (1983, zit. in Agarwal, 1988a) untersuchte die Korrelation mehrerer Indikatoren der Strukturwiedererkennung. Er fand eine sehr hohe Redundanz der Maßzahlen. Agarwal (1988a) verwendet aus diesem Grund in seiner Monte-Carlo-Simulationsstudie nur einen Ähnlichkeitsindikator. Es handelt sich in diesem Fall um die *Produkt-Moment-Korrelation zwischen den geschätzten und den wahren TNW*. Weitere denkbare Maßzahlen wären auch *SSE (Sum of Square Error)* *MSE (Mean Square Error)* zwischen dem wahren und dem geschätzten TNW-Modell.

Abbildung 5.3 veranschaulicht die Vorgehensweise bei der Normierung und der Ähnlichkeitsmaßzahlbildung bei zwei dreifach gestuften Attributen schematisch. Hierzu werden die TNW zunächst innerhalb der Attribute zentriert. Es folgt die Standardabweichungsnormierung (StaNorm) durch Multiplikation der TNW mit einem PSM-spezifischen Faktor so, daß die Standardabweichung der TNW einen festgelegten Wert erreicht. Dieser Wert wurde willkürlich mit $10 \cdot (\text{Anzahl der TNW}) = 60$ festgelegt. Anschließend kann eine Maßzahl der Ähnlichkeit des (normierten) geschätzten und (normierten) wahren Modells berechnet werden (z.B. SSE; Sum of Square Error).

Das Prinzip der Meßgütebestimmung durch die Strukturwiedererkennung wurde von uns, neben Computersimulationen, erstmalig auch in einer empirischen Untersuchung eingesetzt. Die Beschreibung der genauen Vorgehensweise ist im folgenden Abschnitt 6 enthalten.

5.6 Fazit

Die Meßgüte der Verfahren zur Nutzenschätzung kann nach vielfältigen Kriterien überprüft werden. Im allgemeinen belegen empirische Untersuchungen zur CA eine hohe Meßgüte im Sinne der verwendeten Kriterien. Möchte man die Meßgüte eines conjointanalytischen Verfahrens überprüfen, dann sollte man bedenken, daß jeder Güteindikator nur einen bestimmten Teilaspekt der Meßgüte abbildet. In jedem Fall ist man deshalb gut beraten, wenn mehrere Meßgütekriterien erfaßt werden.

Eines der größten Probleme der bisherigen Forschung zur Meßgüte der CA dürfte nach unserer Auffassung im Reliabilitätsproblem der Kriteriumsaufgabe (Abschnitt 5.4.1) begründet liegen. Da die meisten Forschungsarbeiten zur CA die kriterienorientierte Validität der (des) eingesetzten Verfahren(s) bestimmen, und jene die Beantwortung der gestellten Forschungsfrage oft besonders beeinflußt, kommt der Beachtung des Reliabilitätsproblems der Kriteriumsaufgabe eine besondere Bedeutung zu.

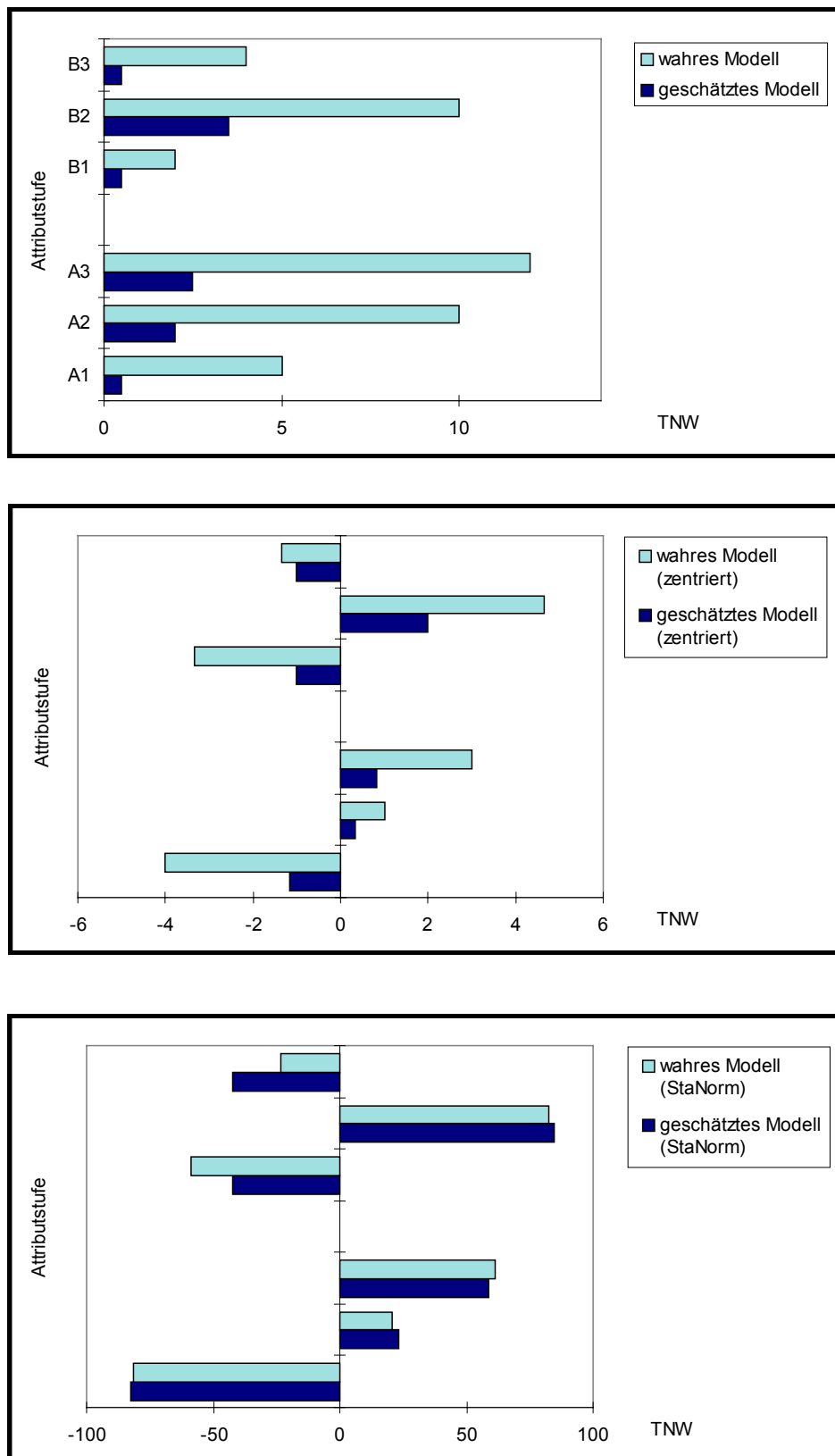


Abbildung 5.3: Zentrierung und Standardabweichungsnormierung eines wahren und eines geschätzten TNW-PSM: $\text{StdDev}(\text{TNW})=10 \cdot (\text{Anzahl der TNW})=60$. Anschließend kann eine Maßzahl der Strukturwiedererkennungslleistung (z.B. SSE) berechnet werden.

6 Empirischer Teil:

Optimierung der adaptiven Conjoint Analyse

Zieht man aus den Verfahrensmöglichkeiten und Erkenntnissen des Abschnitts 3 'Prozeß der Conjoint Analyse' Bilanz, dann verfügt ein derzeit optimales Verfahren der CA über folgende Charakteristika:

- Das CA-Interview wird computergestützt durchgeführt.
- Vor der dekompositorischen Nutzenschätzung erfolgt eine kompositorische Initialschätzung.
- Das Verfahren kann mit einer maximalen Attributanzahl von 5-10 umgehen. Es verwendet daher einen Teilprofilansatz.
- Als Antwortmodus wird ein Konstant-Summen-Paarvergleich oder ein Dollar-Metrik-Paarvergleich eingesetzt. Als Alternative hierzu kann auch das Rating oder der abgestufte Paarvergleich verwendet werden, wenn Vorkehrungen gegen Ankereffekte getroffen werden. Dies ist z.B. beim ACA-Paarvergleichsdifferenzendesign der Fall, indem den Apn adaptiv und individuell nutzenähnliche Paare vorgegeben werden.

Begutachtet man diese Charakteristika, dann stellt man fest, daß ACA 4.0 dem 'optimalen Verfahren' sehr nahe kommt. Insbesondere die ACA-Phase III (Paarvergleiche von adaptiv vorgegebenen Teilprofilobjekten) scheint die Methode der Wahl bei der Feinschätzung von multiattributiven Präferenzstrukturmodellen zu sein.

Wir wollen nicht zuletzt auch aufgrund der hohen Verbreitung (Wittink et al., 1994) hier die Phase III von ACA 4.0 als status quo des Forschungsstandes ansehen und versuchen, die CA von diesem Referenzpunkt aus weiter zu optimieren. Eine Forschungsfrage, die sich im Zusammenhang mit ACA stellt, soll hier aufgegriffen und bearbeitet werden:

Welche Objektprofilstärke ist bei Verwendung eines adaptiven Differenzendesigns und beim Antwortmodus 'abgestufter Paarvergleich' als optimal zu betrachten?

Die bisherigen Befunde zu dieser Forschungsfrage sind in Abschnitt 3.5.3.3 dargestellt. Wie dort deutlich wird, steht die Frage nach der optimalen Profil-

stärke in einem gewissen Zusammenhang mit der Frage nach der optimalen Anzahl der Paarvergleiche, die man seinen Apn vorgeben sollte. Die bisherigen Befunde zur optimalen Anzahl der Paarvergleiche wurden in den Abschnitten 3.5.3.3 und 3.7.2.4 diskutiert.

6.1 Forschungssoftware ALASCA

Das Softwaresystem ACA 4.0 eignet sich nur bedingt zu Forschungszwecken. Ein Grund hierfür liegt in dem nur sehr vage beschriebenen Algorithmus des Programms zur Bildung des adaptiven Differenzendesigns (vgl. Abschnitt 3.5.3.3). Wesentlich gewichtiger ist jedoch die Tatsache, daß der Ablauf des Interviews bei ACA 4.0 nur in sehr engen Grenzen vom Untersucher beeinflussbar ist. Beispielsweise kann die kompositorische Modellbildung (Phase I+II) nicht vollständig weggelassen werden. Ein anderes Beispiel wäre, daß die Obergrenze der Anzahl der Paarvergleiche (Gleichung 3.18) nicht vom Untersucher beeinflussbar ist.

Aus diesem Grunde war die Entwicklung von spezieller Software notwendig, mit der computergestützte CA-Interviews durchführbar sind (Jütting & Holling, 1996). Dieses Programm soll hier den Arbeitsnamen ALASCA tragen. ALASCA erlaubt, wie auch ACA 4.0, die Durchführung eines abgestuften Paarvergleichs. Das Paarvergleichsdifferenzendesign wird zur Vermeidung von Ankereffekten ebenfalls adaptiv vorgegeben. Die hierzu notwendige laufende Parameterschätzung des TNW-Modells erfolgt durch OLS-Regression analog zu dem Verfahren, das in Abschnitt 3.8.2 dargestellt ist. Abbildung 6.1 zeigt die Bildschirmoberfläche von ALASCA beim abgestuften Paarvergleich.

Wie weiter unten durch Computersimulationsstudien belegt wird, ist das Verhalten des adaptiven Designalgorithmus von ACA 4.0 und ALASCA sehr ähnlich. Das Paarvergleichsdifferenzendesign wird bei ALASCA nach folgenden Vorschriften gebildet:

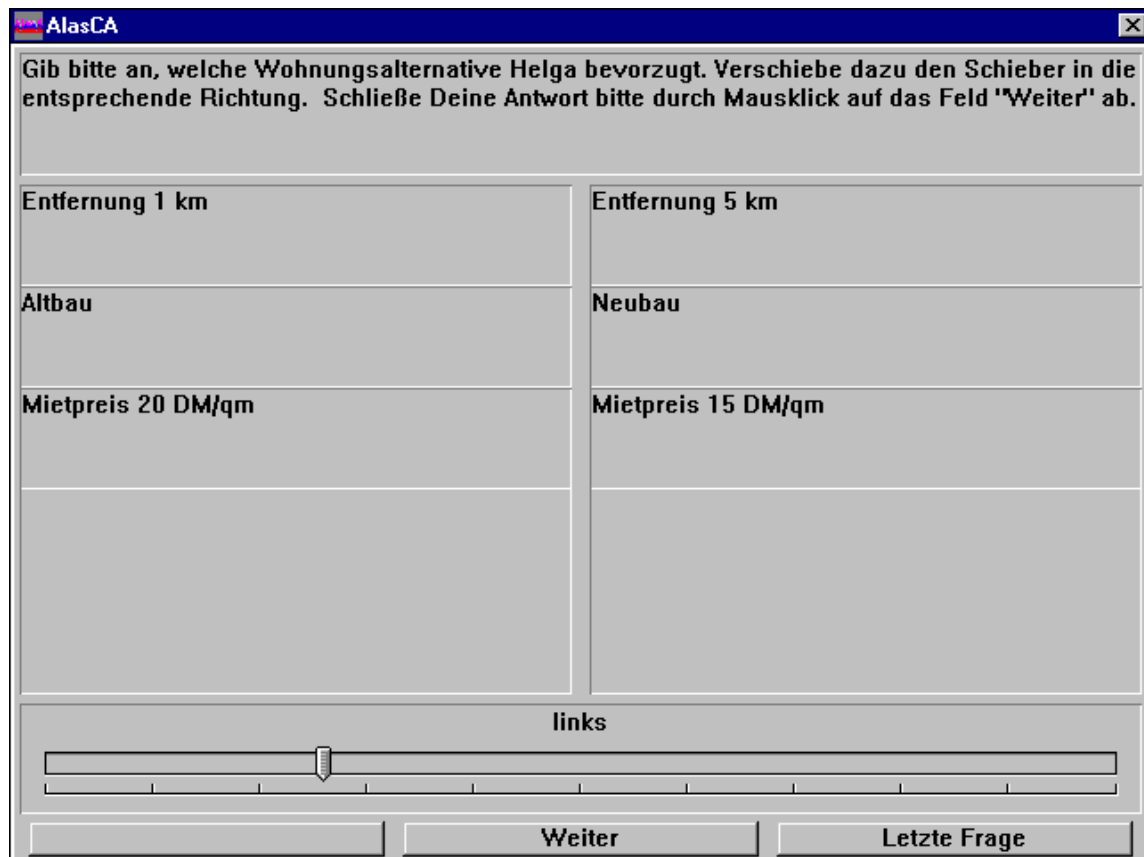


Abbildung 6.1: Die Bildschirmoberfläche des Programms ALASCA beim abgestuften Paarvergleich (Profilstärke 3). Die Ap gibt die Antwort, indem sie mit der Maus den Schieber entsprechend ihrer Präferenz bewegt. Die jeweilige Schieberstellung wird automatisch verbal verankert.

1. Zunächst werden die Attribute eines Paarvergleichs ausgewählt. ALASCA wählt diese jeweils so, daß a) die Attribute möglichst gleich häufig verwendet werden, und daß b) die Attribute innerhalb der Beurteilungsobjekte möglichst gleich häufig miteinander kombiniert werden. Kann durch diese Kriterien nicht zwischen mehreren Attributen entschieden werden, wird das Attribut per Zufall ausgewählt.
2. Im nächsten Schritt werden die Attributstufen bestimmt. Es werden die Attributstufen ausgewählt, die bisher am seltensten verwendet wurden. Bei Gleichheit entscheidet der Zufall.
3. Die gewählten Attributstufen werden innerhalb des Paarvergleichs so angeordnet, daß die bis dahin geschätzte Nutzendifferenz der Beurteilungsobjekte minimal ist.

Das ALASCA-Paarvergleichsdesign folgt damit dem Grundprinzipien des ACA-Designs. Bei beiden Designs gehen die Attribute und die Attributstufen innerhalb der Attribute möglichst gleich häufig in die Paarvergleichsaufgaben ein. Die Attributstufen innerhalb der Paare werden dabei so angeordnet, daß die Beurteilungsobjekte möglichst nutzenähnlich sind (vgl. hierzu die Ausführungen zum abgestuften Paarvergleich in den Abschnitten 3.6.2.4).

6.1.1 Computersimulationen: Vergleich von ALASCA und ACA 4.0

Bevor die von uns entwickelte Software ALASCA zu Forschungszwecken eingesetzt werden kann, sollte sie den Nachweis erbringen, daß sie mit ACA 4.0 vergleichbar ist. Gelingt der Nachweis, dann sollten die Untersuchungsergebnisse, die mit ALASCA gewonnen werden, auch auf ACA 4.0 übertragbar sein.

Die Darstellung der Beurteilungsobjekte auf dem Bildschirm, der Antwortmodus 'abgestufter Paarvergleich' und das Parameterschätzverfahren sind in dieser Hinsicht unproblematisch, da sie dem Vorgehen von ACA 4.0 folgen. An dieser Stelle soll dagegen die Behauptung überprüft werden, daß das Verhalten des adaptiven Paarvergleichsdesigns von ALASCA dem des ACA-Designs vergleichbar ist. Dies ist deswegen schwierig zu beurteilen, weil der genaue ACA-Designalgorithmus nicht veröffentlicht ist und ein bloßes Nachprogrammieren somit nicht möglich ist.

Auf der Grundlage der veröffentlichten Designbildungsprinzipien und der Erfahrung mit ACA konstruierten wir das ALASCA-Design. Die Prinzipien der Designkonstruktion sind in Abschnitt 6.1 beschrieben.

Um die Ähnlichkeit der geschätzten PSM zwischen ACA 4.0 und ALASCA zu belegen, wurden eine Reihe von Computersimulationen durchgeführt. Ihr Prinzip besteht darin, daß ein Computerantworter die Antworten im CA-Interview simuliert. Die Antworten werden aufgrund einer 'wahren Nutzenstruktur' und verschiedenen Antwortfehlerstufen berechnet. Bei dieser Art der Simulation interessiert, inwieweit die CA (in diesem Fall ACA 4.0 bzw. ALASCA) in der Lage ist, die wahren Präferenzstrukturen wiederzuerkennen (vgl. Abschnitt 5.5). Als Indikator der Strukturwiedererkennungslleistung verwendeten wir die Maßzahl SSE (z.B. Judd & McClelland, 1989) nach Normierung (StaNorm) der

PSM. Die Vorgehensweise ist ausführlich in Abschnitt 5.5.1 und Anhang C beschrieben.

Die Methode und die Ergebnisse der Computersimulationsstudien sind ausführlich in Anhang C dargestellt. Zusammenfassend kann festgehalten werden, daß das adaptive Paarvergleichsdesign von ALASCA und ACA 4.0 zu nahezu identischen Strukturwiedererkennungsleistungen führt. Lediglich bei speziellen wahren Nutzenstrukturen ergeben sich bei der ersten Paarvergleichsfrage geringfügige Unterschiede. Diese egalisieren sich jedoch nach wenigen Paarvergleichen. Wir ziehen aus den Computersimulationen den Schluß, daß sich die folgenden Befunde zur gestellten Forschungsfrage, die mit der Forschungssoftware ALASCA gewonnen wurden, uneingeschränkt auf ACA 4.0 übertragen lassen sollten.

6.2 Welche Effekte sind theoretisch zu erwarten?

Unsere Grundüberlegung bei der Modellierung der Präferenzstruktur einer Ap durch Paarvergleiche der beschriebenen Art ist wie folgt:

Hypothese 1: Die *Güte des geschätzten PSM* hängt von den Faktoren *Anzahl der Paarvergleiche*, *Profilstärke* und *Antwortfehler* ab. Die Zusammenhänge sind monoton und werden in ihrer Richtung durch die Abbildungen 6.2, 6.3 und 6.4 spezifiziert.

Aus statistischen Gründen ist davon auszugehen, daß die Güte der PSM mit der Profilstärke und mit der Anzahl der Paarvergleiche zunimmt (vgl. die Ausführungen in Abschnitt 3.5.3.3). Bei der Anzahl der Paarvergleiche und der Profilstärke handelt es sich um objektive Eigenschaften der ALASCA-Aufgabe. Der 'Antwortfehler' ist dagegen ein theoretisches Konzept, das uns zur Erklärung und Vorhersage der empirischen Befunde dienlich sein soll. Er beschreibt die Genauigkeit, mit der eine Ap gemäß ihrer wahren Präferenzstruktur die Paarvergleiche bearbeitet. Mit zunehmendem Antwortfehler sollte sich eine abnehmende Meßgüte des geschätzten PSM zeigen (Abbildung 6.4).

Abbildung 6.2: Der Zusammenhang zwischen der Anzahl der Paarvergleiche und der Meßgüte.

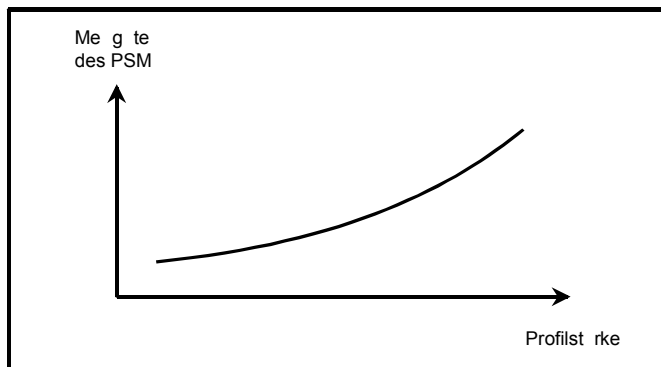


Abbildung 6.3: Der Zusammenhang zwischen der Profilstärke und der Meßgüte.

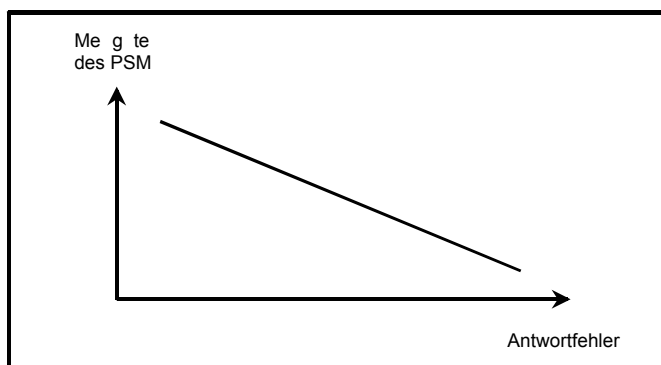


Abbildung 6.4: Der Zusammenhang zwischen dem Antwortfehler und der Meßgüte.

Wovon hängt der Antwortfehler einer Ap ab? Allgemein ausgedrückt hängt er von den Charakteristika der Auskunftssituation, der Auskunftsfähigkeit und der Auskunftsmotivation einer Person ab. Die Charakteristika der Auskunftssituation umfassen beispielsweise die Schwierigkeit der gestellten Aufgabe, die Art, wie sehr die gestellte Aufgabe zur Beantwortung motiviert oder die subjektive Wahrscheinlichkeit, mit der die Ap glaubt, die richtige Antwort geben zu können.

Die Auskunftsfähigkeit und -motivation ist z.B. durch die Auswahl der speziellen Apn für die Untersuchung oder durch die Untersuchungsrahmenbedingungen bestimmt (vgl. Abschnitt 3.1). Sie wird jedoch, und das ist hier von Interesse, auch von der Art der Beurteilungsaufgaben des CA-

Interviews beeinflusst. So kann man beispielsweise annehmen, daß die Auskunftsmotivation mit zunehmender Interviewdauer bzw. mit der Anzahl zuvor beantworteter Paarvergleiche abnimmt. Dies sollte zu einer Zunahme des Antwortfehlers führen (Abbildung 6.5).

Des Weiteren sollte die Auskunftsfähigkeit der Apn mit zunehmender Profilstärke deutlich geringer werden. Diese Annahme ist dadurch zu rechtfertigen, daß mit zunehmender Profilstärke mehr kognitive Ressourcen zur Bearbeitung der Urteilsaufgabe notwendig werden, weil die Aufgabenschwierigkeit zunimmt. Theoretisch erklärbar ist diese Erwartung durch die Beschränktheit der menschlichen

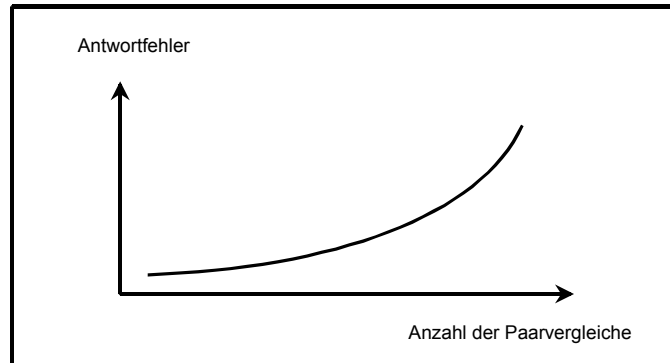


Abbildung 6.5: Der Zusammenhang zwischen der Anzahl der Paarvergleiche und dem Antwortfehler.

Informationsverarbeitungskapazität im Arbeitsgedächtnis. Da bei kompensatorischen Urteilsprozessen zwischen zwei Alternativen gleichzeitig mehrere Attributstufen im Arbeitsgedächtnis aktiviert sein müssen, dürften spätestens ab der Profilstärke 3 Kapazitätsprobleme auftreten (Brown & Kirsner, 1980). Je stärker die Objektprofile sind, desto niedriger sind damit die zur Verfügung stehenden kognitiven Ressourcen in Relation zur Aufgabenschwierigkeit. Dies führt zu einem hohen Antwortfehler. Den Zusammenhang zwischen Antwortfehler und Profilstärke zeigt Abbildung 6.6 schematisch.

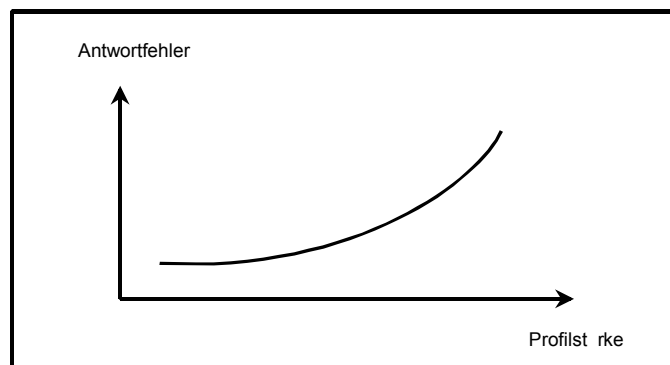


Abbildung 6.6: Der Zusammenhang zwischen der Objektprofilstärke und dem Antwortfehler.

Hypothese 2: Der *Antwortfehler* hängt von der *Auskunftsfähigkeit* und der *Auskunftsmotivation* der Apn ab. Die Auskunftsfähigkeit nimmt mit zunehmender Profilstärke ab. Die Auskunftsmotivation sinkt mit der Anzahl bereits bearbeiteter Paarvergleiche. Die Zusammenhänge sind monoton.

Der Antwortfehler dient in Hypothese 1, neben der Objektprofilstärke und der Anzahl der Paarvergleiche, als erklärendes Konzept für die Güte des geschätzten PSM. Aus den Ausführungen zu Hypothese 2 und den Abbildungen 6.5 und 6.6 wird jedoch deutlich, daß der Antwortfehler auch von der Objektprofilstärke und der Anzahl der Paarvergleiche abhängig ist. Die Profilstärke und die Anzahl der Paarvergleiche wirken sich gemäß diesem Denkmodell einerseits direkt und positiv auf die Meßgüte des PSM aus (statistische Perspektive; Hypothese 1). Andererseits bewirken sie über die Beeinflussung des Antwortfehlers einen negativen Einfluß auf die Meßgüte des PSM (psychologische Perspektive; Hypothese 2; vgl. auch Abschnitt 3.5.3.3). Welcher der Einflüsse überwiegt, kann nur durch eine empirische Untersuchung beantwortet werden.

Die Objektprofilstärke sollte sich neben der Meßgüte des PSM auch auf die Antwortzeiten auswirken. Stärkere Objektprofile führen zu einer längeren Antwortzeit, weil mehr gelesen werden muß und die Beurteilungsaufgabe komplexer ist. Es folgt aus dem Gesagten:

Hypothese 3: Mit zunehmender Profilstärke (und damit zunehmender Auslastung kognitiver Ressourcen/ abnehmender Auskunftsfähigkeit) vergrößert sich der Antwortfehler. Die Meßgüte des PSM wird dadurch negativ beeinflusst. Darüber hinaus verlängert sich die Antwortzeit.

Die Überprüfung der Antwortzeit-Vorhersagen soll, neben der Überprüfung der Meßgüte-Vorhersagen und Antwortfehler-Vorhersagen, einer weiteren Stützung unserer Denkweise dienen. Aus den Hypothesen 1 bis 3 werden bei der Darstellung der empirischen Untersuchung II (Abschnitt 6.4) Teilhypothesen abgeleitet und nachfolgend überprüft.

6.3 Begründung und Darstellung des Forschungsdesigns

Unsere Bemühungen dienen zwei Zielen: Erstens soll durch diese Arbeit die gestellte Forschungsfrage beantwortet werden. Darüber hinaus besteht das zweite Ziel unserer Anstrengungen darin, einige methodische Defizite bisheriger Forschung zur CA aufzuzeigen und eine Möglichkeit zu deren Beseitigung in Form eines besonderen Forschungsdesigns vorzuschlagen. Unsere Darstellung zur Lösung der gewählten Forschungsfragen kann daher auch als beispielhafter Vorschlag angesehen werden, wie ein willkürlich gewähltes Forschungsproblem im Feld der CA effektiv bearbeitet werden kann.

6.3.1 Anforderungen durch das Problem der Kriterienreliabilität

Nahezu alle empirischen Forschungsarbeiten zur CA gehen nach folgendem Muster vor: Zwei oder mehr Variationen eines (CA-)Nutzenschätzverfahrens werden eingesetzt, um die gestellte Forschungsfrage zu beantworten. Fragt man sich beispielsweise, ob Präferenzen effektiver durch ein kompositorisches Nutzenschätzverfahren oder eine Vollprofil-CA modellierbar sind, gibt man seinen Apn beide Verfahren vor. Für jede Ap wird ein PSM (between subject design) oder mehrere PSM (within subject design) geschätzt. Zusätzlich zu den Objekten, auf deren Beurteilung die Parameterschätzwerte des PSM beruhen, werden den Apn Holdout-Objekte vorgegeben. Die kriterienorientierte Validität (Abschnitt 5.4) der PSM hinsichtlich der Beurteilung der Holdout-Objekte wird herangezogen, um zu beurteilen, welches Nutzenschätzverfahren effektiver arbeitet. Ist diese beim Verfahren A höher als beim Verfahren B, dann führt man dies auf den Unterschied zwischen den Verfahren zurück. Würde im obigen Beispiel die Vollprofil-CA eine höhere Validität erbringen, dann würde dies z.B. als Überlegenheit dekompositorischer Nutzenmessungen interpretiert werden.

Die Logik dieser Denkweise ist grundsätzlich überzeugend. Trotzdem haftet diesem Vorgehen ein großes Problem an: Es ist das Problem der Kriterienreliabilität, das in Abschnitt 5.4.1 beschrieben wurde. Insbesondere bei Versuchsplänen, bei welchen die CA-Verfahren zwischen den Apn variiert werden (between subject design), kann dies dazu führen, daß die ermittelten Validitätsmaßzahlen nicht nur von der kriterienorientierten Validität der PSM,

sondern hauptsächlich von der individuellen Schwierigkeit und der daraus folgenden Reliabilität der Kriteriumsaufgabe beeinflusst werden.

In experimentellen Untersuchungen mit randomisierter Zuordnung der Apn zu den Versuchsgruppen sollten die durchschnittlichen Kriterienreliabilitäten der Versuchsgruppen vergleichbar sein. Dann hat das Kriterienreliabilitätsproblem eine Steigerung der Fehlervarianz zur Folge (in varianzanalytischer Sprache ausgedrückt: Es führt zu einer Steigerung der Varianz innerhalb der Gruppen). Vorhandene Effekte würden dann leichter übersehen (Fehler 2. Art). Schwerwiegender ist, wenn die Kriterienreliabilitäten der Versuchsgruppen nicht vergleichbar sind, was bei den häufig kleinen Stichprobengrößen von Forschungsarbeiten zur CA durchaus vorkommen kann. In diesem Falle kann die Interpretation, daß die Nutzenschätzverfahren PSM unterschiedlicher Validität bilden, ein Fehlschluß sein.

Hypothese 4: Die kriterienorientierte Validität bezüglich der Beurteilungen von Holdout-Konzepten wird von der Schwierigkeit der Holdout-Aufgabe maßgeblich beeinflusst. Verwendet man einen solchen Indikator der kriterienorientierten Validität als abhängige Variable, dann führt das dazu, daß vorhandene Effekte übersehen werden können (Problem der Kriterienreliabilität).

6.3.2 Empirische Untersuchung I

Um Hypothese 4 zu überprüfen und die Effektstärke der mangelnden Kriterienreliabilität auf die kriterienorientierte Validität abschätzen zu können, führten wir eine empirische Untersuchung durch.

6.3.2.1 Methode

Wir untersuchten mit ACA 4.0 die Präferenzen von Studierenden der Psychologie (n=66) gegenüber fiktiven Prüfungsordnungen im Diplomstudiengang Psychologie. Wir entschieden uns bei dieser Untersuchung für fünf Attribute. Die Attribute und die Attributstufen können dem Beispiel zur konjunktivkompositorischen Modellbildung in Anhang B entnommen werden. Um die Validität der geschätzten PSM auf individueller Ebene abzuschätzen, erhielten

die Apn nach der Erledigung des ACA-Interviews die Aufgabe, acht Vollprofil-Holdout-Objekte gemäß ihrer Präferenz zu ordnen. Bei diesen Objekten handelte es sich um jeweils eine fiktive Prüfungsordnung, die auf allen fünf Attributen spezifiziert war. Die Auskunfts motivation der Apn war hoch, weil zum Untersuchungszeitpunkt die Prüfungsordnung geändert wurde und die Apn einen Einfluß ihrer Meinung auf die Änderungen erwarteten.

6.3.2.2 Ergebnisse

Für jede Ap berechneten wir die Rangkorrelation (r_s) zwischen der Rangreihe der Holdout-Objekte und der Rangreihe, die das PSM vorhersagen würde. Ein Histogramm dieser Maßzahl der kriterienorientierten Validität ist in Abbildung 6.7 dargestellt. Es zeigt sich, daß die Rangkorrelationen hoch zwischen den Apn variieren.

Eine übliche Interpretation dieser Variabilität wäre, daß die PSM eine unterschiedliche Validität besitzen, was beispielsweise auf eine unterschiedliche Antwort motivation der Apn oder sonstige Unterschiede der Interviewsituation zurückzuführen wäre.

Aus der Sicht des Problems der Kriterienreliabilität würde man jedoch behaupten, daß mindestens ein Teil der Variabilität der kriterienorientierten Validität durch Unterschiede in der Schwierigkeit der Kriteriumsaufgabe erklärt werden kann. Um ein Maß für diese Schwierigkeit zu entwickeln, normierten wir zunächst die PSM (Attributwichtigkeitsnormierung [Summe der Attributwichtigkeiten]=1; vgl. Abschnitt 4.1). Anschließend berechneten wir den Ap-spezifischen Gesamtnutzen eines jeden Holdout-Objekts und bestimmten ein Maß für die Schwierigkeit der Kriteriumsaufgabe (SK; Schwierigkeit der Kriteriumsaufgabe). Hierzu berechneten wir auf der Grundlage der in Abschnitt 5.4.1 dargestellten und in Abbildung 5.2 visualisierten theoretischen Überlegungen die Summe der Nutzendifferenzen zwischen allen Holdout-Objekt-Paaren:

$$SK = |u(X_i) - u(X_{i^*})| \quad (6.1)$$

Je höher der Wert dieses Schwierigkeitsindikators SK, desto einfacher ist die Kriteriumsaufgabe für eine Ap, da die Beurteilungsobjekte eine höhere durchschnittliche Nutzendifferenz besitzen.

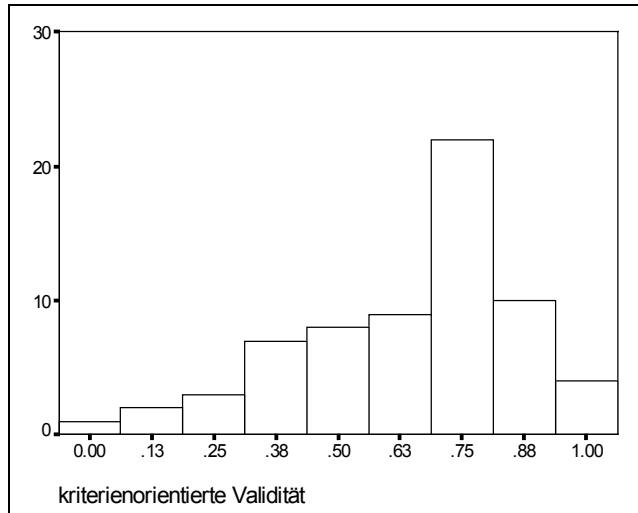


Abbildung 6.7: Histogramm der kriterienorientierten Validität gegenüber einem Holdout-Set. Dargestellt sind Rangkorrelationen ($n=66$).

Die Argumente, die zum Kriterienreliabilitätsproblem geführt haben, begründen die Hypothese, daß die kriterienorientierte Validität des PSM einer Ap um so niedriger sein sollte, je schwieriger die Kriteriumsaufgabe für die Ap aufgrund ihrer besonderen Präferenzstruktur ist. Die Maßzahl der kriterienorientierten Validität und die Maßzahl der Kriteriumsschwierigkeit korreliert mit $r_{r_s-SK}=0,54$ ($p<0,01$ mit

dem Fisher-Z transformierten Rangkorrelationskoeffizienten als Indikator der Validität). Hypothese 4 wird durch diesen Befund bestätigt. Er führt sehr deutlich die großen Effekte vor Augen, die durch die unterschiedliche Kriterienreliabilität entstehen.

6.3.2.3 Diskussion

Wie kann dem Problem der Kriterienreliabilität begegnet werden? Nach unserem Dafürhalten mindern folgende Ansätze die Gefahr eines Forschungsartefaktes durch das Kriterienreliabilitätsproblem:

1. *Verwendung multipler Meßgütekriterien und mehrerer Maßzahlen.* Dieser Vorgehensweise liegt die Annahme zugrunde, daß sich die mangelnde Reliabilität des Kriteriums verschieden auf die Meßgütekriterien und deren Maßzahlen auswirken sollte. Stimmen die Antworten verschiedener Meß-

gütekriterien und deren Maßzahlen auf die gestellte Forschungsfrage überein, dann sollte dies ein Hinweis auf die Richtigkeit dieser Antworten sein. Multiple Meßgütekriterien/ Maßzahlen werden häufig (aber vermutlich meist recht intuitiv) bei empirischen Untersuchungen zur CA eingesetzt.

2. *Verwendung von Meßwiederholungsuntersuchungsdesigns ('within subject designs')*. Die unterschiedlichen Kriterienreliabilitäten bestehen zwischen den Apn zwar weiterhin. Da die Wirkung der experimentellen Variation der CA-Verfahren aber jeweils innerhalb einer Ap bestimmt wird, bleibt bei diesem Vergleich die Kriterienreliabilität konstant.
3. *Bestimmung der Reliabilität der Kriteriumsmessung*. Diese Vorgehensweise macht zumindest die Auswirkungen des Problems deutlich. Inwieweit die Annahmen einer folgenden Minderungskorrektur (vgl. Abschnitt 5.4.1) gerechtfertigt sind, bleibt problematisch.
4. *Verwendung der Strukturwiedererkennung als Meßgüteindikator*. Bei Computersimulationen ist die wahre Nutzenstruktur dem Untersucher bekannt. Ein CA-Verfahren wird danach beurteilt, wie gut es in der Lage ist, diese wiederzuentdecken (vgl. Abschnitt 5.5). Wenn es möglich wäre, auch bei empirischen Untersuchungen die wahre Nutzenstruktur zu kennen, dann hätte dies zwei große Vorteile im Umgang mit dem Kriterienreliabilitätsproblem. Erstens kann die wahre Struktur direkt als fehlerfreies Kriterium angesehen werden. Man kann ein PSM danach beurteilen, inwieweit es die wahre Nutzenstruktur abbildet. Zweitens können bei Bekanntheit der wahren Nutzenstruktur Holdout-Objektsets individuell so gebildet werden, daß deren Schwierigkeit über die Apn oder innerhalb einer Ap über mehrere Holdout-Sets vergleichbar ist. Am einfachsten geht dies natürlich, wenn es möglich wäre, in alle Apn die gleiche wahre Nutzenstruktur zu induzieren. In diesem Fall sind objektiv gleiche Holdout-Beurteilungsaufgaben auch subjektiv für jede Ap gleich schwierig. Auf diese Weise wird zwar nicht die Reliabilität der Kriteriumsaufgabe direkt kontrolliert. Eine vergleichbare Schwierigkeit der Kriteriumsaufgabe sollte jedoch indirekt auch zu einer vergleichbaren Reliabilität der Kriteriumsmessung beitragen. Die Beurteilungsschwierigkeit mehrerer Holdout-Objektsets könnte z.B. dadurch ähnlich gemacht werden, indem man die Nutzendifferenzen der Objekte konstant hält.

Punkt 4 sehen wir als die entscheidende Strategie an, den Nachteilen der herkömmlichen Vorgehensweise bei empirischen Untersuchungen zur CA - Vergleich mehrerer Nutzenschätzverfahren anhand der kriterienorientierten Validität zu einer Holdout-Aufgabe - zu begegnen. Woher aber ist bei einer empirischen Untersuchung die wahre Nutzenstruktur bekannt, so daß mit dieser gearbeitet werden kann?

6.3.3 Strukturwiedererkennung bei empirischen Untersuchungen

Wie in Abschnitt 5.5 dargestellt ist, wird das Prinzip der 'Wiedererkennung einer wahren Nutzenstruktur' bei Computersimulationen eingesetzt. Ist es auch auf empirische Untersuchungen übertragbar?

Im Gegensatz zur Simulationsstudie sind bei einer empirischen Untersuchung wahre Nutzenstrukturen der Apn nicht ohne weiteres bekannt. Wie kann man trotzdem valide Annahmen über die wahre Nutzenstruktur einer Ap treffen? Grundsätzlich sind zwei Strategien denkbar. Die erste Möglichkeit besteht darin, daß man mit einem *hohen Meißaufwand die Nutzenstruktur einer Person ermittelt*. Diese dient dann als bestmögliche Schätzung ihrer wahren Präferenzstruktur. Die Ergebnisse eines anderen Nutzenschätzverfahrens können dann mit dem wahren Modell verglichen werden. Ist das PSM des zweiten Verfahrens dem wahren Modell sehr ähnlich, so spricht dies für die Validität dieses PSM. Mit anderen Worten: Das zweite Verfahren liefert trotz niedrigerem Erhebungsaufwand ähnliche Erkenntnisse über die Nutzenstruktur einer Person wie das mit viel Aufwand geschätzte wahre Modell.

Die zweite Möglichkeit besteht darin, *in die Apn eine vorgegebene Nutzenstruktur zu induzieren*. Dies erfolgt z.B. in einer Lernphase, in der die Apn darauf trainiert werden, wie eine fiktive Modellperson zu urteilen und zu entscheiden. Das Urteilsverhalten der Modellperson kann vom Untersuchungsleiter willkürlich manipuliert werden. Eine einfache Vorgehensweise besteht darin, daß ein Computer auf der Basis eines vorgegebenen PSM dieses Verhalten simuliert. In diesem Fall ist die 'wahre Nutzenstruktur' des zu lernenden Verhaltens bekannt. Nach erfolgreicher Induktion dieses Urteilsverhaltens in eine Ap kann man dann ein Nutzenschätzverfahren danach beurteilen, inwiefern das geschätzte PSM das wahre PSM abbildet.

Beurteilt man ein Nutzenschätzverfahren nach seiner Fähigkeit, wahre Nutzenstrukturen wiederzuerkennen, dann kann dies eine wertvolle Erweiterung der üblichen Meßgütekriterien darstellen. Der Hauptnachteil des Verfahrens bei empirischen Untersuchungen dürfte, sowohl bei der Bildung des wahren PSM durch ausführliche Messung als auch bei der Induktion eines wahren PSM, in der relativ hohen Gefahr der Reaktivität liegen. Insbesondere bei der Induktionsmethode sollte man sich gewahr sein, daß nicht nur die Parameter eines wahren PSM, sondern auch die algebraische Form des Modells induziert werden. Trainiert man Apn z.B. darauf, wie ein bestimmtes additives TNW-Modell zu urteilen, dann sollte man sich nicht wundern, wenn ein additives Modell dem späteren 'als-ob-Verhalten' der Apn angemessen ist. Der Umkehrschluß, daß menschliche Urteile und Entscheidungen durch additive Präferenzintegrationsmodelle strukturangemessen (Abschnitt 3.4.2.3) repräsentiert werden können, wäre jedoch falsch.

6.3.4 Das Forschungsdesign

Unser Forschungsdesign erfüllt alle oben genannten Anforderungen. Es hat folgende Eigenschaften:

- Eine wahre Nutzenstruktur wird in die Apn induziert.
- Mehrere Meßgütekriterien stehen zur Verfügung.
- Es ist möglich, die Reliabilität der Kriterien, die bei der Bestimmung der kriterienorientierten Validität eingesetzt werden, abzuschätzen.
- Das experimentelle Treatment erfolgt im Rahmen eines Meßwiederholungsversuchsplans (within subject design).

Abbildung 6.8 zeigt den Aufbau des Forschungsdesigns schematisch. Um nach diesem Ansatz vorgehen zu können, ist es zunächst notwendig, eine wahre Nutzenstruktur festzulegen. Dies kann z.B. durch die Spezifikation eines additiven TNW-Modells erfolgen. Bei der empirischen Untersuchung versucht man in der Lernphase die wahre Nutzenstruktur in die Apn zu induzieren. Durch Lernerfolgskontrollen wird sichergestellt, daß diese Bemühungen erfolgreich waren. In den Abschnitten 6.4.1.4 und 6.4.1.5 ist ausführlich beschrieben, wie die Induktion und Lernerfolgskontrolle bei unserer Untersuchung umgesetzt wurde.

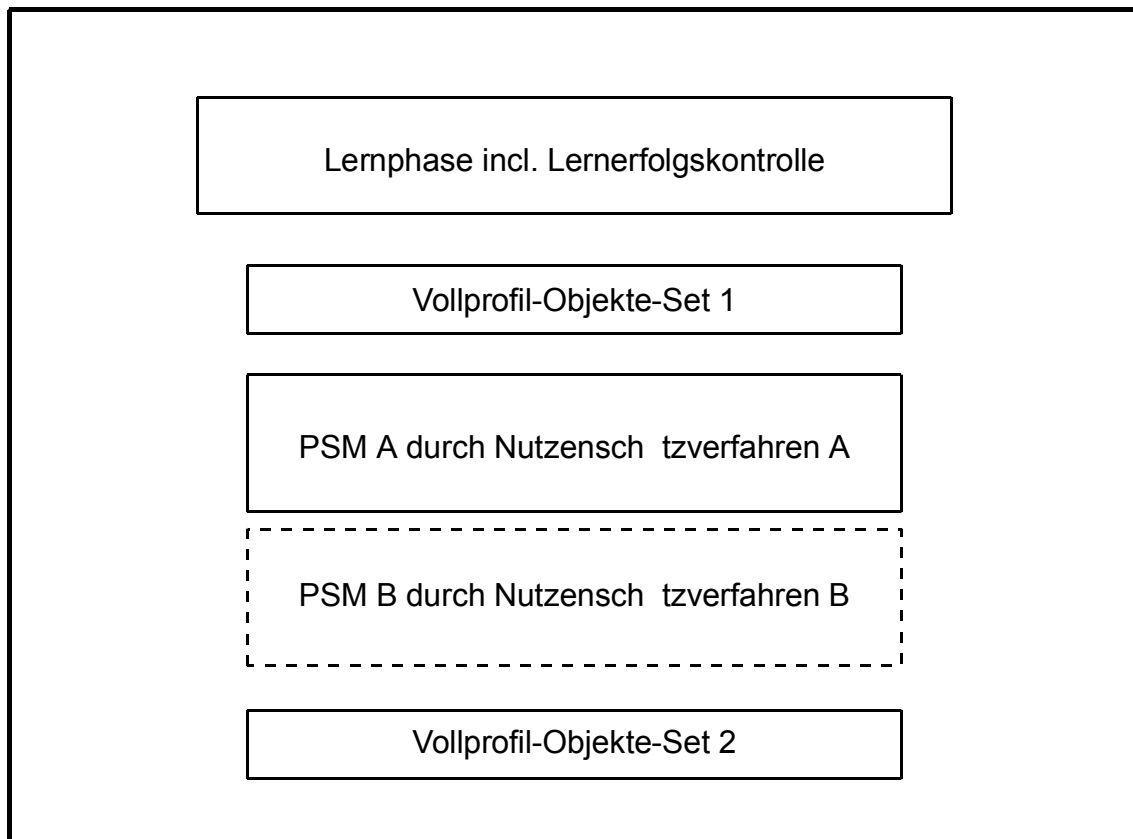


Abbildung 6.8: Schematische Darstellung des Forschungsdesigns. Der gestrichelt dargestellte Kasten ist nur bei einem 'within subject design' vorhanden.

Entsprechend dem Grundgedanken des Experiments schätzt man mindestens zwei PSM A und B. Die Meßgüte des jeweilig eingesetzten Nutzenschätzverfahrens kann man direkt durch die Strukturwiedererkennungslleistung der wahren Nutzenstruktur durch das PSM ermitteln (Abschnitt 5.5).

Prinzipiell ist es möglich, das Nutzenschätzverfahren bei der empirischen Untersuchung zwischen oder innerhalb der Apn zu variieren. Dem gestrichelt gezeichneten Kasten würde entsprechen, daß einer Ap sowohl Verfahren A als auch Verfahren B vorgegeben werden. Wir ziehen bei unserem Forschungsdesign aus den in Abschnitt 6.3.1 genannten Gründen die Variation innerhalb der Apn vor.

Um der Forderung nach mehreren Meßgütekriterien gerecht zu werden, und um die Vergleichbarkeit mit solchen Untersuchungsdesigns zu wahren, die in der Forschung zur CA üblicherweise angewandt werden, verwenden wir in unserem Forschungsdesign ebenfalls Vollprofil-Holdout-Objekte. Diese Objekte müssen von den Apn z.B. in eine Rangreihe gebracht oder auf einer Ratingskala eingeschätzt werden. Prinzipiell sind alle Maße der kriterienorientierten Validität, wie wir sie in Abschnitt 5.4 ausführlich besprochen haben, zur Beurteilung der Güte der PSM einsetzbar.

Unser Forschungsdesign beinhaltet zwei Vollprofil-Objekte-Sets. Dies dient zweierlei Zielen: Zum einen kann dadurch die Reliabilität der Kriteriumsaufgabe durch einen Test-Retest-Ansatz (s.u.) abgeschätzt werden. Der andere Grund liegt im verwendeten Meßwiederholungsdesign begründet. Eine einwandfreie Interpretation der Ergebnisse bei diesen Versuchsplänen ist nur dann gegeben, wenn sichergestellt werden kann, daß sich die induzierte wahre Nutzenstruktur der Apn nicht während der Untersuchung geändert hat. Unterscheiden sich die Präferenzen beim Vollprofil-Objekte-Set 1 sehr von denen des Vollprofil-Objekte-Sets 2, dann ist es problematisch, eventuelle Unterschiede zwischen PSM A und PSM B auf Unterschiede in den Nutzenschätzverfahren A und B zurückzuführen. Der Unterschied könnte auch durch eine veränderte Nutzenstruktur entstanden sein.

Die beiden Vollprofil-Objekte-Sets werden nach zwei Kriterien konstruiert. Erstens handelt es sich dabei um orthogonale Haupteffektdesigns (OHD; vgl. Abschnitt 3.5.3.1), welche eine nicht konfundierte Schätzung der TNW zulassen. In diesem Sinne kann die Holdout-Aufgabe auch als traditionelle Vollprofil-CA aufgefaßt werden. Die PSM A, B, ... der untersuchten Nutzenschätzverfahren können mit den PSM 1 und 2 verglichen werden, die durch eine traditionelle Vollprofil-CA gebildet wurden. Die Vollprofil-CA dient sozusagen als Referenzpunkt.

Zweitens sind einige, aber nicht alle Objekte der beiden OHD identisch. Dadurch können mehrere Indikatoren der Reliabilität der Holdout-Beurteilungen (Kriterienmessung) bestimmt werden (vgl. Abbildung 5.1). Durch den Vergleich der beiden PSM 1 und PSM 2 läßt sich die Reliabilität auf PSM-Ebene abschätzen. Berücksichtigt man nur diejenigen Objekte, die in beiden Holdout-Sets vorkommen, läßt sich ein Indikator der Reliabilität auf Antwortvektorebene

bestimmen. Die Vorgehensweise und die einsetzbaren Maßzahlen haben wir eingehend in Abschnitt 5.2 beschrieben.

Zusammenfassend läßt sich sagen, daß in diesem Forschungsdesign zwei wesentliche Neuerungen gegenüber bisherigen Forschungsansätzen enthalten sind:

1. Erstens wird mit *wahren Nutzenstrukturen* operiert, die die Apn vor der eigentlichen Untersuchung lernen. Die wichtigsten Vorteile dabei sind: a) Das Problems der Kriterienreliabilität wird bei der Bestimmung der kriterienorientierten Validität gemindert. b) Mit der Strukturwiedererkennung der wahren Nutzenstruktur erhält man ein weiteres Meßgütekriterium der geschätzten PSM.
2. Zweitens werden die zu untersuchenden Nutzenschätzverfahren von *zwei besonders konstruierten Vollprofil-Objekte-Sets* 'umrahmt'. Die wichtigsten Vorteile dabei sind: a) Die Reliabilität der Kriteriumsaufgabe kann durch verschiedene Methoden (auf PSM-Ebene und Antwortvektorebene) abgeschätzt werden. b) Ein direkter Vergleich der untersuchten Nutzenschätzverfahren mit dem Referenzverfahren der traditionellen Vollprofil-CA ist möglich. Die Vollprofil-Beurteilungsaufgabe wird also einmal als übliche Holdout-Aufgabe aufgefaßt, um die kriterienorientierte Validität der PSM A, B, ... zu bestimmen. Ein anderes Mal wird sie als traditionelle Vollprofil-CA angesehen, mit der ein eigenes PSM geschätzt werden kann.

Dieses Forschungsdesign hat auch einige Nachteile. Die Anzahl der Attribute, mit der sinnvoll gearbeitet werden kann, ist sowohl durch die Induktion als auch durch die Verwendung der Vollprofil-Objekte begrenzt. Darüber hinaus ist die Durchführung der empirischen Untersuchung sehr zeitaufwendig. Bei fünf Attributen muß damit gerechnet werden, daß allein die Lernphase ca. zwei bis vier Stunden dauert. Dazu kommt der Zeitbedarf für die eigentliche Untersuchung mit den beiden Vollprofil-Aufgaben und mindestens zwei weiteren Nutzenschätzverfahren. Mit den notwendigen Pausen ist damit die Zeitdauer eines ganzen Arbeitstages für die Untersuchung meist erreicht. Unser Design ist damit besonders zur Bearbeitung von grundlagenorientierten Forschungsfragenstellungen geeignet.

6.4 Empirische Untersuchung II: Meßgüte und Antwortfehler in Abhängigkeit der Anzahl der Paare und der Profilstärke

In Abschnitt 6.3.4 ist das Forschungsdesign dargestellt, das zur Bearbeitung der gestellten Forschungsfragen entwickelt wurde. Die empirische Untersuchung II geht nach dem Prinzip dieses Designs vor. In diesem Abschnitt werden die Methode und die Ergebnisse der Untersuchung dargestellt und diskutiert.

6.4.1 Methode

Dieser Abschnitt widmet sich der Beschreibung des genauen Aufbaus der empirischen Untersuchung. Er beschreibt z.B. die verwendete Objektklasse, die Objektattribute, die wahre Nutzenstruktur, die Vorgehensweise bei der Induktion der wahren Nutzenstruktur, die Stichprobe der Apn und den Untersuchungsplan. An passender Stelle werden hier spezifische Teilhypothesen aufgestellt. Diese leiten sich aus den Hypothesen 1 bis 3 ab und sind in Abschnitt 6.2 begründet. Diese Vorgehensweise wurde zur Erhöhung der Übersichtlichkeit gewählt.

6.4.1.1 Die wahre Nutzenstruktur

Bei der Konstruktion der wahren Nutzenstruktur ließen wir uns von dem Ziel leiten, den folgenden Bedingungen zu genügen:

- Das wahre PSM sollte Attribute mit linearem und nicht linearem Nutzenverlauf beinhalten.
- Die relative Wichtigkeit ihrer Attribute sollte verschieden sein und sich innerhalb der üblichen Grenzen bewegen, die man bei CA-Untersuchungen findet.

In Tabelle 6.1 sind verschiedene Transformationen der gewählten wahren Nutzenstruktur dargestellt. Darstellungsart A verdeutlicht die Erfüllung der o.g. Konstruktionsbedingungen. Die zentrierte und normierte Form des Modells (StaNorm; vgl. Abschnitt 5.5.1), die zur Berechnung des Maßes der Strukturwiedererkennung $W(SSE)$ verwendet wird, ist in B dargestellt (eine ausführliche Erklärung der Maßzahl $W(SSE)$ findet sich in Anhang C). In C dagegen wurde das Modell A mit dem Faktor 25 dividiert. Der Skalierungsfaktor wurde so

Tabelle 6.1: Die TNW des wahren PSM in verschiedenen Darstellungsformen (Erläuterungen im Text).

Attribut und Attributstufe wahres PSM	Attribut 1			Attribut 2			Attribut 3			Attribut 4			Attribut 5		
	11	12	13	21	22	23	31	32	33	41	42	43	51	52	53
A	100	50	0	90	45	0	80	8	0	70	49	0	50	25	0
B	224	0	-224	202	0	-202	227	-96	-132	136	42	-178	112	0	-112
C	4	2	0	3,6	1,8	0	3,2	0,32	0	2,8	1,96	0	2	1	0

gewählt, daß die maximale relative Attributwichtigkeit den Wert 4 erreicht, wie dies auch bei ACA 4.0 gehandhabt wird (Abschnitt 3.7.2.4). Dies ist notwendig, um die 'wahre Antwort' innerhalb der Grenzen der Paarvergleichsantwortskala [-4; +4] zu halten.

6.4.1.2 Objektklasse

Da aus Kostengründen eine studentische Apn-Stichprobe angestrebt wurde, sollte die Objektklasse möglichst so gewählt sein, daß alle Studenten mit Beurteilungen und Entscheidungen innerhalb des fraglichen Objektbereichs einigermaßen vertraut sind. Dieses Auswahlkriterium führte uns zu einer Reihe von möglichen Untersuchungsobjektclassen. Aus diesen wählten wir die Objektklasse 'Studentenappartements' aus, weil mehrere andere veröffentlichte CA-Untersuchungen vorliegen, die ebenfalls mit diesem Objektbereich arbeiteten (Agarwal & Green, 1989; Corstjens & Gautschi, 1983; Elrod et al., 1992; Glazer, 1984; Green, Helsen & Shandler, 1988; Green, Krieger & Agarwal, 1991; Green, Krieger & Schaffer, 1993; Johnson & Meyer, 1984; Johnson, Meyer & Ghose, 1989; McCulloch & Best, 1979 und Segal, 1982). Dies war insbesondere in der Phase der Auswahl der Attribute hilfreich.

6.4.1.3 Attribute und Attributstufen

Nach Sichtung vorliegender CA zu Studentenappartements wählten wir fünf Attribute aus, die uns unter Berücksichtigung der lokalen Verhältnisse als die wichtigsten erschienen. Jedes Attribut wurde auf drei Stufen spezifiziert. Die Attribute und die Attributstufen wurden dem zuvor festgelegten wahren additiven Teilnutzenwert-PSM (Tabelle 6.1) zugeteilt. Bei der Zuweisung der inhaltlichen Bedeutung der Attribute und Attributstufen zu den Modellparametern versuchten wir, eine nicht alltägliche aber realistische Präferenzstruktur einer fiktiven Person nachzubilden. Dadurch wollten wir einerseits vermeiden, daß das wahre PSM der individuellen Nutzenstruktur einiger Apn sehr ähnlich ist, was einen verringerten Lernaufwand bedeuten würde. Andererseits sollte das wahre PSM die Apn nicht dadurch irritieren, daß es sehr unrealistisch ist. Die Attribute, die Attributstufen und die wahre Nutzenstruktur sind in Tabelle 6.2 aufgeführt. Die Attribute 1, 2 und 5 haben einen linearen Nutzenverlauf, während das

Tabelle 6.2: Die fünf Attribute mit jeweils drei Attributstufen und die wahre Nutzenstruktur.

Attribut und Attributstufe	'wahrer' TNW	relative Attributwichtigkeit	graphische Veranschaulichung
1. Entfernung zum Institut		100	Teilnutzenwert
1 km	100		
3 km	50		
5 km	0		
2. Größe der Wohnung		90	
35 m ²	90		
25 m ²	45		
15 m ²	0		
3. Art des Hauses		80	
Studentenwohnheim	80		
Einfamilienhaus	8		
Mehrfamilienhaus	0		
4. Zustand der Wohnung		70	
Neubau	70		
renovierter Altbau	49		
Altbau	0		
5. Mietpreis		50	
15 DM je m ²	50		
20 DM je m ²	25		
25 DM je m ²	0		

Präferenzmodell des Attributs 3 konvex und des Attributs 4 konkav verläuft.

Aus den fünf Attributen lassen sich insgesamt 3^5 Objekte konstruieren. Diese sind in Tabelle D des Anhangs aufgeführt. In der Untersuchungsphase wurden die Attribute in Form der kurzen 'characteristics'-Form repräsentiert, wie sie in Tabelle 6.2 aufgeführt sind. Die Apn wurden jedoch in der Lernphase ausführlich mit den Repräsentationsformen 'benefits' und 'image' (vgl. Abschnitt 3.2.5) vertraut gemacht (s. schriftliche Instruktionen in der Lernphase; Anhang E). Durch diese Vorgehensweise sollten die Vorteile der knappen Darstellung der objektiven 'characteristics'-Eigenschaften in der Untersuchung mit der Kontrollierbarkeit der Interpretationen dieser Eigenschaften kombiniert werden (vgl. Abschnitt 3.2.5).

6.4.1.4 Auskunftspersonen und Untersuchungsrahmenbedingungen

An der Untersuchung nahmen 24 Apn teil. Diese sind ohne Ausnahme Studierende der Westfälischen Wilhelms-Universität Münster. Das Alter beträgt im Durchschnitt 21,6 Jahre ($s=3,1$). 79% der Apn sind Studierende im Diplomstudiengang Psychologie.

Die Apn wurden per Aushang und in Lehrveranstaltungen um die Teilnahme an der Untersuchung gebeten. Die Dauer der Untersuchung wurde von 0900 Uhr bis maximal 1800 Uhr (mit Mittagspause) angekündigt. 16 der Apn (67%) entstammen einem Seminar für Erstsemester, das vom Autor zu dieser Zeit abgehalten wurde. Da über 60% der Seminarteilnehmer auch an der Untersuchung teilnahmen, kann davon ausgegangen werden, daß die Stichprobe nicht erheblich durch die abschreckend lange Untersuchungsdauer vorselektiert ist. Diese Studentinnen und Studenten waren innerhalb der ersten drei bis sechs Wochen ihres Studiums.

Die Apn-Stichprobe kann sicherlich nicht als repräsentativ für die Population von CA-Apn gelten. Schon der Median der Abiturnote ($Md=1,8$) weist darauf hin, daß die intellektuellen Fähigkeiten der untersuchten Apn überdurchschnittlich sein dürften. Aufgrund der Dauer der Untersuchung ist es jedoch außerordentlich schwierig, nichtstudentische Apn zur Teilnahme zu bewegen bzw. über die gesamte Untersuchungsdauer zu motivieren. Gerade dieser letzte Punkt ist nicht zu gering einzuschätzen. Da die Untersuchung für die Apn sehr anstrengend ist, können wahrscheinlich aussagekräftige Daten mit diesem

Forschungsansatz nur bei Apn gesammelt werden, die hinsichtlich der intellektuellen Fähigkeiten, der Konzentrationsfähigkeit und der Ausdauer überdurchschnittliche Leistungen erbringen.

Es fanden vier Untersuchungen mit jeweils fünf bis sieben Apn statt. Die Untersuchungstage gliederten sich in die Lernphase und die Untersuchungsphase. In der Lernphase wurden die Apn in der Gruppe unterrichtet. Bei der Untersuchungsphase befanden sich die Apn weiterhin im selben Raum, wobei sie jedoch ihre Tätigkeiten (meist am Computer) selbständig ausführten.

Eine Voruntersuchung (6 Apn) diente der Feinabstimmung der Untersuchungsteile und der Überprüfung der Laufstabilität der eingesetzten Software. Die in ihr gewonnenen Erkenntnisse führten z.B. zu einer Optimierung der Zeiteinteilung, zu einer Verbesserung der Inhalte der Lernphase und zur Bestimmung der optimalen Schwierigkeit der Aufgaben zur Lernerfolgsüberprüfung (s. Abschnitt 6.4.1.6).

6.4.1.5 Lernphase

Eine Hauptschwierigkeit unseres Forschungsdesigns besteht darin, die wahre Nutzenstruktur den Apn exakt vorzugeben, ohne ihnen eine numerische Vorstellung des wahren TNW-Modells nahezulegen. Dies würde wahrscheinlich dazu führen, daß die Apn nicht mehr 'natürlich' urteilen. Vielmehr würden sie dazu verleitet werden, ihre Urteile 'auszurechnen'. Da das Ziel der Präferenzstrukturmodellierung darin besteht, 'natürliche' Präferenzurteile zu modellieren und nicht darin, die Rechenfähigkeit der Apn zu überprüfen, wäre dies nicht wünschenswert.

Um das Ziel der möglichst genauen Induktion des wahren Modells zu erreichen, konstruierten wir die Lernphase in mehreren Stufen, um den Apn die Nutzenstruktur in zunehmender Genauigkeit beizubringen. Die eingesetzten schriftlichen Instruktionstexte sind in Anhang E beigefügt.

In der ersten Stufe der Lernphase erhielten die Apn zunächst ein *Szenario* per schriftlicher Instruktion. Kern dieses Szenarios ist die Beschreibung der fiktiven Person Helga K., in die sich die Apn während der Untersuchung hineinversetzen sollten. Die Apn wurden in dem Glauben gelassen, daß Helga tatsächlich existiert und das nachfolgende Lernmaterial den Urteilen dieser

Person entstammt. Tatsächlich handelt es sich um simulierte Urteile der wahren Nutzenstruktur.

Den Apn wird mitgeteilt, daß Helga auf Wohnungssuche ist. Danach schildert der Instruktionstext ihre groben Wertvorstellungen über Studentenwohnungen. Darin werden die beurteilungsrelevanten Attribute genannt. Die jeweils höchste und niedrigste Ausprägung der Attribute kommt im Text zur Sprache, und die relative Wichtigkeit der Attribute wird in eine Rangreihe gebracht. Das Szenario legt besonderen Wert darauf, Helgas Präferenzstruktur plausibel zu begründen. Dies ist notwendig, um in den folgenden Stufen der Lernphase verständlich argumentieren zu können.

Die zweite Stufe dient der *vergleichenden Quantifizierung* von Helgas Vorstellungen. Hierzu erhalten die Apn zunächst die Aufgabe, die Attributwichtigkeiten auf einer Skala einzuschätzen (s. Anhang E). Nachdem dies erfolgt ist, wird den Apn erzählt, daß Helga einen Makler aufgesucht hat, der sie bat, zunächst einen Fragebogen zur Klärung ihrer Vorstellungen auszufüllen. Darin waren genau dieselben Skalen enthalten, wie sie von den Apn soeben bearbeitet worden waren. Die wahren Ergebnisse werden den Apn rückgemeldet, anhand des Szenariotextes diskutiert und ausführlich begründet.

Genauso wird mit der Quantifizierung des Nutzenverlaufs innerhalb der Attribute verfahren. Hierbei wird besonderer Wert auf die Attribute mit nicht-linearem Verlauf gelegt. Nach Abschluß der Lernstufe der vergleichenden Quantifizierung ist die wahre Nutzenstruktur prinzipiell eindeutig festgelegt.

Die dritte Stufe der Lernphase dient der Einübung der wahren Nutzenstruktur durch *vergleichende Beurteilungen von Wohnungskonzepten*. Hierzu wurde den Apn per Tageslichtprojektor jeweils ein Paar, ein Tripel oder ein Quadrupel von Wohnungskonzepten vorgegeben. Die Apn hatten die Aufgabe, die Objekte gemäß Helgas Präferenzen in eine Rangfolge zu bringen und die Lösung zu begründen. Nach einer kurzen Gruppendiskussion wurden Helgas Entscheidungen den Apn rückgemeldet und bei Bedarf detailliert begründet.

Diese Lernstufe wurde in vier Blöcken mit jeweils fünf bzw. sechs vergleichenden Beurteilungsaufgaben durchgeführt. Tabelle F im Anhang gibt Aufschluß darüber, welche Apartmentkonzepte für die insgesamt 21 Beurteilungsaufgaben herangezogen wurden. Die Reihenfolge dieser Aufgaben ist dabei so gewählt, daß es immer schwieriger wird, Helgas Präferenz zu finden. Die Lernziele dieser Stufe sind:

- ❑ (alle Aufgaben) Vertrautmachen der Apn mit den Urteilsaufgaben: Lesen und Vergleichen der Alternativen.
- ❑ (alle Aufgaben) Einprägen der relativen Wichtigkeit der Attribute und Einprägen der Beurteilung der Merkmalsstufen.
- ❑ (ab Lernblock b) Verdeutlichung der Relationen der Merkmalsnutzen.

6.4.1.6 Lernerfolgsüberprüfung und -rückmeldung

Nach jedem der vier Aufgabenblöcke findet eine Kontrolle des Lernerfolgs statt. Bei diesen Lernchecks erhielten die Apn sieben Karten auf denen jeweils ein Objekt auf allen Attributen spezifiziert war (Vollprofile). Sie hatten die Aufgabe, die Karten in Einklang mit Helgas Vorstellungen (also gemäß der gelernten Nutzenstruktur) in eine Rangreihe zu bringen. Die mündlich gegebenen Instruktionen zu dieser Aufgabe sind in Anhang G aufgeführt.

Nach der Rangreihung gaben die Apn die Rangfolge der Karten, auf denen eine Codenummer notiert war, am Computer ein und erhielten eine Rückmeldung über ihre Leistung. Hierzu wurde die Rangkorrelation zwischen der empirischen Rangreihe und der wahren Rangreihe berechnet. Den Apn wurde die Güte ihrer Rangreihe in Abhängigkeit der Rangkorrelation rückgemeldet. Zusätzlich zum Leistungsgrad erhielten die Apn auch eine Rückmeldung über Helgas (die wahre) Rangreihe. Darüber hinaus wurden sie aufgefordert, ihre Rangreihe mit Helgas Urteilen zu vergleichen und die Ursachen für eventuelle Fehleinschätzungen zu suchen und zu verstehen.

Die primäre Funktion der Lernerfolgsmessung ist die beschriebene Leistungsrückmeldung und der Beleg des Lernerfolgs. Zusätzlich dient sie dem Zweck abzuschätzen, in welchem Maße die Nutzenähnlichkeit der Beurteilungsobjekte zu einer Steigerung der Schwierigkeit der Beurteilungsaufgabe führt. Damit sollten die Befunde der empirischen Untersuchung I (CA zu Prüfungsordnungen) mit einer anderen Methode noch einmal bestätigt werden.

Hypothese 5: Je höher die Nutzenähnlichkeit von Beurteilungsobjekten ist, desto schwieriger ist die Beurteilungsaufgabe. Eine hohe Schwierigkeit führt zu einer Abnahme der Reliabilität der Rangreihe.

Sollte sich diese Vermutung bestätigen, unterstreicht dies noch einmal die Bedeutung des Reliabilitätsproblems der Kriterienaufgabe (siehe Abschnitte

5.4.1 und 6.3.2), was eine weitere Rechtfertigung des Untersuchungsansatzes 'Induktion von wahren Werten' bedeuten würde.

Die vier Lernchecksets aus je sieben Vollprofilkonzepten wurden folgendermaßen konstruiert: Wir ordneten alle möglichen Beurteilungsobjekte in absteigender Reihenfolge ihres wahren Nutzens an (Tabelle D im Anhang). Drei Lernchecksets A, B und C wurden gebildet, indem zunächst an einer zufälligen Stelle im oberen Bereich der Rangliste ein Objekt ausgewählt wurde. Das Objekt, das 25 Stellen tiefer in der Liste stand, wurde als zweites Objekt hinzugewählt. Weitere 25 Stellen tiefer in der Liste wurde das dritte Objekt entnommen usw., bis die sieben Objekte eines Lernchecksets vollständig waren. Der Listenabstand 25 ergab sich aus der Voruntersuchung, in der sich der mittlere Schwierigkeitsgrad bei der Rangreihung von solchen Lernchecksets gezeigt hatte. Die wahre Nutzendifferenz benachbarter Lerncheckobjekte liegt im Mittel bei $\bar{u}_A=23,3$ ($\bar{u}_A=24,2$ $\bar{u}_B=22,3$ $\bar{u}_C=23,3$).

Das vierte Objektset D wurde so konstruiert, daß nicht jedes 25. Objekt, sondern jedes 18. Objekt der Liste ausgewählt wurde. Das arithmetische Mittel der Nutzendifferenz benachbarter Lerncheckobjekte ist $\bar{u}_D=16,7$ und ist damit um ca. 28% geringer als die Nutzendifferenz der anderen Lernchecksets. Die Bildung der Rangfolge innerhalb dieses Lernchecksets sollte aufgrund der geringeren Nutzendifferenz schwieriger sein, was in einer verminderten Übereinstimmung der empirischen Rangfolge mit der wahren Rangfolge resultieren sollte. Der Hälfte der Apn (zufällige Apn-Zuordnung innerhalb eines jeden Untersuchungstages) wurden die Lernerfolgsaufgaben in der Reihenfolge A, B, C, D vorgegeben, die andere Hälfte hatte diese in der Reihenfolge A, B, D, C zu bearbeiten. Durch die Balancierung der seriellen Position der Sets C und D sollten Reliabilitätsunterschiede auf die verminderte Nutzendifferenz des Sets D rückführbar sein. Abbildung 6.9 skizziert das Prinzip der Lernphase Stufe III schematisch.

Die Lernphase mit seinen drei Stufen und den vier Lernerfolgsüberprüfungen während der dritten Stufe dauerte ca. 2,5 bis 3 Stunden. An die Lernphase schließt sich die eigentliche Untersuchung an.

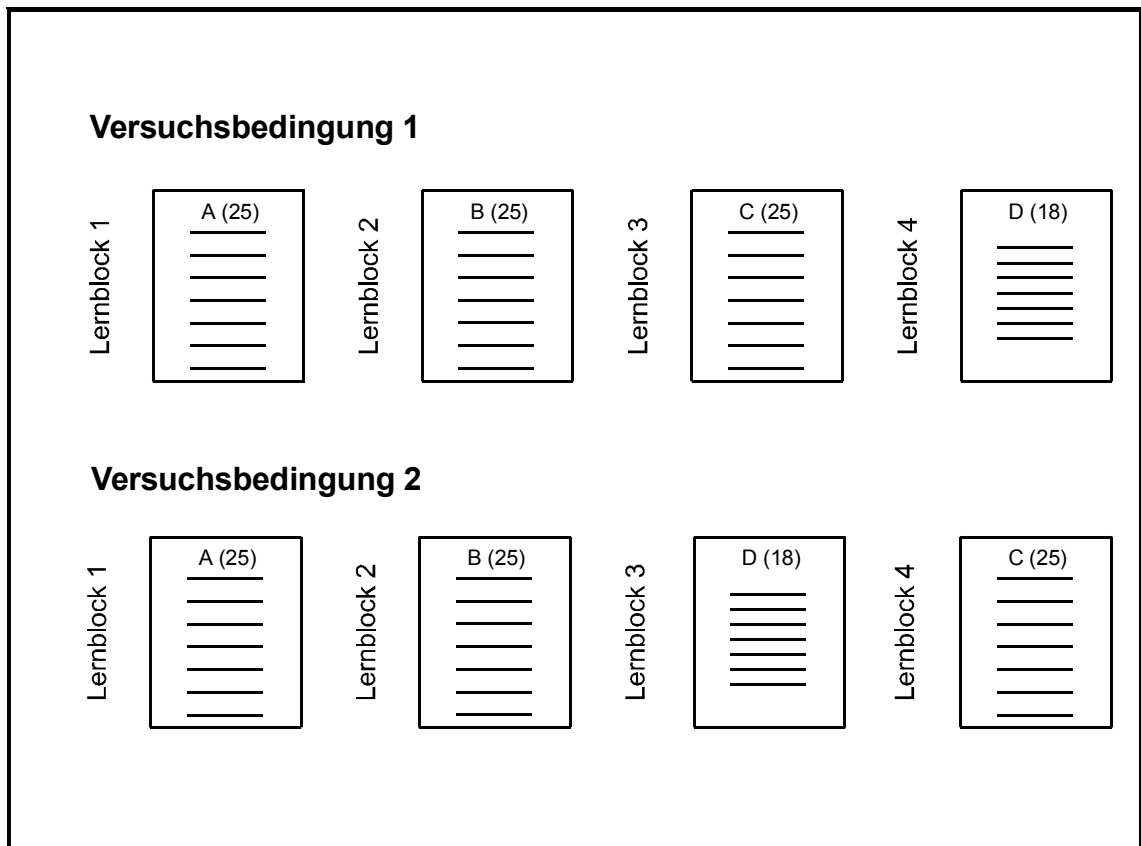


Abbildung 6.9: Schematische Darstellung der dritten Stufe der Lernphase (vergleichende Beurteilung von Wohnungskonzepten) mit seinen vier Lernblöcken und den vier Lernerfolgsmessungen A, B, C und D. Die Objekte des Lernchecksets D haben eine geringere mittlere Nutzendifferenz als die Objekte der Sets A, B und C. Die Reihenfolge der Sets wurde experimentell variiert.

6.4.1.7 Untersuchungsaufbau und Versuchsplan

Dieser Abschnitt beschreibt den Aufbau der Untersuchung. Dabei werden Teilhypothesen formuliert und Hinweise auf die Ergebnisauswertung gegeben. Abbildung 6.10 gibt einen Überblick über den Aufbau der Untersuchung.

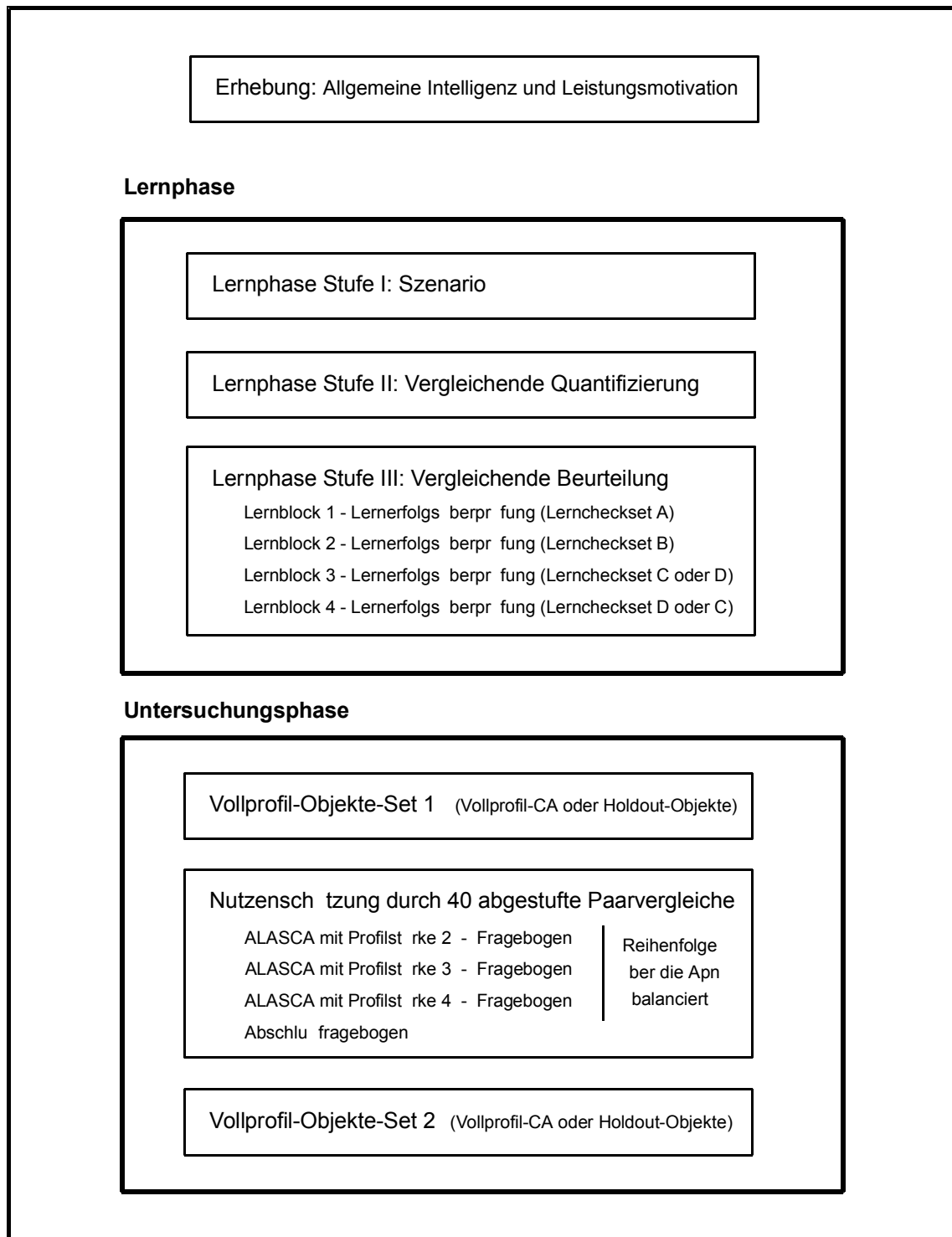


Abbildung 6.10: Schematischer Aufbau der empirischen Untersuchung. Erläuterungen im Text.

In Abschnitt 6.3.4 haben wir unser spezielles Forschungsdesign ausführlich dargestellt und begründet. Bei diesem Design werden die zu untersuchenden Nutzenschätzverfahren von zwei Vollprofil-Objekte-Sets umrahmt. Den Zweck

dieser Vorgehensweise haben wir an dieser Stelle erläutert. Die beiden Vollprofil-Objekte-Sets wurden mit der SPSS-Prozedur ORTHOPLAN erstellt. Es handelt sich dabei um nahezu orthogonale Designs mit jeweils 15 Objekten. Fünf dieser Objekte sind in beiden Sets vorhanden. Tabelle D im Anhang zeigt die Objekte die beiden Vollprofil-Objekte-Sets.

Die Apn haben die Aufgabe, die Objekte gemäß Helgas Vorstellungen in eine Rangreihe zu bringen. Die mündliche Instruktion zu dieser Aufgabe ist in Anhang H abgedruckt. Die Apn sind mit dieser Art von Rangreihungsaufgaben zu diesem Zeitpunkt bereits durch die vorangegangenen vier Lernerfolgsmessungen vertraut, bei denen sie sieben Objekte in eine Rangreihe brachten. Bei der jetzigen Aufgabe müssen nicht sieben, sondern 15 Karten sortiert werden. Da dies schwieriger ist, erhalten die Apn eine Instruktion zur strategischen Vorgehensweise: Sie sollen die Objekte zunächst in drei Kategorien einteilen (gute, mittlere und schlechte Wohnungen) und dann die Objekte der drei Kategorien in eine Rangreihe bringen. Anschließend können die drei Rangreihen zu einer Reihe zusammengefügt werden, indem die Kategorien-grenzen noch einmal überprüft werden. Per Instruktion sind die Apn auf die Aufgabendauer von maximal 30 Minuten eingestellt. Wir erfaßten die tatsächliche Dauer der Aufgabe.

Aus dem orthogonalen Design und der Antwortrangreihe lassen sich die Parameter eines additiven Teilnutzenwertmodells schätzen. Beide Aufgaben kann man somit als eine eigenständige klassische Vollprofil-CA (Abschnitt 3.7.2.1) oder auch als Holdout-Aufgabe (Abschnitt 5.4) auffassen. Zunächst sollen zwei Annahmen überprüft werden, die man als notwendige Bedingung für die Interpretierbarkeit der weiteren Befunde ansehen kann (vgl. Abschnitt 6.3.4):

Annahme 1: In der Lernphase wird die wahre Nutzenstruktur erfolgreich induziert.

Die Überprüfung dieser Annahme ist über den Vergleich der theoretisch zu erwartenden Rangreihe mit der empirischen Rangreihe möglich. Eine andere Vorgehensweise wäre der Vergleich des geschätzten PSM mit dem wahren PSM (Strukturwiedererkennung).

Annahme 2: Die gelernte Nutzenstruktur bleibt über die Zeitspanne, in der die zu untersuchenden Paarvergleichs-Nutzenschätzverfahren (ALASCA) bearbeitet werden, stabil.

Diese Annahme läßt sich über den Vergleich der zweiten Vollprofil-Aufgabe mit der ersten Aufgabe überprüfen. Die beiden PSM, ihre Strukturwiedererkennungsleistung und Übereinstimmungsmaße zwischen der theoretischen und empirischen Rangreihung sollten ähnlich sein. Darüber hinaus sollte die Rangreihe der fünf Objekte, die in beiden Sets vorkommen, übereinstimmen. Fallen diese Kennwerte nicht im Sinne einer *Stabilität der wahren Nutzenstruktur* aus, so kann dies entweder an einer Veränderung der gelernten Nutzenstruktur zwischen den beiden Vollprofil-Aufgaben oder an einer niedrigen Reliabilität der Aufgaben liegen.

Die zu untersuchenden Paarvergleichs-Nutzenschätzverfahren führten wir mit dem Softwaresystem ALASCA (Abschnitt 6.1) durch. In einem Meßwiederholungsdesign bearbeitete jede Ap drei verschiedene CA-Interviews. Die Apn beantworteten hierzu jeweils 40 abgestufte Paarvergleiche der Profilstärke 2, 3 und 4. Die Apn wurden per Instruktion darauf eingestellt, maximal eine halbe bis eine Minute für jeden Paarvergleich zu verwenden (s. mündliche Instruktionen in Anhang H). Dieser Zeitansatz entspricht bei 40 Paarvergleichen einer Gesamtdauer von durchschnittlich maximal 30 Minuten. Die instruierte Gesamtdauer einer ALASCA-Aufgabe entspricht damit derjenigen einer Vollprofil-Aufgabe. Wir erhoben die tatsächliche Dauer eines jeden ALASCA-Paarvergleichs. Neben der Information, in welcher Relation die Gesamtdauer der ALASCA-Aufgabe zur Dauer der Vollprofil-CA steht, dient dies der Überprüfung folgender Hypothese:

Hypothese 3a: Die Paarvergleichsaufgaben konsumieren mit zunehmender Profilstärke mehr Zeit.

Dies ist eine Teilhypothese der Hypothese 3. Sie wird im Zusammenhang mit der Hypothese 3 begründet.

Nach jedem ALASCA-Durchgang wurde computergestützt ein Fragebogen ('Zwischenfragebögen') vorgegeben. Die Fragen waren für die verschiedenen Profilstärkebedingungen identisch. Nach Abschluß aller drei ALASCA-Paarvergleiche und der dazugehörigen Fragen beantworteten die Apn noch mehrere allgemeine Fragen ('Abschlußfragebogen'). Die Items der Zwischenfragebögen und des Abschlußfragebogens sind in den Tabellen I-1 und I-2 des Anhangs aufgeführt. Welche Information sollte mit den Fragebögen erhoben werden und welche Ergebnisse sind zu erwarten?

Hypothese 3b: Die Paarvergleichsaufgaben nehmen mit zunehmender Profilstärke mehr kognitive Ressourcen in Anspruch.

Die Hypothese 3b überprüft eine Annahme der Hypothese 3. Die Items 1, 6 und 10 (Einschätzung der Schwierigkeit der Paarvergleichsaufgaben; Einschätzung, wie schwierig es fiel, sich zu konzentrieren; Einschätzung, wie schwierig es fiel, eine Antwort zu finden; s. Tabelle I-1) sind Indikatoren der kognitiven Ressourcen, die zur Aufgabenbewältigung eingesetzt werden.

Den Apn wurde fälschlicherweise glaubhaft gemacht, daß die Dauer des Interviews sowohl von der Anzahl der abgefragten Paarvergleiche als auch von der Qualität der Antworten abhängig sei (vgl. Anhang H). Wir wählten diese Instruktion, um die Auskunftsmotivation zu steigern. Nimmt man an, daß die Apn danach streben, die Zeitdauer der Untersuchung möglichst zu minimieren, dann sollte die gewählte Instruktion zur Folge haben, daß die Apn die Paarvergleiche sowohl zügig als auch gewissenhaft bearbeiten. Dies sollte insbesondere dazu beitragen, daß die Auskunftsmotivation zwischen den Profilstärkebedingungen weitgehend konstant bleibt. Um die Konzentrationsfähigkeit der Apn zu erhalten, legten die Apn zwischen den 40 ALASCA-Paarvergleichen und den dazugehörigen Fragen jeweils 10 Minuten Pause ein.

Durch appellhafte Instruktionen und die o.g. Fehlinstruktion wurde versucht, die Auskunftsmotivation der Apn über die ALASCA-Bedingungen möglichst hoch und möglichst konstant zu halten. Da die direkte Vergleichbarkeit der drei ALASCA-Aufgaben aber trotzdem fraglich ist (z.B. könnte man erwarten, daß die erste ALASCA-Aufgabe die Apn mehr anspricht, weil die Aufgabe neu ist und die Apn noch nicht so ermüdet sind), balancierten wir die Reihenfolge der Profilstärkebedingungen über die Apn (vollständige Permutation).

Bevor die Apn die ALASCA-Interviews und die Fragebögen bearbeiteten, wurden sie per Beispielprogrammlauf mit der Bedienung beider Softwaresysteme vertraut gemacht.

Welche Ergebnisse lassen sich aus den ALASCA-Paarvergleichen ableiten und welche Ergebnisse sind zu erwarten?

Hypothese 1a: Die kriterienorientierte Validität und die Strukturwiedererkennungslleistung der geschätzten PSM hängt von der Profilstärke ab.

Hypothese 1b: Die kriterienorientierte Validität und die Strukturwiedererkennungslleistung der geschätzten PSM hängt von der Anzahl der Paarvergleiche ab.

Die Begründung ist in Zusammenhang mit der Hypothese 1 in Abschnitt 6.2 zu finden. Dort wurde die Wirkung der Objektprofilstärke und der Anzahl der Paarvergleiche aus der statistischen und der psychologischen Perspektive dargestellt. Da die Wirkungen gegenläufig sind, kann die genaue Art des Zusammenhangs zwischen den Meßgütekriterien und diesen beiden Faktoren nicht vorhergesagt werden. Diese ist somit empirisch zu bestimmen. Die kriterienorientierte Validität der ALASCA-PSM hinsichtlich der beiden Holdout-Aufgaben kann auf individueller Ebene bestimmt werden. Diese Bestimmung entspricht der weitverbreiteten Vorgehensweise zur Bestimmung der kriterienorientierten Validität (Abschnitt 5.4). Das Verfahren der Wiedererkennungslleistung der wahren Nutzenstruktur ist in Abschnitt 5.5.1 beschrieben. Bei der Beurteilung der Meßgüte hat auch der *Zeitbedarf* für deren Erreichung eine wichtige Bedeutung.

Neben der Beantwortung dieser Hauptfragestellung soll hier auch die *Meßgüte der ALASCA-Verfahren mit der Meßgüte der beiden Vollprofil-CA-Verfahren verglichen* werden. Hierzu dienen sowohl Maßzahlen der Strukturwiedererkennungslleistung als auch der kriterienorientierten Validität (bei der Vollprofil-CA jeweils hinsichtlich des anderen Holdout-Sets). Neben der Meßgüte ist auch der Zeitbedarf der Nutzenschätzverfahren von Bedeutung.

Die Antworten der Apn auf die Paarvergleiche können mit den wahren Antworten verglichen werden, die durch die wahre Nutzenstruktur festgelegt sind. Dadurch lassen sich Rückschlüsse auf den Antwortfehler ziehen. Dies dient der Überprüfung einer Teilhypothese der Hypothese 3. Sie ist dort begründet:

Hypothese 3c: Mit zunehmender Profilstärke vergrößert sich der Antwortfehler.

In unserer theoretischen Sichtweise, wie sie in Abschnitt 6.2 dargelegt wurde, hängt der Antwortfehler des weiteren von der Auskunftsmotivation und vom

Verhältnis der kognitiven Ressourcen zur Aufgabenschwierigkeit ab. Zu Beginn der Untersuchung führten wir zwei psychometrische Testverfahren durch. Dies diente dem Zweck, ein Bild über die Eigenschaften der Apn hinsichtlich ihrer allgemeinen Intelligenz (Indikator der kognitiven Ressourcen) und ihrer Leistungsmotivation (Indikator für die Grundtendenz, in Leistungssituationen motiviert zu reagieren) zu erhalten. Da die Dauer des gesamten Untersuchungstages ohnehin schon kritisch war, entschieden wir uns dabei für Verfahren, die mit sehr geringem Zeitaufwand durchführbar sind.

Als Indikator der kognitiven Ressourcen wählten wir einen Subtest des LPS (Leistungsprüfsystem; Horn, 1962), die LPS-Skala 12, weil dieser eine sehr hohe Korrelation mit dem Gesamttest besitzt ($r=0,86$; Horn, 1962). Zusätzlich zu dieser LPS-Skala nehmen wir an, daß auch die Abiturnote (Item 28 des Abschlußfragebogens) als Indikator der kognitiven Leistungsfähigkeit aufgefaßt werden kann.

Das Konstrukt der Leistungsmotivation erklärt eine allgemeine Disposition, motiviert in Leistungssituationen zu reagieren. Als Instrument zur Erfassung der Leistungsmotivation diente uns eine deutsche Fassung der MARPS (Mikula, Schwinger & Uray 1976; im Original: Mehrabian Achievement Risk Preference Scale, Mehrabian, 1968, 1969), die zur Unterscheidung von hoch und niedrig leistungsmotivierten Personen (Atkinson, 1964) konstruiert wurde. Das Instrument MARPS und die Instruktionen zu den beiden Verfahren sind in Anhang J und K beigefügt.

Welche Befunde sind hinsichtlich der kognitiven Ressourcen und der Leistungsmotivation zu erwarten?

Hypothese 6: Die PSM von Apn mit *hoher kognitiver Leistungsfähigkeit zeigen eine bessere Meßgüte*, als solche mit niedriger Leistungsfähigkeit. Die PSM von Apn mit *hoher Leistungsmotivation zeigen eine bessere Meßgüte* als solche mit niedriger Leistungsmotivation.

Die Untersuchungsphase dauerte ca. 3,5 Stunden. Ein gesamter Untersuchungstag dauerte mit der Lernphase und einer einstündigen Mittagspause damit etwa 7 bis 7,5 Stunden. Die Mittagspause wurde nach dem dritten Lerncheck abgehalten. Diesen Zeitpunkt erachten wir als günstig, weil so die Apn

nach einer kurzen Auffrischung des Lernmaterials möglichst ausgeruht die Untersuchungsphase beginnen.

6.4.1.8 Parameterschätzung

Bei der Schätzung der Modellparameter wurde davon ausgegangen, daß zu Beginn keine Initialschätzung vorliegt. Damit sofort ab dem ersten Paarvergleich eine Schätzung der TNW möglich ist, wurde die Parameterschätzmethode nach Gleichung 3.27 eingesetzt. Die Elemente des Vektors der initialen (bzw. expliziten) TNW wurden gleich Null gesetzt (vgl. Gleichung 3.27 in Abschnitt 3.8.1). Darüber hinaus bleibt durch die Aufnahme der expliziten TNW die Vergleichbarkeit von ALASCA mit ACA 4.0 gewahrt.

Bei dieser Art der Vorgehensweise berücksichtigt man bei der nachfolgenden Schätzung der TNW (wie auch bei ACA) weiterhin die expliziten TNW. Dies ist dann sinnvoll, wenn man annimmt, daß die expliziten TNW valide Information bereitstellen. Enthalten sie jedoch wie in unserem Falle keine Information, dann ist die Parameterschätzung auf der Grundlage nachfolgender Paarvergleiche nicht optimal, da Fehlervarianz 'mitgeschleppt' wird. Dies sollte sich jedoch nicht auf unsere Untersuchungsergebnisse auswirken, weil diese Art der Parameterschätzung über alle Manipulationen konstant gehalten wird.

6.4.2 Ergebnisse

In diesem Abschnitt sind die Ergebnisse der empirischen Untersuchung dargestellt. Dabei werden die Hypothesen einer Prüfung unterzogen.

6.4.2.1 Lernerfolg und Abhängigkeit der Ergebnisse vom Objektset

In der Lernphase wurde viermal der Lernerfolg gemessen und an die Apn zurückgemeldet. Die Lernchecksets wurden speziell ausgewählt, um Hypothese 5 zu überprüfen.

Abbildung 6.11 zeigt für die vier Lernerfolgsmessungen die mittleren Rangkorrelationen der empirischen Rangreihen mit den wahren Rangreihen getrennt für die Versuchsbedingungen 1 und 2 (Rangkorrelationen Fisher-Z

transformiert, gemittelt und rücktransformiert; Faverge, 1980)⁶. Zunächst ist festzuhalten, daß die Korrelationen im allgemeinen sehr hoch sind. Dies bedeutet, daß das Ziel der Lernphase, die Induktion der wahren Nutzenstruktur, erreicht wurde. Dies war der Hauptzweck der Lernerfolgsmessungen.

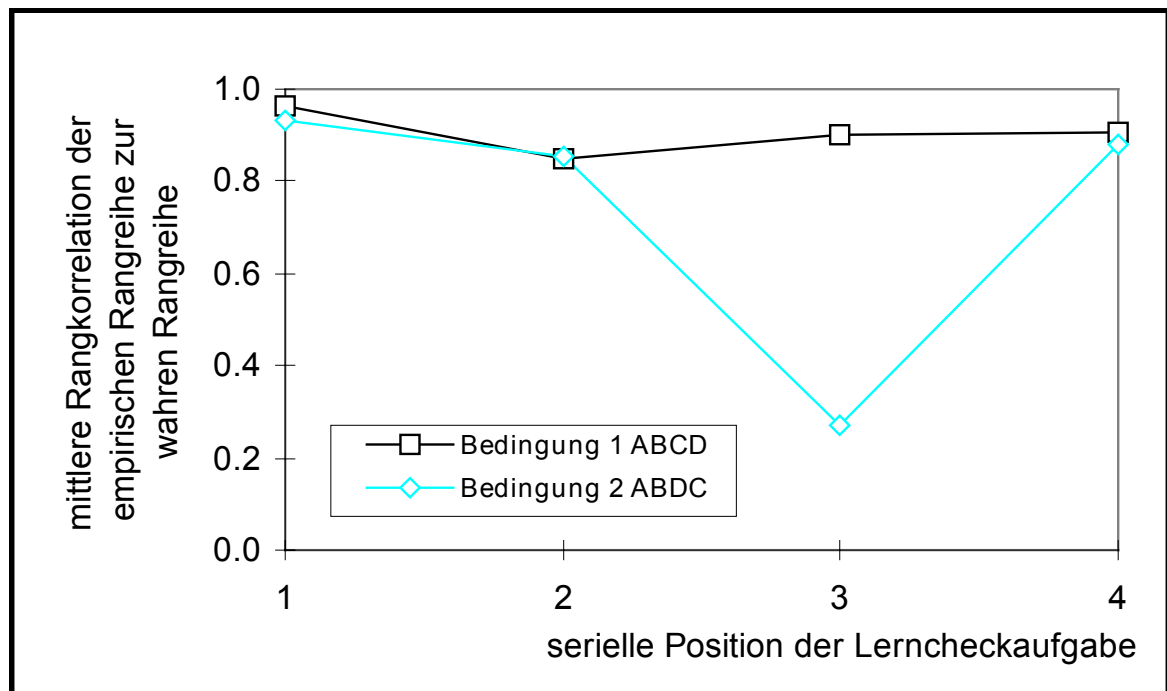


Abbildung 6.11: Mittlere Rangkorrelation der empirischen mit den wahren Rangreihen getrennt für die Versuchsbedingungen 1 und 2.

Daneben zeigen sich noch andere Effekte. Beim schwierigeren Lerncheckset D erreichten die Apn durchschnittlich eine deutlich geringere Rangkorrelation ($r_{s(D)}=0,59$) als beim Set C, das mit D hinsichtlich der seriellen Position balanciert dargeboten wurde ($\bar{r}_{s(C)}=0,89$; $F_{1,22}=13,54$ $p<0,01$; mit dem Lerncheckset als Meßwiederholungsfaktor und der Fisher-Z transformierten Korrelation als abhängige Variable). Dieser Effekt belegt die Behauptung der Hypothese 5, daß sich die wahre Nutzendifferenz der Beurteilungsobjekte eines Objektssets auf die Schwierigkeit der Beurteilungsaufgabe und damit auch auf die Reliabilität der Messung auswirkt.

⁶ Die Ergebnisse lassen sich auch replizieren, wenn nicht mit Fisher-Z transformierten Korrelationskoeffizienten gerechnet wird. Die Maßzahlen unterscheiden sich in der zweiten Nachkommadezimalen um maximal 0,03. Dies gilt auch für alle nachfolgenden Berechnungen.

Der Effekt der verminderten Rangkorrelation tritt beim Lerncheckset D nur bei der Versuchsbedingung 2 (ABDC), nicht aber bei der Versuchsbedingung 1 (ABCD) auf. Technisch gesprochen heißt dies: Es besteht eine Wechselwirkung zwischen der Versuchsbedingung (1, 2) und dem Lerncheckset (C, D) ($F_{1,22}=14,96$ $p<0,01$; mit dem Lerncheckset als Meßwiederholungsfaktor und der Fisher-Z transformierten Korrelation als abhängige Variable). Dieser Effekt wurde nicht erwartet. Wir versuchen deshalb im nachhinein zu erklären, weshalb sich die Rangkorrelation des Objektsets D in Versuchsbedingung 1 gegenüber den übrigen Objektsets kaum verringert:

- Die Apn der Bedingung 2 erledigen die schwierige Aufgabe des Sets D nach drei Lerndurchgängen. Dagegen haben die Apn der Bedingung 1 bereits vier Lerndurchgänge erhalten, wenn sie die schwierige Aufgabe des Sets D zu lösen haben. Zu diesem Zeitpunkt haben die Apn deshalb einen höheren Kenntnisstand, der zur Lösung der schwierigen Aufgabe ausreicht.
- Die Apn der Bedingung 2 erhielten das schwierige Set D völlig unvermutet, nachdem sie bereits zwei Lernerfolgsmessungen mit guter Leistung bestanden haben. Die Apn der Bedingung 1 dagegen haben bereits von der Minderleistung einiger Apn erfahren und sind somit vorgewarnt. Zudem mußte der Versuchsleiter nach den schlechten Leistungen motivierend auf die Gruppe einwirken, um die Leistungsbereitschaft der Apn zu erhalten. Vermutlich führte dies zu einer Steigerung der Auskunftsmotivation.

Zusammenfassend läßt sich folgendes festhalten: Die Befunde zur Lernerfolgsmessung bestätigen die starken Effekte, die eine geringfügige Veränderung der Schwierigkeit der Beurteilungsaufgabe auf die Qualität der Beurteilungsergebnisse zur Folge haben kann. Die Rangfolge von Vollprofilen war der wahren Rangfolge dann sehr ähnlich, wenn der Durchschnitt der Nutzen-differenzen der Objekte $\bar{U}=23,3$ betrug. Vermindert man diese um 28%, dann sinkt die Ähnlichkeit der empirischen Rangfolge zur wahren Rangfolge drastisch ab. Dieser Befund macht noch einmal deutlich, wie ernst das Problem der Kriterienreliabilität zu nehmen ist.

Die Wechselwirkung zwischen der Versuchsbedingung und dem Lerncheckset zeigt dagegen, daß dieser Effekt nicht immer auftreten muß. Die deutlich verminderte Leistung tritt bei schwierigeren Aufgaben dann nicht auf,

wenn die Apn besondere Beurteilungsfähigkeiten und eine hohe Motivation besitzen, die Objekte genau zu beurteilen.

6.4.2.2 Stabilität der induzierten Nutzenstruktur

In Annahme 1 wurde die Forderung aufgestellt, daß das erste Vollprofil-Objekte-Set erkennen lassen sollte, daß die wahre Nutzenstruktur erfolgreich induziert

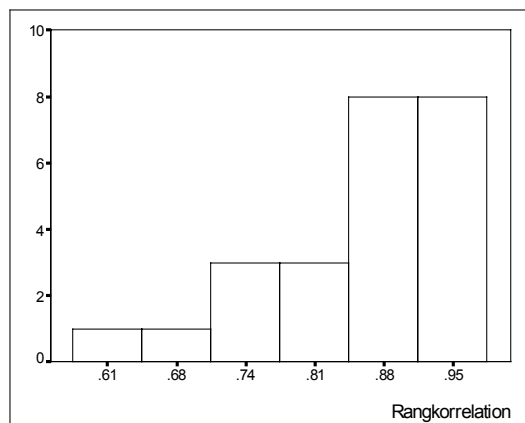


Abbildung 6.12: Histogramm der Rangkorrelation zwischen der wahren und der empirischen Objekt-rangreihe beim Holdout-Set 1.

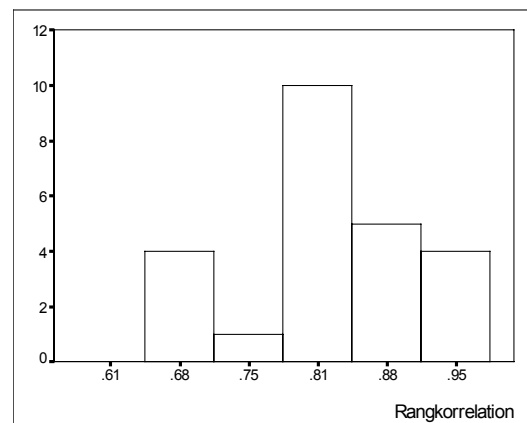


Abbildung 6.13: Histogramm der Rangkorrelation zwischen der wahren und der empirischen Objekt-rangreihe beim Holdout-Set 2.

wurde. Abbildung 6.12 zeigt ein Histogramm der Rangkorrelationen zwischen der empirischen Rangreihe und der wahren Rangreihe der Holdout-Aufgabe 1. Der mittlere Rangkorrelationskoeffizient liegt bei $\bar{r}_{\text{s-Set1-wahr}}=0,88$ (Mittelung nach Fisher-Z Transformation). Neben den Ergebnissen der Lernerfolgsmessungen ist dies ein weiterer klarer Hinweis auf die erfolgreiche Induktion der wahren Nutzenstruktur.

Auf der Basis dieser empirischen Daten lassen sich individuelle TNW-PSM schätzen. Dies wurde mit der SPSS-Prozedur CONJOINT durchgeführt. Die Maßzahl $W(\text{SSE})$ (s. Anhang C) ist ein Indikator der Strukturwiedererkennung des wahren PSM. Die mittlere Strukturwiedererkennungsleistung beträgt $\overline{W(\text{SSE})}_{\text{Set1-wahr}}=286,17$ ($s=60,28$). Um sich eine Vorstellung von der Güte dieser Strukturwiedererkennungsleistung zu bilden, mag der Hinweis hilfreich sein, daß sie einer mittleren Produkt-Moment-Korrelation zwischen den geschätzten und den wahren TNW von $\bar{r}_{\text{TNW1-TNWwahr}}=0,88$ entspricht. Auch der Vergleich des

geschätzten PSM mit der wahren Nutzenstruktur belegt damit deutlich den Erfolg der Lernphase.

Annahme 2 stellte die Forderung nach der Stabilität der Nutzenstruktur, um die Befunde des Meßwiederholungsdesigns eindeutig interpretieren zu können. Die mittlere Rangkorrelation der empirischen Rangreihe mit der wahren Rangreihe des Holdout-Sets 2 beträgt $\bar{r}_{s\text{-Set2-wahr}}=0,84$. Die Verteilung dieser Rangkorrelationen ist in Abbildung 6.13 dargestellt. Die Rangkorrelationen des ersten und des zweiten Sets unterscheiden sich geringfügig, aber statistisch signifikant ($F_{1,23}=6,23$ $p<0,05$; mit dem Vollprofil-Objekte-Set als Meßwiederholungsfaktor). Dieser Effekt dürfte insbesondere daran liegen, daß sehr gute Rangreihen bei der zweiten Holdout-Aufgabe seltener auftreten.

Das auf der Grundlage der zweiten Vollprofil-CA geschätzte PSM erkennt die wahre Nutzenstruktur im Mittel mit $\overline{W(SSE)}_{\text{Set2-wahr}}=284,39$ ($s=60,17$) wieder (dies entspricht einer mittleren Produkt-Moment-Korrelation zwischen den geschätzten und den wahren TNW von $\bar{r}_{\text{TNW2-TNWwahr}}=0,88$). Dieser Wert ist fast identisch mit der Strukturwiedererkennungsmaßzahl der ersten Vollprofil-CA. Die Meßgütekriterien der beiden Rangreihen bzw. der beiden Vollprofil-CA unterscheiden sich nicht ($F_{1,23}=0,02$ $p>0,8$; mit dem Vollprofil-Objekte-Set als Meßwiederholungsfaktor).

Diese Inkonsistenz der Befunde - ein Unterschied der Rangkorrelationen zwischen der empirischen und der wahren Rangreihe und kein Unterschied der Strukturwiedererkennungsleistung der geschätzten PSM - ist insbesondere deswegen verwunderlich, weil die Maße $\bar{r}_{s\text{-Set1-wahr}}$ und $\bar{r}_{s\text{-Set2-wahr}}$ bzw. $\overline{W(SSE)}_{\text{Set2-wahr}}$ jeweils sehr hoch korreliert sind ($r_1=-0,95$ bzw. $r_2=-0,95$).

Man kann die dargestellten Befunde dahingehend interpretieren, daß die Leistung der Apn in der zweiten Vollprofil-Rangreihung gegenüber der ersten geringfügig abnimmt. Diese geringfügig verminderte Leistung muß jedoch nicht unbedingt an der Veränderung der induzierten Nutzenstruktur liegen. Da die Rangreihung des zweiten Vollprofil-Objekte-Sets die letzte Aufgabe nach einem langen Untersuchungstag ist, könnte sie auch durch ein Nachlassen der Auskunfts motivation zustande gekommen sein. Die mittleren Beantwortungszeiten liegen bei 27,8 Minuten ($s=3,7$) für die Vollprofil-Rangreihungsaufgabe 1 und bei 16,8 Minuten ($s=3,8$) für die Aufgabe 2. Sie unterscheiden sich damit beträchtlich voneinander. Selbst unter Berücksichtigung der Tatsache, daß die Apn bei der zweiten Vollprofil-Aufgabe einen Vorteil an Erfahrung und an Übung

durch die erste Rangreihung und durch 120 ALASCA-Paarvergleiche mitbringen und diese damit sicherer und schneller erledigen können, deuten die Antwortzeiten darauf hin, daß die Apn die zweite Aufgabe mit einer niedrigeren Antwortmotivation durchführten. Dies könnte zu der geringfügigen Verringerung der Meßgüte des zweiten Vollprofil-PSM geführt haben.

Insgesamt lassen die Befunde erkennen, daß die induzierte Nutzenstruktur über die gesamte Untersuchungsphase hinweg stabil bleibt. Dies ist insofern wichtig, als verschiedene Leistungen bei den ALASCA-Aufgaben auch auf deren manipulierte Unterschiede und nicht auf Änderungen der Präferenzstrukturen der Apn rückführbar sind.

6.4.2.3 Retest-Reliabilität der Holdout-Aufgabe

Eine wichtige Eigenschaft unseres Forschungsdesigns ist die Beibehaltung der Vergleichbarkeit zu anderen Untersuchungen. Es verwendet weiterhin Holdout-Aufgaben, die im allgemeinen dazu herangezogen werden, die kriterienorientierte Validität der PSM zu bestimmen (vgl. Abschnitt 5.4). In Abschnitt 6.3.1 wurde die Gefährdung der Abschätzung der kriterienorientierten Validität durch das Problem der Kriterienreliabilität dargestellt. Eine Möglichkeit, mit diesem Problem umzugehen, besteht in der individuellen Bestimmung der Reliabilität des Kriteriums.

Die im letzten Abschnitt ermittelten Maßzahlen zur Prüfung der Stabilität der Präferenzstruktur lassen sich auch als Indikatoren der Reliabilität der Holdout-Aufgaben auffassen. Die Retest-Reliabilität der Holdout-Aufgabe läßt sich darüber hinaus auf der Ebene der PSM und auf der Ebene der Antwortvektoren hinsichtlich der fünf identischen Objekte abschätzen.

Als Übereinstimmungsmaß der PSM kann beispielsweise die Produkt-Moment-Korrelation der TNW individuell bestimmt werden (vgl. Abschnitt 5.5.2). Diese beträgt $\bar{r}_{\text{TNW1-TNW2}}=0,87$ (nach Fisher-Z Transformation der Korrelationskoeffizienten) mit einer Streubreite von $r=0,68$ bis $r=0,97$. Die Reliabilität ist damit im Durchschnitt hoch und über die Apn homogen.

Da fünf Objekte der Vollprofil-Objekte-Sets 1 und 2 identisch sind, läßt sich auf der Ebene des Antwortvektors deren Rangreihe miteinander vergleichen. Die mittlere Rangkorrelation bei diesen Objekten beträgt $\bar{r}_{\text{s-Set1-Set2}}=0,91$ (nach Fisher-Z Transformation der Rangkorrelationskoeffizienten) mit einer Streubreite

von $r_s=0,6$ bis $r_s=1,00$. Damit bestätigt die Ebene des Antwortvektors die Befunde auf der Ebene der PSM.

Alle Apn bearbeiteten die Holdout-Aufgaben nach derselben induzierten Nutzenstruktur. Nach den Überlegungen, wie sie durch Abbildung 5.2 in Abschnitt 5.4.1 zusammengefaßt werden, sollte die Reliabilität der Holdout-Rangreihe deswegen nicht sehr hoch zwischen den Apn variieren. Dies zeigen auch die Befunde. Verwendet man die kriterienorientierte Validität der ALASCA-PSM hinsichtlich der Holdout-Rangreihen, um zu beurteilen, welche Profilstärke der zu vergleichenden Objekte die besseren Ergebnisse erbringen, dann kann vorsichtshalber überprüft werden, ob die unterschiedlichen Validitätskoeffizienten nicht auch durch die Unterschiede in der Kriterienreliabilität erklärbar sind.

Zusammenfassend ist festzuhalten, daß die Befunde nicht nur auf die Stabilität der induzierten Nutzenstruktur, sondern auch auf eine durchweg hohe Reliabilität der Präferenzmessung durch die Holdout-Aufgaben hinweisen.

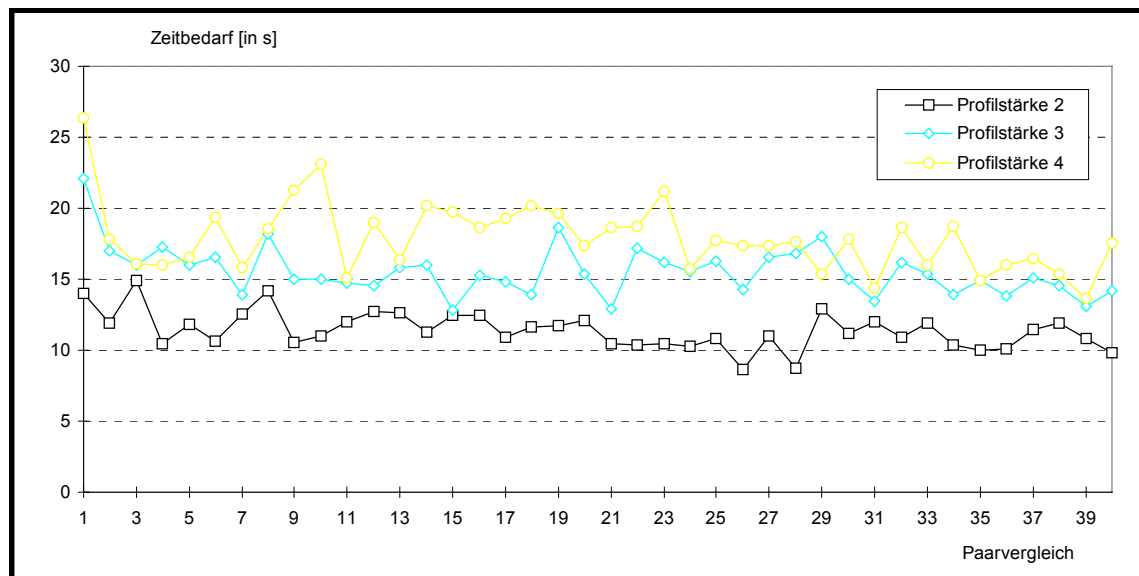
6.4.2.4 Beantwortungszeiten der ALASCA-Paarvergleiche

Bei den abgestuften ALASCA-Paarvergleichen erfaßten wir die Antwortzeiten. Die Beantwortung eines Paarvergleichs der Profilstärke 2 dauerte im Mittel 11,40 Sekunden, eines 3er-Paarvergleichs 15,56 Sekunden und eines 4er-Paarvergleichs 17,88 Sekunden. Dies entspricht einer mittleren Beantwortungsdauer der gesamten Paarvergleichsaufgabe von 7,60 Minuten ($s=1,81$) beim 2er-Paarvergleich, von 10,37 Minuten ($s=2,62$) beim 3er-Paarvergleich und von 11,92 Minuten ($s=2,86$) beim 4er-Paarvergleich. Der Unterschied zwischen 2er- und 3er-Paarvergleich und zwischen 3er- und 4er-Paarvergleich ist statistisch signifikant ($F_{1,23}=25,63$ $p<0,01$ bzw. $F_{1,23}=6,83$ $p<0,02$ mit der Profilstärke als Meßwiederholungsfaktor). Dieser Befund bestätigt Hypothese 3a.

Tabelle 6.3: Zeitbedarfsäquivalenzen von Paarvergleichsaufgaben der Profilstärke 2, 3 und 4.

Bezug	Anzahl	sind hinsichtlich des Zeitbedarfs äquivalent mit		
2er Paarvergleiche	10	10,00	7,33	6,38
3er Paarvergleiche	10	13,65	10,00	8,87
4er Paarvergleiche	10	15,69	11,49	10,00

2er Paarvergleichen 3er Paarvergleichen 4er Paarvergleichen

**Abbildung 6.14:** Die Beantwortungszeit [in s] der Paarvergleichsfragen gemittelt über alle Apn. Die Antwortzeiten sind getrennt für die Profilstärkebedingungen dargestellt.

Die 3er-Paarvergleichsaufgabe dauerte damit durchschnittlich das 1,36-fache der Zeit des 2er-Paarvergleichs. Für die 4er-Paarvergleichsaufgabe benötigten die Apn das 1,57-fache der Zeit, die sie für die 2er Paarvergleichsaufgabe verwendeten. In anderen Worten ausgedrückt heißt dies beispielsweise, daß die Beantwortungszeit von zehn 4er-Paarvergleiche mit 15,70 2er-Paarvergleichen äquivalent ist. Tabelle 6.3 veranschaulicht diese Zeitbedarfsäquivalenzen zwischen 2er-, 3er- und 4er-Paarvergleichen. Abbildung 6.14 zeigt den Verlauf der Antwortzeiten differenziert für die 40 Paarvergleiche und die Profilstärke.

6.4.2.5 Validität und Strukturwiedererkennung der Vollprofil-CA

Die Vollprofil-Objekte-Sets 1 und 2 kann man entweder als Vollprofil-CA oder als Holdout-Objekte-Set ansehen. Damit läßt sich aus den geschätzten PSM sowohl die Strukturwiedererkennung als auch die kriterienorientierte Validität hinsichtlich der jeweils anderen Holdout-Aufgabe bestimmen.

Ein Indikator der kriterienorientierten Validität auf individueller Ebene ist z.B. die Rangkorrelation der vorhergesagten mit der empirischen Rangreihe. Die durchschnittliche Rangkorrelation bei der Vorhersage von Holdout-Set 2 durch die Vollprofil-CA 1 beträgt $\bar{r}_{s\text{-PSM1-Set2}}=0,76$. Die Rangfolge des Holdout-Sets 1 durch die Vollprofil-CA 2 kann mit $\bar{r}_{s\text{-PSM2-Set1}}=0,76$ vorhergesagt werden (Mittelung jeweils nach Fisher-Z Transformation). Damit zeigt sich eine hohe kriterienorientierte Validität der Vollprofil-CA auf individueller Ebene. Die Validität der Aufgaben 1 und 2 unterscheidet sich nicht.

Die Wiedererkennungsleistung der wahren Nutzenstruktur durch die Vollprofil-CA 1 und 2 liegt bei $\overline{W(SSE)}_{\text{Set1-wahr}}=286,17$ ($s=60,28$) bzw. $\overline{W(SSE)}_{\text{Set2-wahr}}=284,39$ ($s=60,17$). Diese Wiedererkennungsleistung entspricht einer mittleren Korrelation zwischen den wahren und den geschätzten TNW von $\bar{r}_{\text{TNW1-TNWwahr}}=0,88$ bzw. $\bar{r}_{\text{TNW2-TNWwahr}}=0,88$. Das Meßgütekriterium der Strukturwiedererkennungsleistung führt zu denselben Schlüssen wie das der kriterienorientierten Validität. Beide Verfahren sind hinsichtlich der Meßgüte äquivalent. Die absolute Höhe der Meßgüteindikatoren kann mit den ALASCA-Paarvergleichs-CA (Abschnitte 6.4.2.6 und 6.4.2.7) verglichen werden.

6.4.2.6 Verlauf der Validität der ALASCA-Paarvergleichs-CA

Genauso wie man nach der Vollprofil-CA die kriterienorientierte Validität der PSM bezüglich einer Rangreihe von Holdout-Objekten bestimmen kann, ist dies auch mit den PSM der ALASCA-Paarvergleichs-CA möglich. Der einzige Unterschied besteht darin, daß die Paarvergleichs-CA eine Schätzung des PSM nach jedem Paarvergleich erlaubt. Bestimmt man die Validität der PSM nach jedem Paarvergleich, dann erhält man einen Verlauf der kriterienorientierten Validität auf individueller Ebene über die Anzahl der Paarvergleiche.

In Abbildung 6.15 ist dieser Validitätsverlauf als mittlere Rangkorrelation zwischen der vorhergesagten und der empirischen Rangreihe des Holdout-Sets 1 graphisch über die Paarvergleiche abgetragen. Die Darstellung ist für die

Paarvergleichsaufgaben der Profilstärken 2, 3 und 4 getrennt. Abbildung 6.16 veranschaulicht die kriterienorientierte Validität hinsichtlich des Holdout-Sets 2 (alle Mittelwertbildungen der Korrelationskoeffizienten nach Fisher-Z Transformation).

Die Rangkorrelationsmaßzahlen der kriterienorientierten Validität zeigen allgemein mit den ersten Paarvergleichen einen steilen Anstieg. Die Zunahme der Validität ist negativ beschleunigt. Ab ca. 20 Paarvergleichen wächst die Validität mit weiteren Paarvergleichen nur noch geringfügig.

Um Hypothesen des Validitätsverlaufs mit zunehmender Anzahl der Paarvergleiche prüfen zu können, wurden Indikatoren der Validität nach 10, 20, 30 und 40 Paarvergleichen gebildet. Um die Zufallsschwankungen dieser Validitätsmaße zu glätten, wurden jeweils fünf Rangkorrelationen innerhalb der Apn gemittelt (vgl. Abzissenbeschriftung der Abbildungen 6.17 und 6.18). Abbildung 6.17 zeigt die über alle Apn gemittelten Indikatoren der kriterienorientierten Validität hinsichtlich des Holdout-Sets 1, Abbildung 6.18 hinsichtlich des Holdout-Sets 2 (alle Mittelwertbildungen der Korrelationskoeffizienten nach Fisher-Z Transformation).

Die ausführlichen Ergebnisse der nachfolgenden Varianzanalysen sind tabellarisch in Anhang M dargestellt. Im Text sind nur die zentralen Kennwerte der statistischen Prüfung enthalten. Die Anwendungsvoraussetzungen der Varianzanalysen wurden geprüft. Es ergaben sich keine groben Verletzungen.

In einem ANOVA-Design mit den Meßwiederholungsfaktoren 'Holdout-Set', 'Anzahl der Paarvergleiche' und 'Profilstärke' (2X4X3; Tabelle L-1) und der Fisher-Z transformierten Rangkorrelation als abhängige Variable zeigt sich kein Haupteffekt auf dem Faktor 'Holdout-Set' ($F_{1,23}=0,05$ $p>0,8$). Dies bedeutet inhaltlich, daß die ALASCA-PSM hinsichtlich beider Holdout-Sets eine durchschnittliche Validität in vergleichbarer Höhe zeigen.

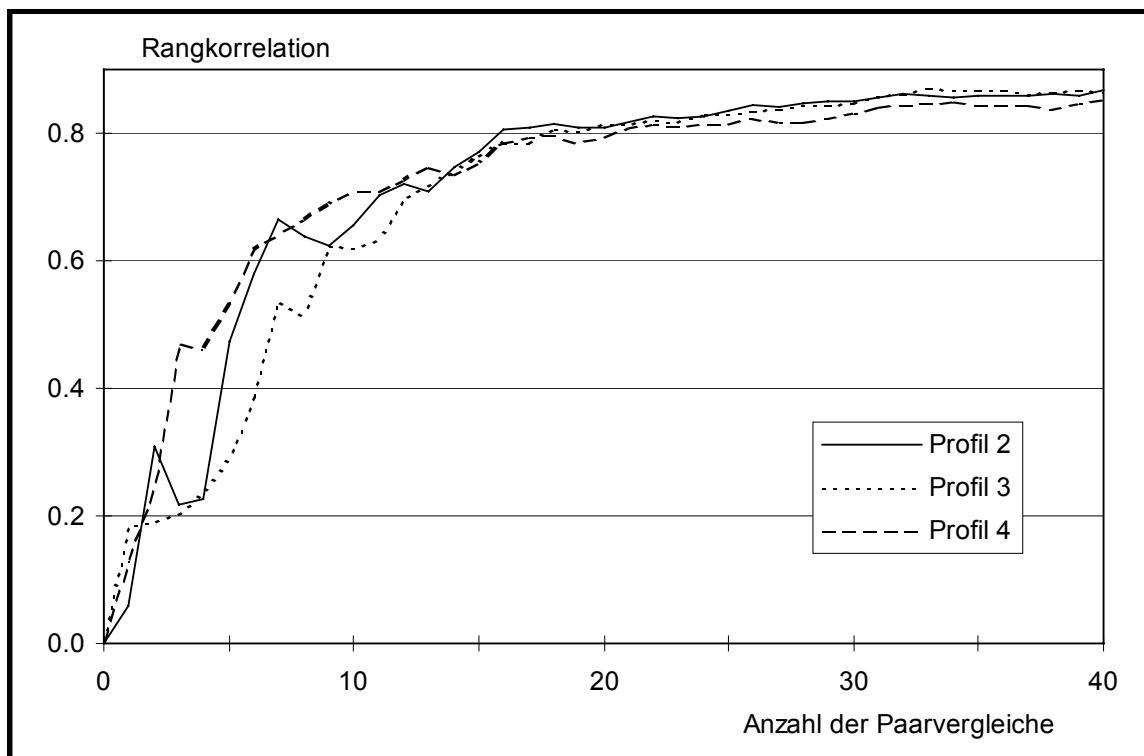


Abbildung 6.15: Kriterienorientierte Validität der ALASCA-PSM hinsichtlich des Holdout-Sets 1.

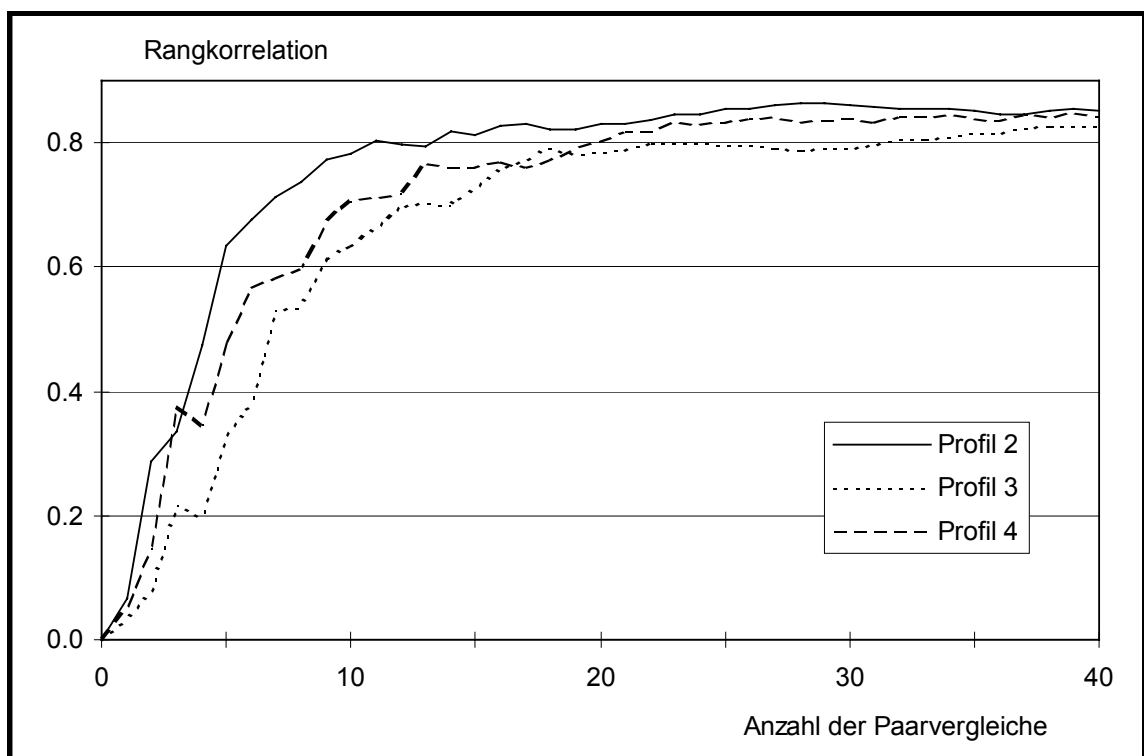


Abbildung 6.16: Kriterienorientierte Validität der ALASCA-PSM hinsichtlich des Holdout-Sets 2.

Die erwartete Zunahme der kriterienorientierten Validität mit der Anzahl der Paarvergleiche (Hypothese 1b) bestätigt sich durch den Haupteffekt des Faktors 'Anzahl der Paarvergleiche' ($F_{3,69}=93,3$ $p<0,01$).

Vergleicht man die Abbildungen 6.17 und 6.18, dann ist augenfällig, daß sich der Verlauf der Graphen für beide Holdout-Sets unterscheidet. Eine Varianzanalyse offenbart dies dadurch, daß alle Wechselwirkungen, an denen der Faktor 'Holdout-Set' beteiligt ist, statistisch signifikant sind (Holdout-Set X Anzahl der Paarvergleiche: $F_{3,69}=6,23$ $p<0,01$; Holdout-Set X Profilstärke: $F_{2,46}=4,0$ $p<0,05$; Holdout-Set X Profil X Anzahl der Paarvergleiche: $F_{6,138}=4,06$ $p<0,01$). Offensichtlich reagiert die kriterienorientierte Validität eines Nutzenschätzverfahrens sensibel auf das spezielle Holdout-Set, das als Kriterium Verwendung findet. Dieser Effekt ist seit längerem bekannt und führte uns zu dem speziellen Forschungsdesign (vgl. Abschnitte 5.4.2 und 6.3).

Betrachtet man die kriterienorientierte Validität getrennt für beide Holdout-Sets (4X3-ANOVA-Design mit den Meßwiederholungsfaktoren 'Anzahl der Paarvergleiche' und 'Profilstärke' und der Fisher-Z transformierten Rangkorrelation als abhängige Variable, Tabelle L-2), dann zeigt sich beim *Holdout-Set 1* der erwartete Haupteffekt für den Faktor 'Anzahl der Paarvergleiche' ($F_{3,69}=91,05$ $p<0,01$). Dieser Effekt ist auch bei allen Einzelvergleichen zwischen den Stufen des Faktors 'Anzahl der Paarvergleiche' zu finden (univariate F-Tests; alle $F_{1,23}>31,4$ und alle $p<0,01$ für die Einzelvergleiche 10-20, 20-30 und $F_{1,23}=5,29$ $p<0,05$ für den Einzelvergleich 30-40). Es besteht dagegen kein Effekt auf dem Faktor 'Profilstärke' ($F_{2,46}=0,15$ $p>0,85$).

Die gleichen Berechnungen lassen sich auch für die kriterienorientierte Validität hinsichtlich des *Holdout-Sets 2* durchführen (Tabelle L-3). Wieder ist der Haupteffekt für den Faktor 'Anzahl der Paarvergleiche' vorhanden ($F_{3,69}=47,56$ $p<0,01$). Auch hier ist der Effekt bei den Einzelvergleichen zwischen den Stufen des Faktors 'Anzahl der Paarvergleiche' zu finden (univariate F-Tests; alle $F_{1,23}>31,4$ $p<0,01$ für die Einzelvergleiche 10-20, 20-30 und $F_{1,23}=5,02$ $p<0,05$ für den Einzelvergleich 30-40). Hier liegen ähnliche Befunde vor wie beim Holdout-Set 1. Anders sieht es dagegen beim Faktor 'Profilstärke' aus. Im Gegensatz zum Holdout-Set 1 bewirkt dieser beim Holdout-Set 2 einen Haupteffekt ($F_{2,46}=5,33$ $p<0,01$). Der Effekt der Profilstärke ist jedoch nur für den

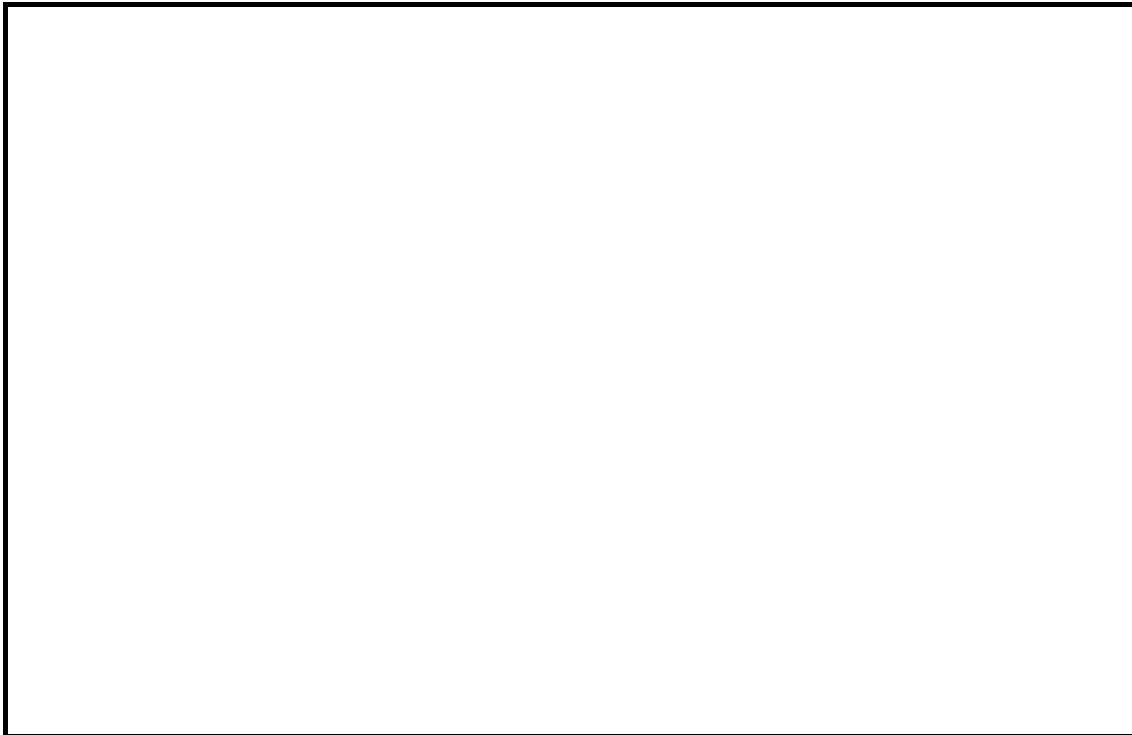


Abbildung 6.17: Kriterienorientierte Validität der ALASCA-PSM nach 10, 20, 30 und 40 Paarvergleichen hinsichtlich des Holdout-Sets 1 (über je 5 Paarvergleiche geglättet).

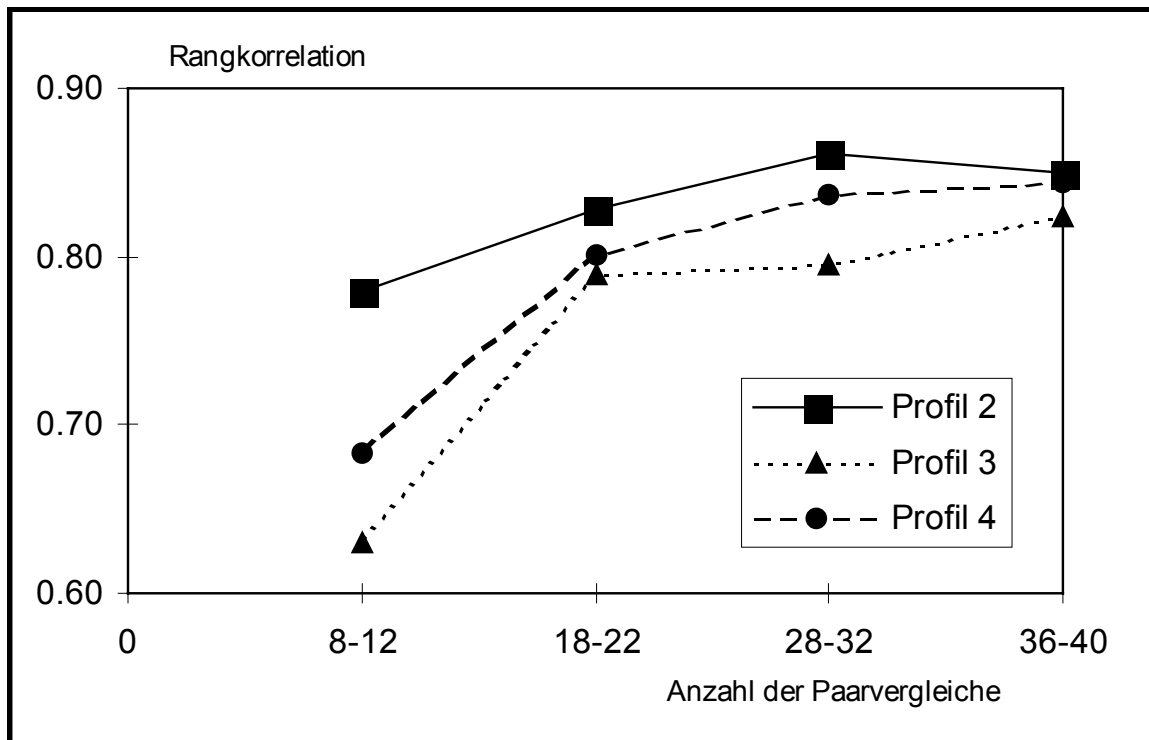


Abbildung 6.18: Kriterienorientierte Validität der ALASCA-PSM nach 10, 20, 30 und 40 Paarvergleichen hinsichtlich des Holdout-Sets 2 (über je 5 Paarvergleiche geglättet).

größten Unterschied (zwischen der Profilstärke 2 und 3) vorhanden (2er- vs. 3er-Profil: $F_{1,23}=9,43$ $p<0,01$). Die Unterschiede zwischen den anderen Profilen sind nicht signifikant. (2er- vs. 4er Profil: $F_{1,23}=2,71$ $p>0,12$; 3er- vs. 4er Profil: $F_{1,23}=1,78$ $p>0,18$).

Die Befunde zur Frage, welche Objektprofilstärke optimal ist, und wie viele Paarvergleiche verwendet werden sollten, sind verwirrend. Erstens widersprechen sich die Befunde, was den Einfluß der Objektprofilstärke der ALASCA-Paarvergleiche auf die kriterienorientierte Validität betrifft. Die Profilstärke zeigt keinen Einfluß, wenn das Holdout-Set 1 als Validitätskriterium dient. Nimmt man aber das Holdout-Set 2, dann unterscheidet sich die Validität der ALASCA-PSM, die durch Paarvergleiche der Profilstärke 2 und 3 gebildet wurden. Ob die Profilstärke 2 eine unterschiedliche Validität gegenüber der Profilstärke 4 erbringt, muß offen bleiben. Somit sind die Befunde keinesfalls klar interpretierbar.

Ist diese uneindeutige Befundlage dieser Untersuchung ein Einzelfall? Wir behaupten nein. Bei unserer Untersuchung tritt diese Problematik lediglich dadurch deutlich zutage, weil wir mit zwei Holdout-Sets arbeiten. Hätten wir nur ein Holdout-Set verwendet, dann wären die Ergebnisse scheinbar eindeutig. Daß sie tatsächlich sehr sensibel auf das spezielle Holdout-Set reagieren, läßt sich aber nur aufgrund der Kenntnis der Literatur vermuten.

Die Hypothese 1a behauptet, daß die Validität der PSM von der Profilstärke abhängt. Sie kann mit der kriterienorientierten Validität als abhängige Variable nur teilweise bestätigt werden⁷.

Wodurch kommen diese unklaren Ergebnisse zustande? Wird die kriterienorientierte Validität wieder von der nicht perfekten Reliabilität des Kriteriums beeinflusst? Tabelle 6.4 zeigt die Korrelation zwischen Indikatoren der Reliabilität und der kriterienorientierten Validität hinsichtlich der Holdout-Sets 1 und 2.

Als Indikatoren für die Reliabilität dienen hier $r_{s-Set1-wahr}$ bzw. $r_{s-Set2-wahr}$ (Rangkorrelation der gelegten Rangreihe 1 bzw. 2 mit der wahren Rangreihe; s. Abschnitt 6.4.2.2; Fisher-Z transformiert), $W(SSE)_{Set1-wahr}$ bzw. $W(SSE)_{Set2-wahr}$ (Wiedererkennung der wahren Nutzenstruktur durch Vollprofil-CA 1 bzw. 2, s. Abschnitt 6.4.2.2) und $r_{TNW1-TNW2}$ (Korrelation der TNW der Vollprofil CA 1 und 2; s. Abschnitt 6.4.2.3; Fisher-Z transformiert).

Es zeigt sich ein ausgeprägter Zusammenhang zwischen der Kriterienreliabilität und der kriterienorientierten Validität. Mit einer höheren Kriterienreliabilität ist auch eine höhere Validität der ALASCA-PSM zu finden. Die Effekte sind bei den Reliabilitätsindikatoren, die die Holdout-Rangreihe direkt mit wahren Werten vergleichen, besonders ausgeprägt und lassen sich mittels der Test-Retest-Reliabilität ($r_{TNW1-TNW2}$) nur z.T. nachweisen.

⁷ Würde man auch den unterschiedlichen Zeitbedarf für Paarvergleiche unterschiedlicher Profilstärke berücksichtigen, wäre die Befundlage klarer. In Abschnitt 6.5.2.9 wird dieser Gedanke aufgegriffen.

Tabelle 6.4: Korrelation zwischen Indikatoren der kriterienorientierten Validität hinsichtlich der Holdout-Sets 1 und 2 und der Reliabilität des Kriteriums (Erläuterung der Reliabilitätsindikatoren im Text).

Korrelationskoeffizienten	Kriterienorientierte Validität zu							
	Holdout-Set 1				Holdout-Set 2			
	nach 10	nach 20	nach 30	nach 40	nach 10	nach 20	nach 30	nach 40
Reliabilitätsindikator	Paarvergleichen der Profilstärke 2							
$r_{s\text{-Set1/2-wahr}}$.46*	.53**	.49*	.56**	.74**	.61**	.64**	.67**
$W(\text{SSE})_{\text{Set1/2-wahr}}$	-.45*	-.46*	-.39	-.42*	-.83**	-.74**	-.78**	-.79**
$r_{\text{TNW1-TNW2}}$.12	.35	.47*	.47*	.16	.07	.14	.23
Reliabilitätsindikator	Paarvergleichen der Profilstärke 3							
$r_{s\text{-Set1/2-wahr}}$.29	.44*	.57**	.63**	.59**	.76**	.83**	.83**
$W(\text{SSE})_{\text{Set1/2-wahr}}$	-.25	-.41*	-.54**	-.61**	-.63**	-.82**	-.85**	-.86**
$r_{\text{TNW1-TNW2}}$.24	.45*	.36	.34	.25	.25	.24	.20
Reliabilitätsindikator	Paarvergleichen der Profilstärke 4							
$r_{s\text{-Set1/2-wahr}}$.06	.44*	.50*	.33	.09	.66**	.59**	.56**
$W(\text{SSE})_{\text{Set1/2-wahr}}$	-.04	-.47*	-.46*	-.34	-.03	-.61**	-.59**	-.55**
$r_{\text{TNW1-TNW2}}$.36	.42*	.50*	.37	.05	.21	.02	.21

* $p < 0,05$ ** $p < 0,01$; zweiseitiger Test

Wir möchten darauf hinweisen, daß der Zusammenhang zwischen der kriterienorientierten Validität und der Reliabilität des Kriteriums in dieser Untersuchung auftritt, obwohl in die Apn dieselbe Nutzenstruktur induziert wurde. Dies sollte zu einer Homogenisierung der Kriterienreliabilität führen, weil die Apn Holdout-Sets mit einer subjektiv gleichen Schwierigkeit erhalten. Bei CA, die individuelle Nutzenstrukturen modellieren, sollte der Einfluß der Kriterien-

reliabilität daher noch stärker vorhanden sein (vgl. die Ausführungen zum Reliabilitätsproblem der Kriterienaufgabe in Abschnitt 5.4.1).

Zusammenfassend läßt sich folgendes sagen: Steht bei Experimenten zu CA-Verfahren nur die kriterienorientierte Validität der verschiedenen PSM zu Holdout-Objekten als Entscheidungskriterium darüber zur Verfügung, ob sich die CA-Verfahren unterscheiden, dann treten häufig Interpretationsprobleme auf, weil die kriterienorientierte Validität sensibel auf das spezielle Holdout-Set reagiert. Insbesondere die nicht perfekte Reliabilität der Kriteriumsmessung ist hierfür verantwortlich. Diese Schwierigkeit haben wir antizipiert. Im Gegensatz zu den üblichen Forschungsansätzen auf dem Gebiet der CA besitzt unser Forschungsdesign neben der kriterienorientierten Validität noch das Meßgütekriterium der Wiedererkennungseistung der wahren Nutzenstruktur. Im nächsten Abschnitt möchten wir die Befunde mit diesem Kriterium darstellen.

6.4.2.7 Verlauf der Strukturwiedererkennung der ALASCA-Paarvergleichs-CA

ALASCA schätzt nach jedem Paarvergleich ein PSM. Die Güte, mit der diese PSM die wahre Nutzenstruktur wiedererkennen, sollte von der Anzahl der Paarvergleiche und der Objektprofilstärke abhängen (Hypothesen 1a und 1b). Als Maßzahl der Strukturwiedererkennungseistung dient $W(SSE)$ nach Normierung des wahren PSM und der geschätzten Modelle (vgl. Anhang C).

Abbildung 6.19 zeigt den Verlauf der Strukturwiedererkennung in Abhängigkeit von der Objektprofilstärke und der Anzahl der Paarvergleiche über alle A_{pn} gemittelt. Wie erwartet zeigt sich eine Verminderung von $W(SSE)$ mit der Anzahl der Paarvergleiche. Analog zur kriterienorientierten Validität (vgl. Abbildungen 6.15 und 6.16) nimmt die Strukturwiedererkennungseistung mit den ersten Paarvergleichen stark zu. Die Zunahme ist negativ beschleunigt und flacht nach 20 Paarvergleichen stark ab.

Zur Prüfung verschiedener Hypothesen zur Profilstärke und zur Anzahl der Paarvergleiche wurden (analog zur Vorgehensweise in Abschnitt 6.4.2.6) Indikatoren der Strukturwiedererkennung nach 10, 20, 30 und 40 Paarvergleichen gebildet (Abbildung 6.20). Jeweils fünf $W(SSE)$ -Maßzahlen wurden hierzu wieder innerhalb der A_{pn} gemittelt, um Zufallsschwankungen der Maßzahlen zu glätten (vgl. Abzissenbeschriftung der Abbildung 6.20).

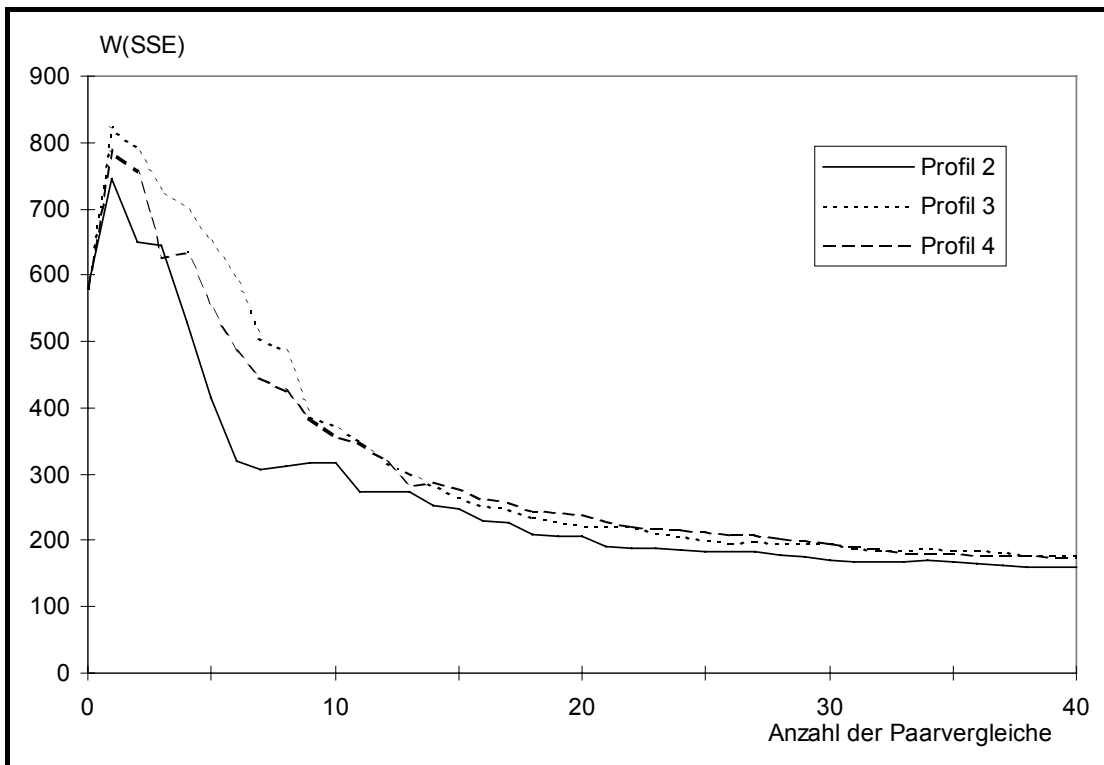


Abbildung 6.19: Die durchschnittliche Strukturwiedererkennungsleistung der geschätzten PSM für die Profilstärken 2, 3 und 4 in Abhängigkeit der Anzahl der Paarvergleiche.

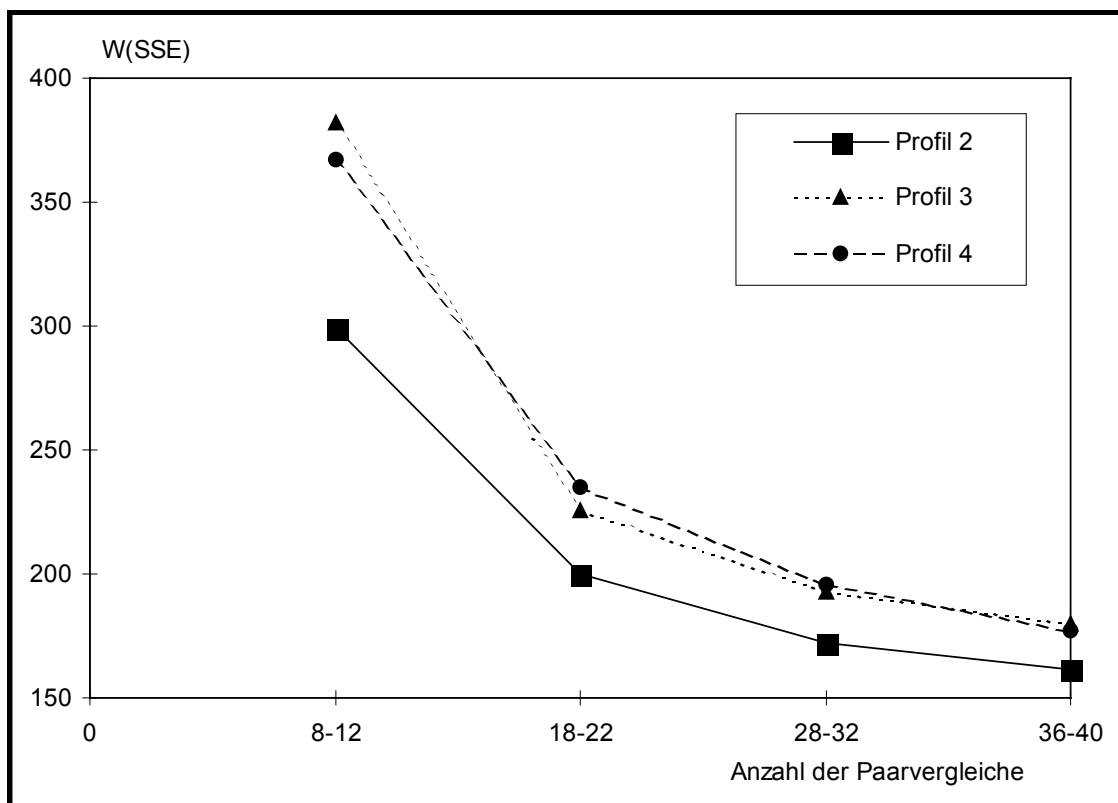


Abbildung 6.20: Strukturwiedererkennung der ALASCA-PSM nach 10, 20, 30 und 40 Paarvergleichen (über je 5 Paarvergleiche geglättet).

Eine (4X3)-ANOVA mit den Meßwiederholungsfaktoren 'Anzahl der Paarvergleiche' und 'Profilstärke' (s. Tabelle L-4) zeigt einen starken Haupteffekt auf dem Faktor 'Anzahl der Paarvergleiche' ($F_{3,69}=214$ $p<0,01$). Dieser bleibt bei allen Einzelvergleichen vorhanden (univariate F-Tests; $F_{1,23}>27,0$ $p<0,01$ für alle Einzelvergleiche der Stufen des Faktors 'Anzahl der Paarvergleiche').

Der Faktor 'Profilstärke' hat einen Haupteffekt auf die Wiedererkennung der wahren Nutzenstruktur ($F_{2,46}=3,93$ $p<0,05$). Dieser rührt vom Unterschied zwischen der Profilstärke 2 und den beiden anderen Profilstärkebedingungen her (Einzelvergleiche der 2er-/3er-Profile $F_{1,23}=13,56$ $p<0,01$ und 2er-/4er-Profile $F_{1,23}=4,21$ $p<0,06$; univariate F-Tests). Die ALASCA-Paarvergleiche der Profilstärken 3 und 4 unterscheiden sich nicht in ihrer Strukturwiedererkennung (Einzelvergleich der 3er-/4er-Profile $F_{1,23}=0,01$ $p>0,9$).

Eine statistisch signifikante Wechselwirkung zwischen den Faktoren 'Profilstärke' und 'Anzahl der Paarvergleiche' bedeutet inhaltlich, daß der Verlauf der Strukturwiedererkennungsleistung mit der Anzahl der Paarvergleiche für die unterschiedlichen Profilstärkebedingungen verschieden ist ($F_{6,138}=4,03$ $p<0,01$). Insbesondere zu Beginn ist die Wiedererkennungsleistung der 2er-Paarvergleiche sehr hoch. Der Wechselwirkungseffekt der Faktoren 'Profilstärke' X 'Anzahl der Paarvergleiche' zeigt sich bei (3X2)-ANOVA's nur zwischen den 'Anzahl der Paarvergleiche'-Stufen 10-20 ($F_{2,46}=5,27$ $p<0,01$). Zwischen allen anderen Stufen des Faktors 'Anzahl der Paarvergleiche' ist er nicht vorhanden ($F_{2,46}<0,51$ $p>0,6$). Abbildung 6.20 ist zu entnehmen, daß der Effekt dadurch zustande kommt, daß die Paarvergleiche der Profilstärke 2 sehr früh zu einer hohen Meßgüte führen.

Welcher Unterschied ergibt sich, wenn man als Indikator der Meßgüte des ALASCA-Verfahrens die Wiedererkennung einer wahren Nutzenstruktur statt der kriterienorientierten Validität hinsichtlich eines Holdout-Objekte-Sets benutzt?

Die Effekte des Faktors 'Anzahl der Paarvergleiche' sind mit beiden Meßgütekriterien erkennbar. Ganz anders verhält es sich jedoch mit den Effekten des Faktors 'Profilstärke'. Nimmt man die kriterienorientierte Validität als abhängige Variable, dann existiert ein hoch signifikanter Effekt beim Holdout-Set 2, während der Effekt beim Holdout-Set 1 fehlt. Die Strukturwiedererkennungsbefunde bestätigen wiederum unsere theoretischen Erwartungen, daß die Profilstärke einen Einfluß auf die Meßgüte der PSM ausübt. Verwendet man Indikatoren der Validität, dann hängen die Befunde vom speziellen Holdout-

Objekte-Set ab, hinsichtlich dessen die kriterienorientierte Validität ermittelt wird. Dieser schwerwiegende Nachteil des Maßes zeigt deutlich die Überlegenheit der Strukturwiedererkennungsleistung bei Forschungsansätzen.

Was läßt sich über die Befunde der Strukturwiedererkennungsleistung zusammenfassend sagen? Zum einen nimmt die Wiedererkennung der wahren Nutzenstruktur mit der Anzahl der Paarvergleiche zu. Dieser Effekt war zu erwarten (Hypothese 1b), da mehr Paarvergleiche eine effektivere Schätzung der Parameter erlauben. Die Wiedererkennungsleistung verläuft negativ beschleunigt.

Des weiteren erhalten wir eine Überlegenheit von Objekten der Profilstärke 2 bei abgestuften Paarvergleichen gegenüber den Profilstärken 3 und 4. Dieser Befund zeigt sich auch bei den Arbeiten anderer Autoren (vgl. Abschnitt 3.5.3.3). Obwohl die stärkeren Profile eine effektivere Parameterschätzung zulassen, sind diese der Profilstärke 2 hinsichtlich der Meßgüte des geschätzten Modells unterlegen. Offensichtlich sind geringere Objektprofilstärken psychologisch angemessener als starke Profile. Damit werden die statistischen Nachteile geringer Profilstärken durch ihre psychologischen Vorteile mehr als ausgeglichen. Der Erklärung dieses Ergebnisses wendet sich der folgende Abschnitt 6.4.2.8 zu.

6.4.2.8 *Bestimmung des Antwortfehlers*

Warum lassen sich PSM durch Paarvergleiche der Profilstärke 2 mit einer höheren Meßgüte schätzen als durch stärkere Objektprofile? Um die psychologische Wirkung des Nutzenschätzverfahrens auf die Apn faßbar zu machen, haben wir bei den theoretischen Überlegungen des Abschnitts 6.2 das Konzept 'Antwortfehler' herangezogen. Dieser Abschnitt befaßt sich mit der Bestimmung eines Indikators des Antwortfehlers bei den Paarvergleichen.

Hierzu wurden die wahren Antworten, also die Antworten, die durch das wahre PSM determiniert sind, mit den tatsächlichen Antworten der Apn verglichen. Tabelle 6.5 zeigt die Produkt-Moment-Korrelation zwischen den wahren und den empirischen Antworten getrennt nach den Profilstärkebedingungen und über alle Apn gemittelt (Korrelationen Fisher-Z transformiert - gemittelt - rücktransformiert). Neben dem Korrelationskoeffizienten r ist auch r^2 in der Tabelle dargestellt. Diese Maßzahl läßt sich als der Anteil der Varianz interpretieren, den

die tatsächlichen Antworten mit der Varianz der wahren Nutzendifferenzen gemeinsam haben. Ein hohes r bzw. r^2 deutet auf einen niedrigen Antwortfehler der Apn hin.

In Tabelle 6.5 zeigt auch die Produkt-Moment-Korrelation zwischen den wahren und den tatsächlichen Antworten beim 1.-10., 11.-20., 21.-30. und 31.-40. Paarvergleich. Daraus läßt sich der Verlauf des Antwortfehlers über die Anzahl der Paarvergleiche ablesen.

Beim ALASCA-Design werden den Apn nur relativ nutzenähnliche Paare vorgegeben (vgl. Abschnitt 6.1). Da die Nutzenbalancierung auf der Grundlage des jeweils aktuell geschätzten PSM vorgenommen wird, kann es (z.B. bei einem noch sehr ungenauen PSM) vorkommen, daß die wahre Antwort außerhalb der möglichen Antwortskala [-4; +4] liegt. Dadurch würde die Produkt-Moment-Korrelation zwischen der wahren und der tatsächlichen Antwort verringert. Um sicherzugehen, daß unsere Befunde nicht durch diesen Umstand verzerrt sind, berechneten wir zusätzlich r und r^2 nur bei den Paarvergleichsantworten, in denen die wahre Antwort innerhalb des Intervalls [-2; +2] liegt ('korrigiert für Randeffekte').

Tabelle 6.5: Die Produkt-Moment-Korrelation r und r^2 zwischen der wahren und der empirischen Antwort auf die abgestuften Paarvergleiche. Die Maßzahlen sind getrennt nach Objektprofilstärke (2, 3, 4) und nach der Anzahl der Paarvergleiche (bis 10, 11-20, 21-30, 31-40) dargestellt.

Paarvergleiche	Produkt-Moment-Korrelation r			erklärte Varianz r^2		
	Profilstärke	Profilstärke	Profilstärke	Profilstärke	Profilstärke	Profilstärke
0-10	0,88	0,86	0,88	0,77	0,74	0,77
11-20	0,88	0,85	0,81	0,77	0,72	0,65
21-30	0,91	0,72	0,74	0,82	0,51	0,54
31-40	0,79	0,73	0,70	0,62	0,54	0,49
alle	0,85	0,79	0,77	0,73	0,62	0,59
alle (korrigiert für Rande ffek te)	0,73	0,60	0,49	0,53	0,35	0,24

Es zeigt sich, daß die Antworten der Apn sehr hoch mit den wahren Antworten korrelieren, der Antwortfehler somit gering ist. Es bestätigt sich Hypothese 3c, daß der Antwortfehler mit der Objektprofilstärke zunimmt. Insbesondere Objektprofile der Stärke 2 gehen mit einem deutlich niedrigeren Antwortfehler einher als stärkere Profile. Dieser Befund ist sowohl ohne als auch mit Korrektur für Randeffekte zu finden (bei letzterer Berechnung liegen die Korrelationskoeffizienten etwas niedriger, was auf Varianzeinschränkung zurückzuführen ist).

Durch die Betrachtungen der Antwortfehler sind mehrere neue Einblicke möglich. Erstens wird der Befund erklärt, warum die Schätzung von PSM auf der Basis von Paarvergleichen der Profilstärke 2 gegenüber stärkeren Profilen überlegen ist. Paarvergleiche mit geringer Profilstärke führen zu einer weniger fehlerbehafteten Antwort. Im Zusammenhang mit Hypothese 3c wurde dieser Effekt so erklärt, daß geringere Profilstärken weniger kognitive Ressourcen benötigen (diese Erklärung wird in Abschnitt 6.4.2.11 überprüft). Die statistischen Nachteile stärkerer Profile werden somit durch ihre psychologischen Vorteile mehr als kompensiert.

Beim Vergleich der Profilstärken 3 und 4 wurde kein Unterschied hinsichtlich der Meßgüte ihrer PSM festgestellt (Abschnitt 6.4.2.7). Offensichtlich werden die statistischen Vorteile, die das 4er-Profil bei der Parameterschätzung besitzt, durch den geringeren Antwortfehler des 3er-Profils gerade ausgeglichen. PSM, die durch Paarvergleiche der Profilstärke 3 und 4 geschätzt wurden, unterscheiden sich daher in ihrer Meßgüte kaum.

Begutachtet man den Verlauf des Antwortfehlerindicators über die 40 Paarvergleiche, dann wird deutlich, daß der Antwortfehler tendenziell gegen Ende der Paarvergleichsaufgabe zunimmt. Dieser Befund ist konsistent mit den Annahmen, die bei den theoretischen Betrachtungen in Abschnitt 6.2 getroffen wurden (vgl. Abbildung 6.5). Der Effekt wurde mit der Abnahme der Antwortmotivation im Verlauf der Paarvergleichsaufgaben theoretisch begründet.

6.4.2.9 Validität und Strukturwiedererkennung in Relation zur Antwortzeit

Zur Lösung der Forschungsfrage, welche Profilstärke als optimal zu betrachten ist, wurden PSM durch Paarvergleiche verschiedener Profilstärke geschätzt und deren Meßgüte miteinander verglichen. Dabei diente als Vergleichspunkt die

Anzahl der Paarvergleiche. Beispielsweise haben wir gefragt: Welches PSM zeigt eine höhere Meßgüte nach 10 Paarvergleichen? Das PSM, das durch Paarvergleiche der Profilstärke 2 geschätzt wurde, oder das PSM, das 3er-Profile verwendete? Den Zeitbedarf für die Paarvergleichsaufgaben haben wir bisher völlig außer acht gelassen.

Dies hätte keinerlei Konsequenzen, wenn die verschiedenen Aufgaben gleich lange dauern würden. In Abschnitt 6.4.2.4 wurde jedoch deutlich, daß dies nicht der Fall ist (vgl. Tabelle 6.3 und Abbildung 6.14). Die Beantwortung von Paarvergleichen der Profilstärke 4 dauert etwa 1,57mal so lange wie die Beantwortung von 2er-Profilen. 3er-Profile benötigen das 1,36-fache der Zeit, welche die Apn zur Bearbeitung der 2er-Profile verwenden.

In vielen Fällen dürfte nicht interessieren, ob es - hinsichtlich der Meßgüte der resultierenden PSM - sinnvoller ist, die Apn um 20 Paarvergleiche der Profilstärke 2, 3 oder 4 zu bitten. Vielmehr steht zur Debatte, welche Profilstärke man bei fester Interviewzeit wählen soll. Wieder wollen wir, wie bereits in den Abschnitten 6.4.2.6 und 6.4.2.7, die Meßgüte der PSM unter den verschiedenen Profilstärkebedingungen betrachten. Zuvor soll jedoch ein profilspezifischer Korrekturfaktor eingeführt werden, der die unterschiedlichen Bearbeitungszeiten berücksichtigt. Hierzu bestimmen wir aus den Ergebnissen des Abschnitts 6.4.2.4 die Anzahl der Paarvergleiche einer Profilstärke, die eine durchschnittliche Ap in einer bestimmten Zeiteinheit erwartungsgemäß beantwortet. Tabelle 6.6 zeigt die gerundeten Ergebnisse. Danach kann man die mittlere Validität bzw. die mittlere Strukturwiedererkennungsleistung über der Zeit abtragen. Die Abbildung 6.21 veranschaulicht die Strukturwiedererkennungsleistung der PSM für die Profilstärkebedingungen (2, 3 und 4) in Abhängigkeit der Zeit⁸. Die Maßzahlen sind wieder über fünf Meßzeitpunkte gemittelt, um Zufallsschwankungen der Meßgütekriterien zu glätten (vgl. rechte Seite von Tabelle 6.6).

⁸ Mit der kriterienorientierten Validität kann analog verfahren werden. Die Ergebnisse sind analog der Strukturwiedererkennungsleistung und sollen hier nicht dargestellt werden.

Tabelle 6.6: Die erwartete Anzahl der Paarvergleiche, die in einer bestimmten Zeiteinheit beantwortet werden. Die Maßzahlen der Meßgüte werden innerhalb der Apn über jeweils 5 Paarvergleiche gemittelt.

Dauer [Minuten]	Erwartete Anzahl der beantworteten Paarvergleiche			Innerhalb der Apn gemittelt über die Paarvergleiche		
	Profilstärke 2	Profilstärke 3	Profilstärke 4	Profilstärke 2	Profilstärke 3	Profilstärke 4
1	5	4	3	3-7	2-6	1-5
2	11	8	7	9-13	6-10	5-9
3	16	12	10	14-18	10-14	8-12
4	21	15	13	19-23	13-17	11-15
5	26	19	17	24-28	17-21	15-19
6	32	23	20	30-34	21-25	18-22
7	37	27	23	35-39	25-29	21-25
8		31	27		29-33	25-29
9		35	30		33-37	28-32
10			34			32-36
11			37			35-39

Eine (3X7)-ANOVA (s. Tabelle L-5) mit den Meßwiederholungsfaktoren 'Profilstärke' und 'Meßzeitpunkt' (nach 1, 2, ..., 7 Minuten) und dem Strukturwiedererkennungsindikator $W(SSE)$ als abhängige Variable belegt einen starken Haupteffekt auf dem Faktor 'Meßzeitpunkt' ($F_{6,138}=721$ $p<0,01$). Dieser korrespondiert mit dem Haupteffekt des Faktors 'Anzahl der Paarvergleiche', der in Abschnitt 6.4.2.7 ohne Berücksichtigung der Antwortzeit beschrieben ist.

Auch der Haupteffekt des Faktors 'Profilstärke' ist wieder deutlich zu finden ($F_{2,46}=30,4$ $p<0,01$). Er rührt vom Unterschied zwischen dem 2er-Profil und den anderen beiden Profilstärkebedingungen her (univariate F-Tests: Einzelvergleiche der 2er-/3er-Profile $F_{1,23}=98,2$ $p<0,01$ und 2er-/4er-Profile $F_{1,23}=40,64$ $p<0,01$). Dagegen gibt es keinen Unterschied in der Strukturwiedererkennungsleistung der ALASCA-Paarvergleiche der Profilstärken 3 und 4 ($F_{1,23}=0,76$ $p>0,39$).

Wie bereits ohne Berücksichtigung der Antwortzeit feststellbar war, gibt es auch hier eine Wechselwirkung zwischen der 'Profilstärke' und dem 'Meßzeitpunkt' ($F_{12,276}=23,0$ $p<0,01$). Sie tritt auf, weil Paarvergleiche der Profilstärke 2 bereits nach kurzer Zeit zu einer hohen Meßgüte führen und der nachfolgende Meßgütezuwachs geringer ausfällt als bei den anderen Profilen.

Die Effekte unter Berücksichtigung der Beantwortungszeiten bestätigen

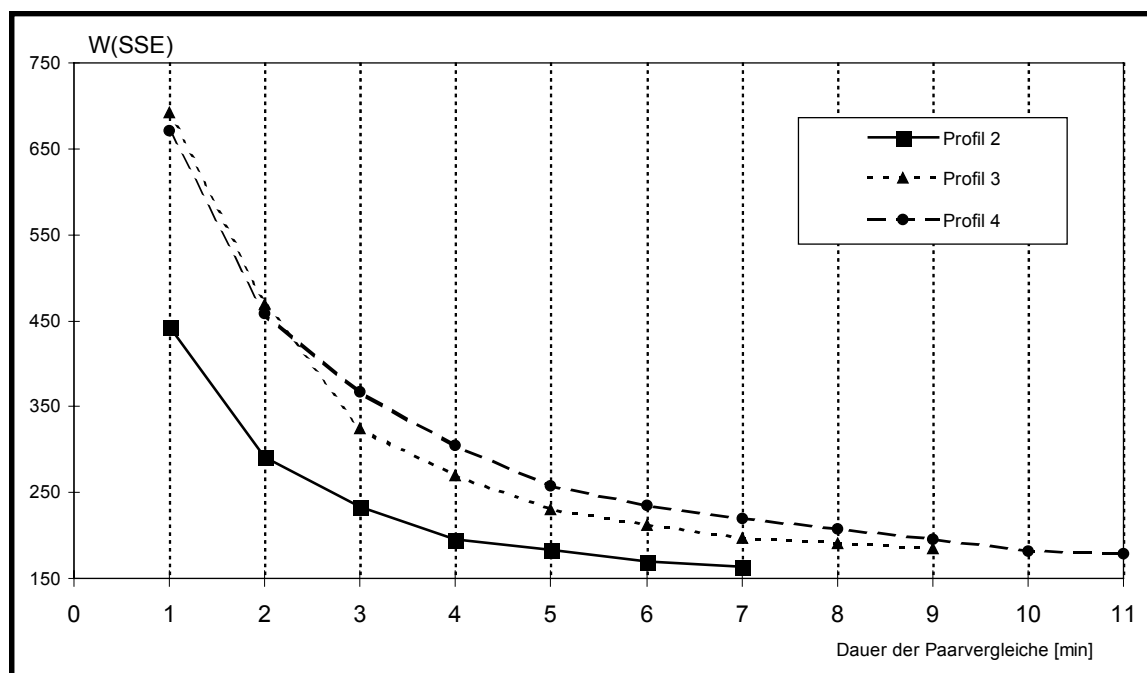


Abbildung 6.21: Strukturwiedererkennung der ALASCA-PSM abgetragen über der durchschnittlichen Antwortzeit (über je 5 Paarvergleiche geglättet).

damit diejenigen Befunde, die keine Antwortzeiten beachten (vgl. Abschnitt 6.4.2.7). Sie fallen zum Teil sogar noch deutlicher aus, weil die überlegenen 2er-Paarvergleiche gleichzeitig die kürzeste Antwortdauer benötigen. Um die Strukturwiedererkennungsleistung eines PSM zu erreichen, das durch ALASCA-2er-Paarvergleiche innerhalb von drei Minuten gebildet wurde, ist im Durchschnitt bei 3er-Paarvergleichen ein ALASCA-Interview von ca. fünf Minuten bzw. bei 4er-Paaren von ca. sechs Minuten erforderlich. Gemäß den Befunden dieser Untersuchung kann man als Faustregel aufstellen, daß man bei Paarvergleichen der Profilstärke 2 etwas mehr als die Hälfte der Zeit benötigt als bei Verwendung von 3er-Paaren (bzw. die Hälfte der Zeit bei 4er-Paaren). Diese Befunde sprechen eindeutig für Paarvergleiche der Profilstärke 2.

6.4.2.10 Vergleich der Meßgüte von Vollprofil-CA und ALASCA-Paarvergleichen

In Abschnitt 6.4.2.5 wurde die Strukturwiedererkennung der Vollprofil-CA 1 und 2 und ihre kriterienorientierte Validität hinsichtlich des jeweils anderen Holdout-Objekte-Sets dargestellt. Diese Maßzahlen lassen sich direkt mit den Meßgüteindikatoren der ALASCA-Paarvergleichs-PSM vergleichen. Diese sind in den Abschnitten 6.4.2.6 und 6.4.2.7 dargestellt. Darüber hinaus kann man auch die CA-Verfahren hinsichtlich ihres Zeitbedarfs vergleichen (Abschnitt 6.4.2.4).

Die mittlere kriterienorientierte Validität der beiden Vollprofil-CA (Vorhersage der Rangfolge der Objekte des jeweils anderen Holdout-Sets) beträgt $\bar{r}_{s-\text{Vollprofil-Set}1/2} = 0,76$. Diesen Kennwert überschreiten die ALASCA-PSM durchschnittlich mit dem 16. 2er-Paarvergleich (bzw. mit dem 18. 3er- und 4er-Paarvergleich). Verwendet man als Indikator der Meßgüte die Wiedererkennungslleistung der wahren Nutzenstruktur, dann 'überholt' die ALASCA-Methode die Vollprofil-CA noch früher. Die ALASCA-Paarvergleiche unterbieten die durchschnittliche Strukturwiedererkennungslleistung der Vollprofil-CA von $\overline{W(SSE)}_{\text{Vollprofil-wahr}} = 285,28$ mit dem 11. 2er-Paarvergleich (bzw. mit dem 14. 3er- und dem 15. 4er-Paarvergleich). Selbst bei einer für die Vollprofil-CA sehr günstigen Einschätzung der Verfahren kommt man zu dem Schluß, daß 18 ALASCA-Paarvergleiche die Schätzung eines PSM zulassen, das durchschnittlich eine ähnliche Meßgüte besitzt wie ein PSM, das aus der Rangreihung von 15 Vollprofil-Objekten geschätzt wurde.

Begutachtet man den Zeitbedarf, den die Apn durchschnittlich für die Vollprofil-Aufgaben und die ALASCA-Aufgaben benötigen, dann ergibt sich ein grobes Mißverhältnis: Die mittleren Beantwortungszeiten zur Rangreihung der 15 Vollprofile liegt bei 27,8 Minuten ($s=3,7$) für die Holdout-Aufgabe 1 und bei 16,8 Minuten ($s=3,8$) für die Aufgabe 2. Dies sind im Mittel 22,3 Minuten. Für die ersten 18 ALASCA-Paarvergleiche einer beliebigen Profilstärke, die hinsichtlich der Meßgüte einer Vollprofil-CA äquivalent sind, verwenden die Apn durchschnittlich 4,68 Minuten. Trotz konservativer Annahmen benötigen die Apn damit für die Vollprofil-CA das 4,8-fache der Zeitdauer, die sie für die ALASCA-Paarvergleiche aufwenden, damit aus den Daten PSM derselben Meßgüte geschätzt werden können.

Die Apn benötigen für die Vollprofil-CA mehr als das Doppelte der Zeit, die sie für alle 40 Paarvergleiche verwenden (durchschnittlich 9,96 Minuten). Bei der Würdigung dieses Effekts ist zu beachten, daß die Apn per Instruktion bei allen Aufgaben auf die gleiche Zeitdauer eingestellt waren (vgl. Abschnitt 6.4.1.7).

Angesichts dieses eindeutigen Befundes muß man sich fragen, warum bisherige Untersuchungen, die einen Vergleich zwischen der Vollprofil-CA und ACA anstellten, keine oder nur geringe Unterschiede zwischen den Verfahrensmöglichkeiten feststellten (vgl. Abschnitt 3.7.3.2). Entweder muß man auch hierfür wieder das Problem der Kriterienreliabilität verantwortlich machen oder unsere Untersuchung muß durch ihren Ansatz 'Induktion einer wahren Nutzenstruktur' angezweifelt werden. Sollte letzteres notwendig sein, dann ist jedoch immer noch nicht klar, warum unsere Lernphase die ALASCA-Aufgabe mehr fördern sollte als die Vollprofil-CA. Immerhin sind die Übungen der dritten Stufe der Lernphase ('vergleichende Beurteilung', s. Abbildung 6.10) und die vier Lernerfolgsüberprüfungen der Vollprofil-Aufgabe sehr ähnlich. Aus diesem Grunde sollte die gewählte Induktionsmethode eher dazu führen, die Vollprofil-CA zu unterstützen als die ALASCA-Paarvergleichs-CA.

Nimmt man die kriterienorientierte Validität gegenüber einem Holdout-Set als Indikator der Meßgüte, dann kann auch Kriterienkontamination durch die Methodenähnlichkeit von Prädiktor- und Kriteriumsaufgabe nicht zur Erklärung unserer Befunde herangezogen werden. Die Kriteriumsaufgabe ist der Vollprofil-CA sehr ähnlich und der ALASCA-Aufgabe vergleichsweise unähnlich. Ein Artefakt durch Methodenähnlichkeit des Kriteriums sollte eher eine Über-

schätzung der Vollprofil-CA zur Folge haben, kann aber unsere Befunde nicht erklären.

Zusammenfassend möchte wir festhalten, daß der Befund der deutlichen Überlegenheit der Paarvergleichs-CA gegenüber der Vollprofil-CA unerwartet ist. Die in der Literatur berichtete Befundlage anderer Untersuchungen werden hier nicht bestätigt. Weitere Forschung sollte sich darum bemühen, eine Erklärung für diese Divergenz zu finden.

6.4.2.11 Wirkung der Profilstärke auf die kognitiven Ressourcen

Nach jedem ALASCA-Durchgang beantworteten die Apn einen Fragebogen (vgl. Abschnitt 6.4.1.7). Die Items 1, 6 und 10 des Fragebogens (Einschätzung der Schwierigkeit der Paarvergleichsaufgaben; Einschätzung, wie schwierig es fiel, sich zu konzentrieren; Einschätzung, wie schwierig es fiel, eine Antwort zu finden; s. Tabelle I-1 im Anhang) sollten dabei Indikatoren dafür sein, wie sehr die kognitiven Ressourcen der Apn durch die Schwierigkeit der Paarvergleiche ausgelastet sind.

Hypothese 3b behauptet, daß stärkere Profile mehr kognitive Ressourcen in Anspruch nehmen. In Tabelle 6.7 sind die Mittelwerte der Indikatoren dieser kognitiven Ressourcen dargestellt. Die Ergebnisse zeigen, daß die Beantwortung stärkerer Profile als schwieriger (Item 1) beurteilt werden, daß es den Apn schwerer fiel, sich zu konzentrieren (Item 6) und eine Antwort zu finden (Item 10). Diese Ergebnisse werden auch durch die Rangfolgen der Items 20, 22 und 23 bestätigt⁹.

⁹ Diese Ergebnisse sollen hier aus Platzgründen nicht dargestellt werden. Sie sind jedoch dahingehend eindeutig, als die Apn glauben, die meisten kognitiven Ressourcen für die 4er-Paare und die wenigsten für die 2er-Paare zu benötigen.

Tabelle 6.7: Mittelwerte der Items 1, 6, 10 und 2.

	2er-Profil	3er-Profil	4er-Profil	Prüfung der Unterschiedshypothese	
Item 1	3,17	2,33	1,71	$F_{2,46}=39,5$	$p<0,01$
Item 6	2,88	2,17	1,79	$F_{2,46}=24,8$	$p<0,01$
Item 10	3,08	2,38	2,17	$F_{2,46}=22,0$	$p<0,01$

Item 1: Die Paarvergleichsaufgaben waren ... schwierig - einfach.

Item 6: Wie schwer ist es Dir gefallen, Dich zu konzentrieren? Schwer - leicht.

Item 10: Eine Antwort zu finden fiel mir insgesamt ... schwer - leicht.

Wie beurteilen die Apn subjektiv ihre Leistung in den ALASCA-Paarvergleichen? Tabelle 6.8 zeigt die mittleren Antworten des Items 18 ('Wie beurteilst Du Deine Leistung bei den Paarvergleichen, gemessen an Deiner Leistung in der letzten 7-Karten-Rangordnungsaufgabe in der Lernphase'? Besser - schlechter) und des Items 19 ('Wie viele falsche Antworten schätzt Du gegeben zu haben?' weniger als 2 - mehr als 7). Die Apn sind der Meinung, daß ihre Leistung mit der Profilstärke abfällt und ihre fehlerhaften Antworten zunehmen. Auch diese subjektiven Reaktionen sprechen, neben den Ergebnissen zur kriterienorientierten Validität und Strukturwiedererkennung, eindeutig für die Verwendung von Objektprofilen der Stärke 2.

Tabelle 6.8: Mittelwerte der Items 18 und 19.

	2er-Profil	3er-Profil	4er-Profil	Prüfung der Unterschiedshypothese	
Item 18	2,75	3,67	4,17	$F_{2,46}=28,7$	$p<0,01$
Item 19	2,38	2,83	3,13	$F_{2,46}=20,0$	$p<0,01$

Item 18: Wie beurteilst Du Deine Leistung bei den Paarvergleichen, gemessen an Deiner Leistung in der letzten 7-Karten-Rangordnungsaufgabe in der Lernphase? Besser - schlechter

Item 19: Wie viele falsche Antworten schätzt Du gegeben zu haben? weniger als 2 - mehr als 7

Zusammenfassend läßt sich festhalten, daß stärkere Objektprofile mehr kognitive Ressourcen von den Apn fordern. Auf der Ebene der subjektiven Reaktionen sprechen sich die Apn eindeutig für geringe Profilstärken aus. Sie

glauben damit bessere Urteile abgeben zu können. Da dies objektiv zutrifft, ist die Einschätzung der Apn realistisch.

6.4.2.12 Auswirkungen der kognitiven Leistungsfähigkeit und der Leistungsmotivation

Die theoretischen Überlegungen in Abschnitt 6.2 weisen den Faktoren 'Auskunftsmotivation' und 'Auskunftsfähigkeit' bei der Bearbeitung der multiattributiven Paarvergleiche eine Schlüsselrolle zu. Die Auskunftsfähigkeit wird bei dieser Untersuchung mit den kognitiven Ressourcen gleichgesetzt, die zur Bearbeitung der Urteilsaufgabe einsetzbar sind.

In Hypothese 6 wird behauptet, daß die Meßgüte der PSM mit der kognitiven Leistungsfähigkeit und der Leistungsmotivation der Apn zunimmt. Um dies zu überprüfen, erhoben wir als Indikatoren der kognitiven Leistungsfähigkeit den Testwert auf der LPS-Skala 12 und die Abiturnote. Als Indikator der Leistungsmotivation ermittelten wir den Testwert der MARPS (s. Abschnitt 6.4.1.7).

Zunächst die Befunde zur kognitiven Leistungsfähigkeit: Eine (4X3)-ANCOVA mit den Meßwiederholungsfaktoren 'Anzahl der Paarvergleiche' (10-20-30-40) und 'Profilstärke' (2-3-4), der Kovariate LPS-Wert und der abhängigen Variablen Strukturwiedererkennungsleistung $W(SSE)$ (jeweils über fünf Paarvergleiche innerhalb der Apn geglättet; vgl. die (4X3)-ANOVA des Abschnitts 6.4.2.7 bzw. Tabelle L-4) zeigt keinen Effekt des LPS-Wertes ($F_{1,22}=0,34$ $p>0,55$). Ebenfalls keinen Effekt bewirkt die Kovariate 'Abiturnote' mit derselben ANCOVA ($F_{1,21}=1,19$ $p>0,25$). Die Ergebnisse sind ausführlich in den Tabellen L-6 und L-7 im Anhang dargestellt.

In einem multiplen Regressionsmodell mit den Prädiktoren 'LPS-Wert' und 'Abiturnote' und dem Kriterium 'Strukturwiedererkennungsleistung' (Wiedererkennungsmaßzahlen $W(SSE)$ der fünf letzten Paarvergleiche innerhalb der Apn geglättet; vgl. Abbildung 6.20) trägt der LPS-Wert in allen Profilstärkebedingungen nichts zur Erklärung des Kriteriums bei (alle $t_{2,20}<0,51$ $p>0,6$). Die Abiturnote allein kann dagegen die Strukturwiedererkennungsleistung bei den 2er-Profilen mit $r=0,39$ ($F_{1,21}=3,87$ $p<0,07$) und bei den 3er-Profilen mit $r=0,36$ ($F_{1,21}=3,17$ $p<0,09$) einigermaßen gut vorhersagen. Bei den 4er-Profilen ist dieser Zusammenhang jedoch nur schwach ($r=0,19$ $F_{1,21}=0,75$ $p>0,40$).

Bei der Interpretation der Irrtumswahrscheinlichkeiten sollte man beachten, daß die zentralen Effekte dieser Untersuchung durch einen Meßwiederholungsversuchsplan nachgewiesen werden sollten. Daher sind die Stichprobengrößen bei diesen 'between subject designs' nicht sehr groß. Darüber hinaus ist die Varianz der kognitiven Leistungsfähigkeit durch die studentische Apn-Stichprobe erheblich eingeschränkt.

Um die Vorhersagen bezüglich der Leistungsmotivation zu überprüfen, wurde eine (4X3)-ANCOVA mit den Meßwiederholungsfaktoren 'Anzahl der Paarvergleiche' (10-20-30-40) und 'Profilstärke' (2-3-4), der Kovariate MARPS-Wert und der abhängigen Variablen Strukturwiedererkennungsleistung $W(SSE)$ (jeweils über 5 Paarvergleiche innerhalb der Apn geglättet; vgl. die (4X3)-ANOVA des Abschnitts 6.4.2.7 bzw. Tabelle L-4) durchgeführt. Es zeigt sich kein Effekt des MARPS-Wertes ($F_{1,22}=0,04$ $p>0,82$). Die Ergebnisse sind ausführlich in Tabelle L-8 im Anhang dargestellt.

Bei der Regression der Kriterien 'Strukturwiedererkennungsleistung nach 10-20-30 oder 40 Paarvergleichen' (Wiedererkennungsmaßzahlen $W(SSE)$ jeweils über 5 Paarvergleiche innerhalb der Apn geglättet; vgl. Abbildung 6.20) auf die MARPS-Werte tragen diese in allen Profilstärkebedingungen nichts zur Erklärung des Kriteriums bei (alle $r>-0,27$ $F_{1,22}<1,66$ $p>0,21$). Dennoch ist anzumerken, daß die Vorzeichen der Korrelationskoeffizienten die erwartete Richtung anzeigen.

Im Zusammenhang mit der Leistungsmotivation lassen sich keine statistisch abgesicherten Effekte erkennen. Auch hier ist jedoch die Stichprobengröße durch die 'between subject designs' gering. Die Varianzeinschränkung durch die spezielle Stichprobe dürfte ebenfalls eine Rolle spielen.

Zusammenfassend kann gesagt werden, daß die Rolle der kognitiven Leistungsfähigkeit und der Leistungsmotivation bei der Bearbeitung von multiattributiven Paarvergleichen durch diese Untersuchung nicht geklärt werden konnte. Die vermuteten Zusammenhänge (Hypothese 6) konnten nicht statistisch abgesichert bestätigt werden. Dennoch deuten alle Zusammenhänge in die vermutete Richtung.

Die Indikatoren der kognitiven Leistungsfähigkeit und der Leistungsmotivation wurden bei dieser Untersuchung nur 'sicherheitshalber' erhoben. Da sie keine der zentralen Fragestellungen berühren, und die Untersuchung ohnehin sehr lange dauerte, wurde auf ihre Messung nur wenig Zeit verwandt. Die

Befunde deuten jedoch an, daß besser kontrollierte Untersuchungen hier fündig werden könnten.

6.5 Implikationen der Befunde und Diskussion

Das Anliegen dieses Abschnitts besteht darin, zunächst die zentralen Ergebnisse der empirischen Untersuchung II in einer übersichtlichen Form darzustellen. Hierbei wollen wir insbesondere auch auf die Implikationen der Befunde für praktische Belange in der Forschung und Anwendung eingehen. Im zweiten Teil soll die Aussagekraft und die Reichweite der Untersuchungsergebnisse kritisch diskutiert werden. Der Abschnitt wird durch einen Ausblick abgeschlossen.

6.5.1 Die zentralen Ergebnisse

Um die Frage nach der optimalen Profilstärke einer effektiven empirischen Bearbeitung zugänglich zu machen, suchten wir einen Weg, mit dem Problem der Kriterienreliabilität umzugehen. Dies war notwendig, weil die Effekte der mangelnden Kriterienreliabilität sich in unserer empirischen Untersuchung I (CA zu Prüfungsordnungen) als sehr eindrucksvoll erwiesen haben. Auch die Befunde zum Lerncheckset D in der empirischen Untersuchung II bestätigen ebenfalls die großen Effekte, die unterschiedliche Schwierigkeiten von Holdout-Aufgaben bewirken (Abschnitt 6.4.2.1).

Eine niedrige Kriterienreliabilität führt zu folgendem Problem: Vergleicht man die kriterienorientierte Validität verschiedener CA-Verfahren, dann sind die Ergebnisse oft uneindeutig. Auch in dieser Untersuchung *zeigt sich bei Verwendung einer Maßzahl der kriterienorientierten Validität als abhängige Variable, daß sich Effekte verwischen bzw. die Effekte vom speziellen Holdout-Set abhängen* (Abschnitt 6.4.2.6). Nach unserer Überzeugung ist dies der Grund, warum bisherige Forschungsbemühungen zu oben genannten Fragen bei sehr vagen Aussagen stehenbleiben mußten (unseren eigenen Voruntersuchungen blieb dieses Schicksal auch nicht erspart).

Bei der Würdigung dieses Ergebnisses darf nicht vergessen werden, daß in unserer Untersuchung dieser Effekt auftritt, obwohl a) alle Apn nach derselben Präferenzstruktur urteilten, und wir b) einen Meßwiederholungsver-

suchsplan verwendeten. Dies führt dazu, daß a) die subjektive Schwierigkeit der Kriteriumsaufgabe zwischen den Apn sehr ähnlich ist und die Reliabilität des Kriteriums damit zwischen den Apn ebenfalls relativ ähnlich sein sollte, und daß b) die PSM der Untersuchungsbedingungen ihre Validität gegenüber demselben Kriterium, also auch gegenüber einem Kriterium mit identischer Reliabilität, erbringen mußten. Unter normalen Umständen ist mindestens der Punkt a) nicht gegeben: Aufgrund der individuell verschiedenen Präferenzstrukturen präsentiert sich die objektiv gleiche Kriteriumsaufgabe den Apn unterschiedlich schwierig. Als Beleg für die Richtigkeit dieser Denkweise haben wir den Befund unserer CA zu Prüfungsordnungen angeführt (Abschnitt 6.3.2). Ein Index für die individuelle Schwierigkeit der Kriteriumsaufgabe korrelierte hoch mit der kriterienorientierten Validität der geschätzten PSM.

Das Resultat dieser Bemühungen ist unser *spezielles Forschungsdesign* (Abschnitt 6.3.4). Es arbeitet mit der Induktion einer wahren Nutzenstruktur. Diese Vorgehensweise muß sicherlich immer kritisch daraufhin überprüft werden, ob sie mit üblichen CA-Interviews vergleichbar ist (externe Validität). In jedem Falle aber überwindet das Forschungsdesign viele Probleme üblicher Untersuchungsansätze zur Erforschung der CA.

Durch die Kenntnis der wahren Nutzenstruktur ist ein neues Meßgütekriterium verfügbar: Die Strukturwiedererkennungsleistung der geschätzten PSM. Vergleicht man die PSM, die in den Untersuchungsbedingungen 'Profilstärke 2', 'Profilstärke 3' und 'Profilstärke 4' geschätzt wurden, dann ergibt sich ein eindeutiger Befund: *Durch abgestufte Paarvergleiche der Profilstärke 2 sind PSM besser schätzbar als bei stärkeren Profilen* (Abschnitt 6.4.2.7). Wichtig dabei ist, daß dieser Befund nicht mit den Befunden in Widerspruch steht, die mit der abhängigen Variablen 'kriterienorientierte Validität' operierten. Bei Verwendung der Strukturwiedererkennungsleistung zeigen sich lediglich deutliche Effekte, die mit der abhängigen Variablen 'kriterienorientierte Validität' z.T. verdeckt geblieben waren. Verwendet man die Validität zu Holdout-Aufgaben als abhängige Variable, dann führt dies zu einer Steigerung des Fehlers 2. Art.

Bei abgestuften, multiattributiven Paarvergleichen bewirken geringere Profilstärken eine deutliche Abnahme der Antwortzeit. Da die empirische Untersuchung gezeigt hat, daß 2er-Profile gegenüber stärkeren Profilen ohnehin zu PSM höherer Meßgüte führen, verstärkt die Berücksichtigung der Antwortzeit

die Befunde noch: *Bei konstanter Interviewzeit zeigen die PSM, die mittels abgestufter Paarvergleiche der Objektprofilstärke 2 gebildet werden, eine deutlich höhere Meßgüte als solche, die durch Paarvergleiche stärkerer Profile geschätzt werden* (Abschnitt 6.4.2.9).

Das spezielle Forschungsdesign erlaubt den Vergleich der tatsächlichen Paarvergleichsantworten mit den Antworten, die durch die wahre Nutzenstruktur determiniert sind. Dadurch enden die Forschungsbemühungen nicht mit der bloßen Feststellung der oben genannten Effekte. Vielmehr können die Befunde über das Konzept des Antwortfehlers erklärt werden. *Die Antworten der Apn enthalten bei 2er-Profilen einen geringeren Antwortfehler als diejenigen Antworten, die bei 3er- und 4er-Profilen abgegeben werden* (Abschnitt 6.4.2.8). Dieser Befund wurde erwartet und durch die begrenzte Informationsverarbeitungskapazität im menschlichen Arbeitsgedächtnis erklärt (Abschnitt 6.2).

Neben der Meßgüte und der Antwortzeit *sprechen auch die subjektiven Reaktionen der Apn für die Verwendung von Beurteilungsobjekten der Profilstärke 2*. Die Apn nehmen diese Aufgaben als weniger schwierig wahr, sie behaupten, daß es ihnen leichter fällt, sich zu konzentrieren und es fällt ihnen leichter, eine Antwort zu finden (Abschnitt 6.4.2.11).

Der Vergleich der drei ALASCA-PSM mit den beiden PSM, die durch eine Vollprofil-CA gebildet wurden, ergibt ein in seiner Höhe unerwartetes Ergebnis: Die Meßgüte der Vollprofil-PSM wird von den ALASCA-PSM schon nach wenigen Paarvergleichen übertroffen. Je nach dem, welchen Indikator der Meßgüte man zugrunde legt und mit welcher Profilstärke ALASCA operiert, geschieht dies nach 11 bis 18 Paarvergleichen. Ein durchschnittliches Vollprofil-CA-Interview dauert 4,8mal so lange wie ein durchschnittliches ALASCA-Interview für 18 Paarvergleiche. *Um PSM der gleichen Meßgüte zu erhalten, dauert das Interview einer Vollprofil-CA sehr viel länger als das Interview einer CA mittels ALASCA* (Abschnitt 6.4.2.10).

6.5.2 Aussagekraft und Reichweite der Befunde

Inwieweit kann man die Befunde unserer Untersuchung generalisieren? Wodurch wird ihre Aussagekraft eingeschränkt? Unser spezieller Forschungsansatz ist, neben seinen vielfältigen Vorzügen, sicherlich auch sehr kritisch daraufhin zu überprüfen, welche externe Validität die Befunde aufweisen. Diese

wird insbesondere durch die Induktion einer wahren Nutzenstruktur bedroht. Vor der eigentlichen Untersuchung mußten die Apn diese Nutzenstruktur lernen. Dabei überprüften wir den Lernerfolg mehrfach, was zu einer starken Bewertungserwartungshaltung der Apn geführt haben könnte. Die ALASCA-Interviews in der nachfolgenden Untersuchung könnten für die Apn die Bedeutung einer herausfordernden Situation angenommen haben, in der das mühsam angeeignete 'Wissen' unter Beweis gestellt werden sollte. Aus diesem Grunde ist fraglich, ob die Auskunftsmotivation der Apn mit derjenigen vergleichbar ist, die bei einer anwendungsorientierten CA (z.B. im Rahmen einer Marktforschungsstudie) angetroffen wird.

Darüber hinaus hat die ausführliche Induktionsphase sicherlich zur Folge, daß den Apn der Prozeß der Bildung menschlicher Präferenzen sehr plastisch und ausführlich vor Augen geführt wurde. Unklar ist, ob diese Sachlage mit einer normalen CA-Untersuchungssituation vergleichbar ist. Informellen Gesprächen mit den Apn war zu entnehmen, daß diese die Interviewaufgaben keineswegs als unnatürlich ansahen. Es könnte eventuell empirisch geklärt werden, ob der Forschungsansatz 'Induktion einer wahren Nutzenstruktur' mit einer 'normalen' Modellierung der individuellen Präferenzstruktur durch eine CA vergleichbar ist. Allerdings hätte eine solche Studie den Nachteil, daß sie mit dem Problem der Kriterienreliabilität zu kämpfen hätte.

Die Apn beschäftigten sich den ganzen Untersuchungstag mit den Wohnungspräferenzen der fiktiven Person (Helga K.). Sie können deshalb wohl mit Recht als Experten für den Objektbereich 'Studentenappartements' und für die Präferenzstruktur der fiktiven Person bezeichnet werden. Das Expertentum ist unproblematisch, weil auch eine anwendungsorientierte CA besonderen Wert darauf legen muß, Experten für den jeweiligen Objektbereich als Apn zu gewinnen. Eine mögliche Schwierigkeit bei der Generalisierung unserer Befunde könnte darin gründen, daß die Apn nicht ihre eigenen Urteile, sondern die Urteile einer fremden Person simulieren. Wir sehen jedoch keinen theoretischen Grund für die Annahme, daß hierbei andere Gesetzmäßigkeiten wirksam sein sollten.

Diese Untersuchung verwendet eine spezielle Objektklasse, eine spezielle wahre Nutzenstruktur (mit fünf dreifach gestuften Attributen). Es stellt sich die Frage, ob die zentralen Befunde der Untersuchung II auch bei Variation dieser Untersuchungseigenschaften zu finden sind. Prinzipiell läßt sich nur durch weitere empirische Bemühungen eine Antwort auf diese Frage finden. Aus

unserer Sicht gibt es keine theoretischen Gründe, warum die Replikation der zentralen Ergebnisse nicht gelingen sollte.

Die Daten stammen von einer speziellen studentischen Apn-Stichprobe. Die intellektuelle Leistungsfähigkeit und Auskunftsmotivation der Apn ist sicherlich - verglichen mit einer typischen CA-Ap - über dem Durchschnitt. Die gesteigerte Auskunftsmotivation ist notwendig, damit eine Untersuchung nach dem speziellen Forschungsdesign überhaupt durchführbar ist. Aber auch diese spezifischen Eigenarten unserer Untersuchung bedrohen vermutlich nicht die erfolgreiche Replikation. Es ist vielmehr sogar zu vermuten, daß die Vorteile von 2er-Profilen bei weniger intelligenten und motivierten Apn noch deutlicher ausfallen (vgl. theoretische Überlegungen in Abschnitt 6.2).

Diese Untersuchung konzentrierte sich auf Fragestellungen, die im Zusammenhang mit der Optimierung der CA bei der Verwendung von abgestuften Paarvergleichen auftreten. Die Auswirkungen unterschiedlich guter Initialschätzungen wurden nicht besprochen. Wir erwarten, daß der Vorteil von 2er-Profilen mit der Güte etwaiger Initialschätzungen zunehmen wird. Dies kann damit begründet werden, daß mit der zunehmenden Güte einer Initialschätzung niedrigere Antwortfehler bei den Paarvergleichen notwendig werden, damit eine direkte Verbesserung der Meßgüte des PSM auftritt. Der empirische Beleg dieser Vorhersage muß jedoch noch erbracht werden.

Zusammenfassend wollen wir festhalten, daß unsere Befunde zur optimalen Objektprofilstärke den Ergebnissen der veröffentlichten Untersuchungen zu dieser Frage nicht widersprechen (vgl. Abschnitte 3.5.3.3 und 3.7.2.4). Dies spricht in jedem Falle für die Vergleichbarkeit unseres speziellen Forschungsdesigns mit anderen Untersuchungsansätzen. Diese Untersuchung findet dagegen Effekte, die bei bisherigen Forschungsarbeiten nicht belegt werden konnten. Daß diese Effekte bisher nicht aufgedeckt wurden, liegt unserer Meinung nach im Problem der Kriterienreliabilität begründet. Sie wären bei dieser Untersuchung, zumindest teilweise, auch nicht gefunden worden, wenn nur die abhängige Variable 'kriterienorientierte Validität' zur Verfügung gestanden hätte.

6.5.3 Ausblick

Diese Untersuchung kommt zu klaren Ergebnissen, soweit die Frage nach der Profilstärke der Beurteilungsobjekte bei abgestuften Paarvergleichen berührt wird. Damit ist das erste Ziel der Untersuchung, die weitere Optimierung von CA-Verfahren, erreicht. Die damit zusammenhängende Frage nach der optimalen Anzahl an Paarvergleichen kann man dagegen nicht allgemeingültig beantworten. Diese Anzahl hängt von vielen Faktoren einer CA-Untersuchung ab. Insbesondere sind die Auskunftsfähigkeit und -motivation der Apn und die Güte einer Initialschätzung zu nennen. In unserer Untersuchung beispielsweise sind 40 oder noch mehr Paarvergleiche anzuraten, weil die daraus geschätzten PSM immer noch einen Zuwachs an Meßgüte verzeichnen. Dies liegt daran, daß wir keine Initialschätzung vornahmen und mit Apn arbeiteten, die eine untypisch hohe Auskunftsfähigkeit und -motivation hatten. Unter anderen Rahmenbedingungen muß jedoch zu einer evtl. völlig anderen Paarvergleichszahl geraten werden. Ist die Initialschätzung bereits sehr gut und ist damit zu rechnen, daß die Apn eine nur geringe Auskunftsmotivation besitzen, dann kann sogar indiziert sein, daß gar keine Paarvergleiche durchgeführt werden sollten, weil sie die Meßgüte der PSM nur verschlechtern würden.

Wir haben ein Forschungsdesign entwickelt, mit dem Fragestellungen zur Optimierung von CA effektiv bearbeitet werden können. Dies war das zweite Ziel dieser Arbeit. Viele bisherige Ansätze zur Untersuchung der CA erbringen wenig eindeutige Ergebnisse. Die Klarheit der Ergebnisse dieser Untersuchung ist dem speziellen Forschungsdesign zu verdanken. Es kann in der vorliegenden Form zur Bearbeitung ähnlich gelagerter Fragestellungen nur empfohlen werden¹⁰.

Diese Untersuchung bediente sich zur Vorhersage und Erklärung der zentralen Befunde des Konzeptes 'Antwortfehler'. Zukünftige Untersuchungen könnten sich mit weiteren Faktoren beschäftigen, die diesen Antwortfehler be-

¹⁰ Aufgrund unserer Erfahrung würden wir in Zukunft drei kleinere Merkmale des Forschungsdesigns ändern. Erstens kann die Messung der Intelligenz und der Leistungsmotivation entfallen. Die Untersuchungsdauer ist auch ohne die beiden Instrumente an der oberen Grenze dessen, was den Apn zugemutet werden kann. Zweitens würden wir auf die Manipulation der Schwierigkeit bei den Lernchecksets verzichten und deren Anzahl auf drei verringern. Die Apn reagieren auf die Schwierigkeitsmanipulation in der Lernphase z.T. sehr irritiert. Sollen Hypothesen im Zusammenhang mit dem Problem der Kriterienreliabilität überprüft werden, empfiehlt sich eine eigens dafür angelegte Untersuchung. Drittens würden wir in Zukunft die Reihenfolge der Vollprofil-CA 1 und Vollprofil-CA 2 balancieren. Es hat sich gezeigt, daß die Apn auf die zweite Vollprofil-CA wesentlich weniger Zeit verwenden. Bei ihr handelt es sich um die letzte Untersuchungsaufgabe. Es ist somit nur verständlich, daß die Apn nach fast acht Stunden Untersuchungsdauer nicht mehr die gleiche Mühe auf die Aufgabenerledigung verwenden wie bei der ersten Vollprofil-CA.

einflussen. Neben den Faktoren dieser Untersuchung, die Objektprofilstärke und die Anzahl vorausgegangener Paarvergleiche, sind z.B. denkbar: Das Ausmaß des Involvements einer Ap, die Höhe der Bezahlung für die Auskunftleistung, die Art der Darbietung der Beurteilungsobjekte, der Antwortmodus etc. Wir halten es für lohnend, diese Bereiche zu erforschen. Durch Verwendung des Antwortfehlers oder ähnlicher Konzepte kann sich die anwendungsorientierte Forschung zur CA den Prinzipien grundlagenorientierter Forschung annähern. Sollte dies zur Auffindung von Gesetzmäßigkeiten führen, die über den jeweiligen spezifischen Untersuchungskontext verallgemeinerbar sind, dann wäre dies ein Gewinn für die Forschung zur multiattributiven Präferenzstrukturmodellierung.

6.6 Zusammenfassung

Diese Arbeit hat zwei Hauptteile. Im ersten, theoretischen Teil (Abschnitte 1 bis 5) wurden die Verfahrensmöglichkeiten der CA und verwandter Methoden zur multiattributiven Präferenzstrukturmodellierung dargestellt und vor dem Hintergrund des Forschungsstandes kritisch diskutiert. Der Kern dieses Unterfangens bestand in der Darstellung des Prozesses einer conjointanalytischen Untersuchung (Abschnitt 3). Hierzu wurde die CA in Stufen aufgeteilt. Die Stufenbildung orientierte sich an den Problemen, die ein Anwender zu lösen hat, wenn er eine empirische Untersuchung nach dem Prinzip der CA oder verwandter Nutzenschätzverfahren durchführen will:

1. Welches Untersuchungsziel und -setting?
2. Welches Attributset?
3. Welche Attributstufen?
4. Welches Präferenzstrukturmodell?
5. Welche Beurteilungsobjekte? (Design)
6. Welche Erhebungsform der abhängigen Variablen? (Antwortmodus)
7. Welches Verfahren zur Nutzenschätzung?
8. Welche Methode der Parameterschätzung?

Eine ausführliche Zusammenstellung dieser Problembereiche und ihrer Verfahrensmöglichkeiten findet sich in Tabelle A des Anhangs.

Abschnitt 4 widmete sich der Weiterverarbeitung der geschätzten PSM. Es kommen Probleme der Normierung, Segmentierung und Aggregation zur Sprache. Ansätze zu ihrer Lösung werden beschrieben und kritisch hinterfragt. In Abschnitt 5 wurden die Verfahren zur Beurteilung der Meßgüte von PSM geordnet und besprochen. Eine besonders kritische Stellung beziehen wir gegenüber der Bestimmung der kriterienorientierten Validität von PSM durch die Beurteilung von Holdout-Objekten, wenn diese Maßzahlen als abhängige Variablen bei der Bearbeitung von Forschungsfragen eingesetzt werden. Andererseits ist es in der Forschung zur CA üblich geworden, die kriterienorientierte Validität zur Entscheidung zwischen conjointanalytischen Verfahrensmöglichkeiten heranzuziehen. Die Qualität der Schlüsse aus solchen Untersuchungen wird jedoch vom *Problem der Kriterienreliabilität* bedroht: Die Reliabilität des Kriteriums beeinflusst die gefundene kriterienorientierte Validität

eines PSM. Dabei ist zu erwarten, daß die Kriterienreliabilität aufgrund individuell verschiedener Präferenzen zwischen den Apn beträchtlich variiert, weil sich die Kriteriums-Holdout-Aufgabe den Apn unterschiedlich schwierig präsentiert.

Der zweite, empirische Teil der Arbeit (Abschnitt 6) greift diesen Kritikpunkt auf. Zunächst wird unsere Denkweise zum Problem der Kriterienreliabilität durch die empirische Untersuchung I (CA zu Prüfungsordnungen) überprüft. Wir erhalten den Befund, daß ein Indikator der individuellen Kriteriumsschwierigkeit hoch mit der kriterienorientierten Validität korreliert. Dieses Ergebnis wurde in der empirischen Untersuchung II (Induktion einer wahren Nutzenstruktur), bei der wir die Schwierigkeit von Vollprofil-Rangordnungsaufgaben im Zusammenhang mit der Lernerfolgsüberprüfung experimentell variierten, nochmals bestätigt.

Wir entwickelten für die empirische Untersuchung II ein spezielles Forschungsdesign, das eine effektive Bearbeitung von Fragestellungen zur CA erlaubt. Das Forschungsdesign vermeidet typische Schwierigkeiten, die durch das Problem der Kriterienreliabilität entstehen, ohne die Vergleichbarkeit mit üblichen Forschungsansätzen zur CA aufzugeben. Bei diesem Design lernen die Apn, Urteile in Einklang mit Nutzenstrukturen abzugeben, die dem Forscher bekannt sind. Auf diese Weise können die Ergebnisse von Nutzenschätzverfahren direkt mit der wahren Nutzenstruktur verglichen werden. Es steht damit ein Indikator der PSM-Meßgüte zur Verfügung, der nicht durch das Kriterienreliabilitätsproblem beeinflusst wird.

Aus der umfangreichen Darstellung und Bewertung conjointanalytischer Ansätze in Abschnitt 3 leiteten wir Charakteristika ab, die ein CA-Verfahren besitzen müßte, um derzeit als optimal gelten zu können. ACA 4.0 ist ein Softwaresystem zur CA, das über viele dieser Eigenschaften verfügt und gleichzeitig eine sehr hohe Verbreitung erfährt. Aus diesem Grunde haben wir in Abschnitt 6 eine aktuelle Forschungsfrage aufgegriffen, die im Zusammenhang mit der Optimierung von ACA 4.0 gestellt werden kann. Es handelte sich dabei um die Frage, welche Profilstärke bei abgestuften, multiattributiven Paarvergleichen bei der CA als optimal zu betrachten ist.

ACA 4.0 ist als Forschungsinstrument nur bedingt geeignet. Deswegen wurde ein eigenes Softwaresystem zur Präferenzstrukturmodellierung (ALASCA) programmiert, das u.a. in der Lage ist, ACA zu imitieren. Die

Vergleichbarkeit von ALASCA zur ACA-Phase III (adaptiv vorgegebene, abgestufte, multiattributive Paarvergleiche) wurde durch Computersimulationen überprüft und als gegeben erachtet. Die Forschungsbefunde, die man mit ALASCA erhält, sollten somit auch auf ACA übertragbar sein.

Zur Bearbeitung der Forschungsfragen nach der optimalen Profilstärke und der günstigen Anzahl der Paarvergleiche führten wir eine empirische Untersuchung (II; Induktion einer wahren Nutzenstruktur) durch. Wie die empirische Untersuchung II zeigt, ist der Einfluß der Objektprofilstärke auf den Antwortfehler so ausgeprägt, daß die statistischen Vorteile starker Profile mehr als ausgeglichen werden. Die höchste Meßgüte erreichen PSM, die auf der Basis von Paarvergleichen der Profilstärke 2 geschätzt wurden. Die Profilstärken 3 und 4 erbringen keinen Unterschied. Es ist zu vermuten, daß sich ihre statistischen und psychologischen Auswirkungen gerade aufheben.

Wie viele Paarvergleiche sollen von den Apn eines CA-Interviews durchgeführt werden? Diese Frage kann nicht allgemeingültig beantwortet werden. Die optimale Anzahl hängt hauptsächlich von der Auskunftsfähigkeit und -motivation der Apn und der Güte der initialen Grobschätzung der PSM ab, die ein gutes CA-Verfahren vor der Paarvergleichsphase einsetzen sollte.

Diese Forschungsarbeit bringt hinsichtlich der optimalen Objektprofilstärke eindeutige Befunde: Paarvergleiche der Profilstärke 2 erbringen eine höhere Meßgüte als stärkere Profile. Darüber hinaus führen geringere Profilstärken zu einer kürzeren Interviewdauer und werden von den Apn als einfacher wahrgenommen.

Anhang

Anhang A Der schrittweise Ablauf einer Conjoint Analyse im Überblick

Tabelle A: Der Prozeß der Conjoint Analyse und Verfahrensmöglichkeiten im Überblick.

Problembereich	Vorgehensweise/ Verfahren
Untersuchungsziel	<p>Modellierung von Einstellungen</p> <p>Modellierung von Präferenzen</p> <p>Modellierung von Verhaltensabsichten (Kaufabsicht, ...)</p> <p>Modellierung von Verhalten (Kaufentscheidungen, ...)</p>
Untersuchungssetting	
Urteils- und Entscheidungsaufgabe	<p>Präferenz (Was bevorzugen Sie?)</p> <p>handlungsgebundene Präferenz (Stellen Sie sich vor, Sie [kaufen; wünschen sich; empfehlen; ...] Was würden Sie bevorzugen?)</p> <p>zielgebundene Präferenz (Stellen Sie sich vor, Sie kaufen [zu einem bestimmten Zweck] Was würden Sie bevorzugen?)</p> <p>situationsgebundene Präferenz (Stellen Sie sich vor, Sie kaufen [in einer bestimmten Situation] Was würden Sie bevorzugen?)</p> <p>zeitorientierte Präferenz (Stellen Sie sich vor, Sie kaufen [heute; in fünf Jahren] Was würden Sie bevorzugen?)</p> <p>stellvertretende Präferenz (Stellen Sie sich vor, [Ihre Mutter, ein Ehepaar, ...] kauft Was würde sie bevorzugen?)</p>
ökologische Validität	<p>Vergleichbarkeit der Untersuchungsentscheidungssituation mit der echten Entscheidungssituation</p> <p>Vergleichbarkeit der Apn mit tatsächlichen Entscheidern</p> <p>Vergleichbarkeit hinsichtlich der Auskunftsfähigkeit/ Objektbereichskennntnisse</p> <p>Vergleichbarkeit hinsichtlich der Auskunftsmotivation</p>

Tabelle A: (Fortsetzung). Der Prozeß der Conjoint Analyse und Verfahrensmöglichkeiten im Überblick.

Problembereich	Vorgehensweise/ Verfahren
Attribute und Attributset	
Art der Attribute	potentiell relevante Attribute determinante Attribute empirisch unabhängige Attribute präferenzunabhängige Attribute relevante Attribute maßgebliche Attribute
Gewinnung potentiell relevanter Attribute	Dokumente - Programmentscheider - Objektentscheider
direkte Verfahren	direkte Befragung Elicitationstechnik Gruppeninterview und -diskussion Kreativitätstechniken Repertory Grid Technik
indirekte Verfahren	lautes Denken beim realen Entscheidungsprozeß Tiefeninterview projektive Verfahren multidimensionale Skalierung
Selektion der Attribute	Einzelurteile Gruppendiskussion und -urteil Stufenverfahren
Repräsentation der Attribute	objektive Eigenschaften (characteristics) subjektiver utilitaristischer Nutzen (benefits) symbolischer Nutzen (imagery)
Attributstufen	
Definition der Attributstufen	diskrete Merkmale (qualitative Attribute) stetige Merkmale (quantitative Attribute) Endpunkteproblem Anzahl der Attributstufen

Tabelle A: (Fortsetzung). Der Prozeß der Conjoint Analyse und Verfahrensmöglichkeiten im Überblick.

Problembereich	Vorgehensweise/ Verfahren
Präferenzstrukturmodell	
Präferenzmodell	lineares Vektormodell
	Idealpunktmodell
	Teilnutzenwertmodell
	gemischtes Modell
Präferenzintegrationsmodell	nichtkompensatorische Verknüpfung
	Dominanzregel
	konjunktive Regel
	disjunktive Regel
	lexikographische Regel
	kompensatorische Verknüpfung
	additives Modell
	multiplikatives Modell
polynomiales Modell	
Prozeßmodell	einstufiger Urteilsprozeß
	sequentieller Urteilsprozeß

Tabelle A: (Fortsetzung). Der Prozeß der Conjoint Analyse und Verfahrensmöglichkeiten im Überblick.

Problembereich	Vorgehensweise/ Verfahren
Konstruktion der Beurteilungsobjekte	
Aufwärmphase	Kennenlernen des Beurteilungssettings und der Attribute/ Stufen durch physische Erfahrung, Präsentation, Text, Film, ... kompositorische Urteilsbildungen
Definition der Beurteilungsobjekte	Vollprofilmethode Trade-off Methode Teilprofilmethode
Konstruktion des Objektssets	faktorielles Design reduziertes faktorielles Design Zufallsdesign Paarvergleichsdifferenzendesign
Objektpräsentation	schriftlich verbal (mündlich/ fernmündlich) bildhaft modellhaft
Unterstützung der Präsentation	Interviewer Telefon-Mail-Telefon Symbole
Erhebung der abhängigen Variablen	
Beurteilungsaufgabe	Präferenz, Nützlichkeit, Kaufabsicht, Geldeinsatz Ähnlichkeit, Nähe, Distanz
Meßskala der abhängigen Variablen	Kategorienzuordnung Rangreihung Ratingskala Magnitudenskala Paarvergleich abgestufter Paarvergleich Konstant-Summen-Paarvergleich Dollar-Metrik-Paarvergleich

Tabelle A: (Fortsetzung). Der Prozeß der Conjoint Analyse und Verfahrensmöglichkeiten im Überblick.

Problembereich	Vorgehensweise/ Verfahren
Verfahren zur Nutzenschätzung	
kompositorische Verfahren	SMART (Edwards, 1977) Standardverfahren (z.B. Green, 1984) Dollar-Metrik-Verfahren (Leigh et al., 1984) konjunktiv-kompensatorische explizite Modellbildung (Srinivasan, 1988)
dekompositorische Verfahren	klassische Conjoint Analyse hierarchische Informationsintegrations-CA (HICA) Hybrid Conjoint Analyse (HCA) adaptive Conjoint Analyse (ACA)
Parameterschätzalgorithmen	
nominale Skalen	CCM
ordinale Skalen	MONANOVA PREFMAP LINMAP JOHNSON
metrisches Skalenniveau	OLS MSAE ANOVA
probabilistische Verfahren	LOGIT

Anhang B Beispiel einer konjunktiv-kompositorischen expliziten Modellbildung nach Srinivasan (1988)

In jedes der stark umrandeten Felder in Spalte A und Spalte B sind nun Zahlen von 0 bis 100 einzutragen. Wir bitten Dich, genau nach folgendem Muster vorzugehen:

1. Wenn es bei den untenstehenden Eigenschaften solche geben sollte, die Du auf gar keinen Fall akzeptieren könntest, dann streiche diese bitte durch. Bitte gehe mit dieser Möglichkeit sehr sparsam um und streiche nur die Eigenschaften, bei welchen Du eine Prüfungsordnung auch dann nicht akzeptieren könntest, wenn sie sonst alle wünschenswerten Eigenschaften hätte. Wenn Du Eigenschaften durchstreichen solltest, dann beachte diese bitte im weiteren Verlauf nicht mehr.
2. Entscheide Dich bitte bei jedem der fünf Attribute (I-V) für die Eigenschaft, welche Dir am ehesten zusagt und für die, welche Dir am wenigsten zusagt. Trage dann in Spalte A jeweils die Zahl 100 für die Eigenschaft ein, die dir am ehesten zusagt. Trage die Zahl 0 für die Eigenschaft ein, die Dir am wenigsten zusagt. In Spalte A stehen also dann bei jedem Attribut (I-V) jeweils eine 0 und eine 100.
3. Beurteile bitte bei jedem Attribut, wie wichtig Dir der Unterschied zwischen der am wenigsten bevorzugten Eigenschaft (0) und der am meisten bevorzugten Eigenschaft (100) ist. Trage in die Kästchen in Spalte B beim wichtigsten Unterschied die Zahl 100 ein. Beurteile danach für die anderen 5 Attribute auf einer Skala von 0 bis 100, wie wichtig Dir der Unterschied zwischen der am niedrigsten und am höchsten bevorzugten Eigenschaft ist. Die eingetragenen Zahlen sollten die Relation der Wichtigkeit zwischen den Attributen widerspiegeln (also z.B. 0: Der Unterschied ist mir überhaupt nicht wichtig; 100: Dieser Unterschied ist mir am wichtigsten).
4. Beurteile nun wieder in Spalte A das Ausmaß Deiner Bevorzugung für die übrigen Eigenschaften. Die Zahlen 0 und 100 hast Du schon in Schritt 2. eingetragen, trage jetzt bitte eventuell fehlende Zahlen nach. Die Zahlen sollten wieder die Relation der Eigenschaften untereinander für jedes Attribut getrennt widerspiegeln. Wenn Du Schritt 2. richtig gemacht hast, liegen diese Zahlen zwischen 0 und 100.

	Spalte A	Spalte B
I Wann Diplomarbeit?		
Erst die Diplomarbeit, dann die Hauptdiplomprüfungen		
Erst die Hauptdiplomprüfungen, dann die Diplomarbeit		
II Wegfall von Prüfungen		
Wegfall der Prüfung zur FoV (forschungsorientierte Vertiefung)		
Wegfall der Prüfung zum nichtpsychologischen Wahlpflichtfach		
Wegfall der Prüfung Physiologie		
Erwerb eines großen Scheins statt VD-Prüfung in einem Fach		
Zusammenfassung der Prüfungen Allgemeine Psychologie I und II		
III Prüfungsart		
Prüfungen des Vor- und Hauptdiploms gestaffelt		
Prüfungen des Vor- und Hauptdiploms in einem Block		
Prüfungen des Vor- und Hauptdiploms in zwei Blöcken		
IV Schwerpunktfächer		
Schwerpunkt Klinische Psychologie		
Schwerpunkt Arbeitspsychologie		
Schwerpunkt Organisationspsychologie		
Schwerpunkt Pädagogische Psychologie		
V Wann Praktika?		
Ein Praktikumssemester		
Kein Praktikumssemester, Praktika vor Studienbeginn anrechenbar		
Kein Praktikumssemester, Praktika nach Studienbeginn anrechenbar		
Kein Praktikumssemester, Praktika nach 2. Semester anrechenbar		
Kein Praktikumssemester, Praktika nach Vordiplom anrechenbar		

Vielen Dank für Deine Mitarbeit!

Anhang C Computersimulationen: Vergleich von ALASCA und ACA 4.0

Methoden

Um mit ACA Computersimulationen durchführen zu können, entwickelten wir ein Antwortprogramm für ACA. Es soll hier ACASIM¹¹ genannt werden. Im Gegensatz zu den Antwortprogrammen früherer Simulationsstudien mit ACA (z.B. Agarwal, 1988a), erfolgt die Erfassung der von ACA gestellten Paarvergleichsfragen und die Rückgabe der Antwort bei ACASIM automatisch. ACASIM benötigt zur Generierung der Antworten die TNW des wahren PSM. Das Programm ist in der Lage, der Antwort einen normal verteilten Antwortfehler zu überlagern. Dieser wird vom Untersucher in Standardabweichungen (s) vorgegeben. Ein Antwortfehler von s=1 bedeutet für ACASIM, daß es zunächst die fehlerfreie Antwort auf der neunstufigen Paarvergleichsskala sucht und dieser einen normal verteilten Antwortfehler mit dem Erwartungswert 0 und der Standardabweichung 1 zuaddiert. Tabelle C-1 zeigt die Wahrscheinlichkeiten der addierten Fehlerwerte für verschiedene Antwortfehlerstufen.

Tabelle C-1: Theoretische Auftretenswahrscheinlichkeit [in %] des zur Antwort addierten Fehlerwertes für verschiedene Antwortfehlerstufen beim Antwortsimulationsprogramm ACASIM.

addierter Fehlerwert	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8
s=0									100								
s=1						0,6	6,1	24,2	38,3	24,2	6,1	0,6					
s=2			0,2	0,9	2,8	6,6	12,1	17,5	19,7	17,5	12,1	6,6	2,8	0,9	0,2		
s=3	0,4	0,9	1,8	3,3	5,5	8,1	10,6	12,5	13,2	12,5	10,6	8,1	5,5	3,3	1,8	0,9	0,4
s=4	1,4	2,2	3,2	4,6	6,0	7,5	8,8	9,6	9,9	9,6	8,8	7,5	6,0	4,6	3,2	2,2	1,4

Die Antwortskala ist beim abgestuften Paarvergleich nach links und nach rechts begrenzt [-4; +4]. Würde eine Antwort aufgrund der wahren Nutzenstruktur und des addierten Fehlers außerhalb der möglichen Antworten liegen (z.B. +6), gibt ACASIM die entsprechende Skalengrenze an ACA zurück (im Beispiel +4).

¹¹ © 1996 AE Holling, Psychologisches Institut IV an der Westfälischen Wilhelms-Universität Münster

Dadurch ist die Streuung des tatsächlich addierten Antwortfehlers etwas kleiner, als dies in Tabelle C-1 dargestellt ist. Dies sollte jedoch deswegen unproblematisch sein, weil das adaptive Paarvergleichsdesign möglichst nutzenähnliche Beurteilungsobjektepaare vorgibt. Die Folge daraus ist, daß die wahre Antwort tendenziell im mittleren Bereich der Antwortskala liegt.

Es wird angenommen, daß vor der Paarvergleichsphase keine Informationen über die TNW des PSM vorliegen. Die Initialschätzwerte aller TNW werden deshalb auf Null gesetzt. Da sowohl das adaptive Paarvergleichsdesign als auch der Antwortgenerator (im Falle $s > 0$) eine zufällige Komponente enthalten, wurden die Ergebnisse von je 200 'gleichen' Simulationsdurchläufen gemittelt. Die Simulationen werden für fünf Fehlerstufen ($s=0$, $s=1$, $s=2$, $s=3$ und $s=4$) und drei wahre Nutzenstrukturen durchgeführt. Die Nutzenstrukturen kann man mit den Namen 'Typ I: homogen und monoton', 'Typ II: homogen und nicht monoton' und 'Typ III: inhomogen' bezeichnen (Erklärung s.u.). Die PSM enthalten 5 Attribute, die jeweils 3-fach gestuft sind. Tabelle C-2 zeigt die 'wahren TNW' der drei Modelle.

Tabelle C-2: Die TNW der simulierten wahren Modelle vom Typ I, II und III.

Attribut und Attributstufe wahres PSM	Attribut 1			Attribut 2			Attribut 3			Attribut 4			Attribut 5		
	11	12	13	21	22	23	31	32	33	41	42	43	51	52	53
Typ I: homogen und monoton	4	2	0	3,6	1,8	0	3,2	0,32	0	2,8	1,96	0	2	1	0
Typ II: homogen und nicht monoton	2	4	0	0	3,6	1,8	3,2	0	0,32	1,96	0	2,8	2	0	1
Typ III: inhomogen	2	4	0	0	3,6	1,8	3,2	0	0,32	0,2	0	0,7	0,4	0	0,2

Der Name 'monoton' des PSM des Typs I erklärt sich daraus, daß der Nutzenverlauf innerhalb der Attribute monoton ist. Die relativen Wichtigkeiten der Attribute (vgl. Abschnitt 2.3.2) betragen 4/ 3,6/ 3,2/ 2,8 und 2. Das wichtigste Attribut ist damit gerade doppelt so wichtig, wie das unwichtigste Attribut. Diese Eigenschaft des PSM nennen wir 'homogen'. Das PSM des Typs II ist ebenfalls homogen. Im Gegensatz zu Typ I wurden nur die TNW der Attributstufen innerhalb der Attribute vertauscht. Die Nutzenverläufe sind damit nicht monoton. Das Modell des Typs III ist eine modifizierte Form des Typ II-Modells. Die relative Wichtigkeit der Attribute 4 und 5 wurde so gemindert, daß das wichtigste Attribut

jetzt das Zehnfache der relativen Wichtigkeit des unwichtigsten Attributs besitzt. Diese Eigenschaft nennen wir inhomogen.

ALASCA besitzt einen eigenen Simulator. Die Funktionsweise ist analog dem Antwortsimulator ACASIM. Die Antworten des Simulators generieren sich aus dem wahren PSM und einem normal verteilten Fehler mit spezifizierter Standardabweichung.

ACA 4.0 führt bei fünf dreifach gestuften Attributen maximal 12 Paarvergleiche durch (Gleichung 3.18). Nach jedem Paarvergleich wird ein Maß der Strukturwiedererkennung des wahren PSM berechnet. Das Prinzip der Berechnung der Wiedererkennungseistung ist in Abschnitt 5.5.1 detailliert beschrieben. Wir normieren das wahre und das geschätzte Modell zunächst durch die Normierungsvorschrift *StaNorm* (Standardabweichung der TNW=150). Anschließend wird als Maßzahl der Strukturwiedererkennungseistung SSE (Sum of Square Error; Judd & McClelland, 1989) zwischen dem geschätzten und dem wahren PSM berechnet. Aus Gründen der einfacheren graphischen Darstellung wird aus SSE die Wurzel gezogen. Die so entstandene Maßzahl bezeichnen wir mit $W(SSE)$. Sie ist mit alternativen Maßzahlen der Strukturwiedererkennung weitgehend redundant (vgl. Abschnitt 5.5.1) und ist proportional mit der ebenfalls häufig verwendeten Maßzahl MSE (Mean Square Error).

Ergebnisse

Nach jedem Paarvergleich kann man die Maßzahl $W(SSE)$ berechnen. Zu ihrer Interpretation ist anzumerken, daß der Wert von $W(SSE)$ beim Initialmodell keine inhaltliche Bedeutung hat, weil die TNW des Initial-PSM alle gleich sind und deshalb keine Normierung des Modells möglich ist.

Die Abbildungen C-1 bis C-12 zeigen die Strukturwiedererkennungseistung des wahren Modell des Typs I, II und III von ACA 4.0 und ALASCA. Dargestellt sind die Abbildungen jeweils für die Profilstärke 2 und 4 und den Antwortfehler $s=0$ und $s=4$. Diese wurden ausgewählt, um die Menge der graphischen Darstellungen klein zu halten. Die Graphen der Profilstärke 3 liegen zwischen denen der dargestellten Profilstärken 2 und 4. Genauso verhält es sich mit den Fehlerstufen $s=1$, $s=2$ und $s=3$. Sie liegen zwischen den Antwortfehlern $s=0$ und $s=4$.

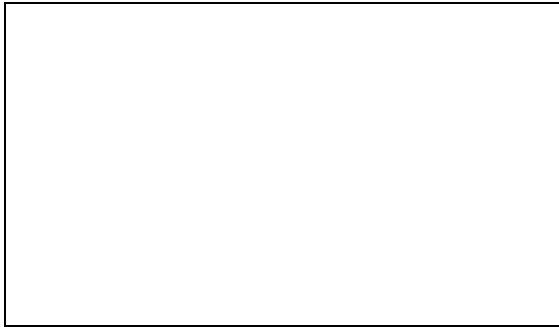


Abb. C-1: Typ I: Profilstärke 2, Antwortfehler $s=0$.

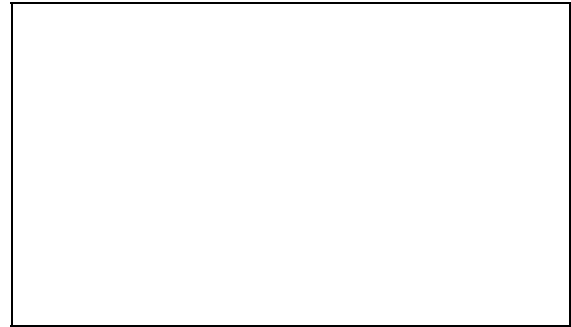


Abb. C-2: Typ I: Profilstärke 4, Antwortfehler $s=0$.

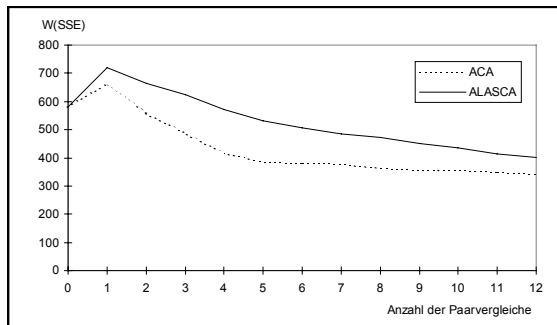


Abb. C-3: Typ I: Profilstärke 2, Antwortfehler $s=4$.

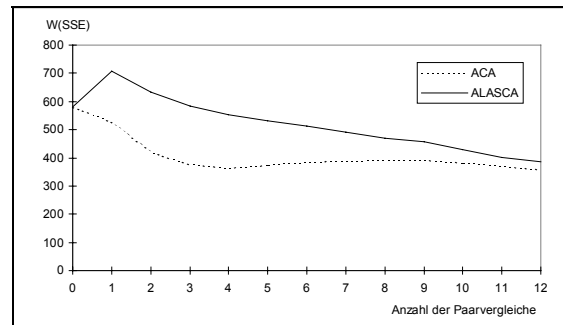


Abb. C-4: Typ I: Profilstärke 4, Antwortfehler $s=4$.

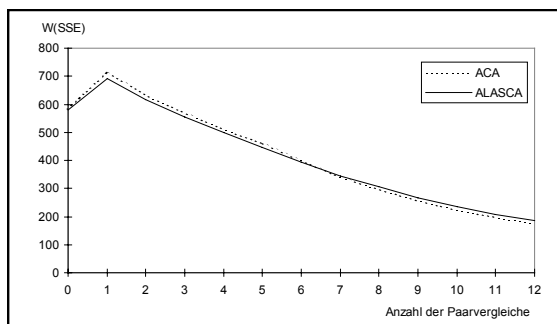


Abb. C-5: Typ II: Profilstärke 2, Antwortfehler $s=0$.

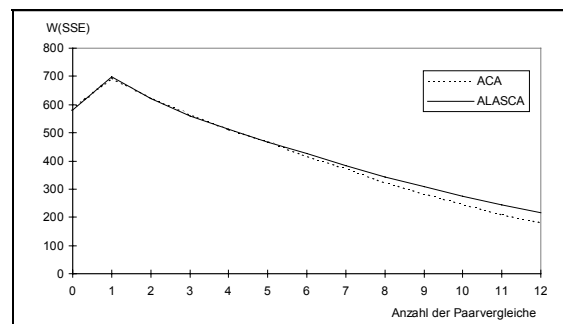
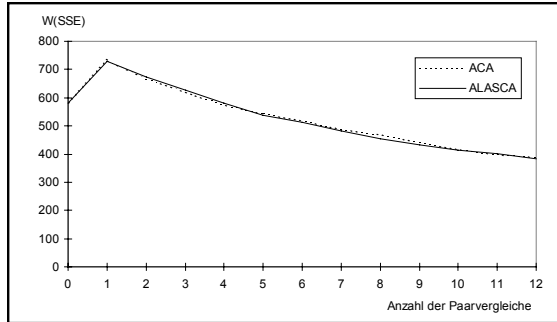
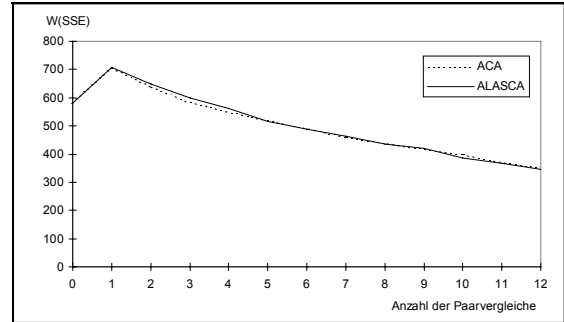
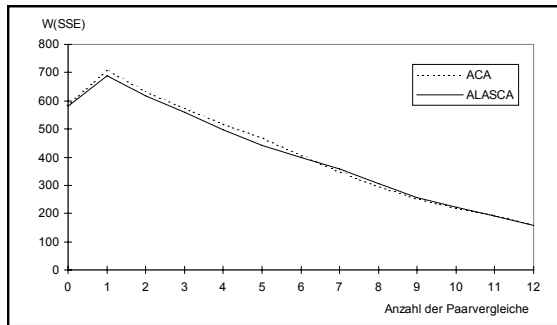
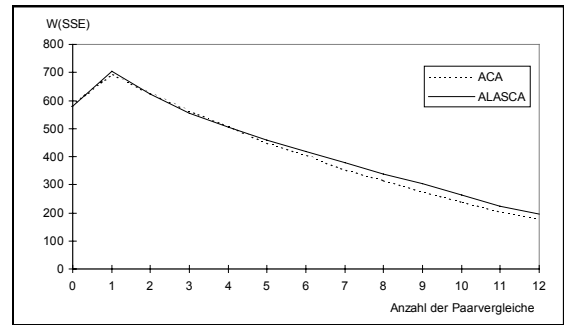
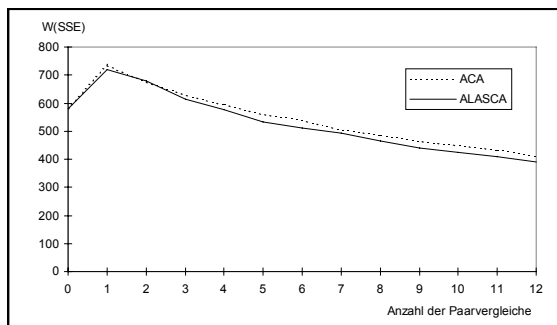
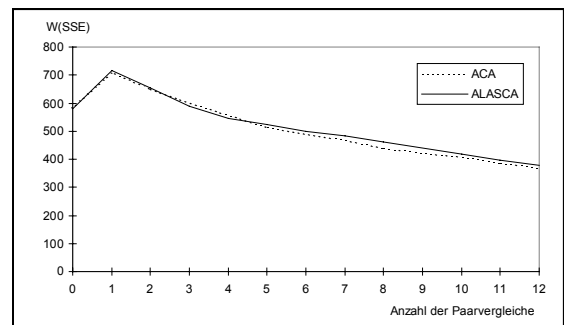


Abb. C-6: Typ II: Profilstärke 4, Antwortfehler $s=0$.

Abb. C-7: Typ II: Profilstärke 2, Antwortfehler $s=4$.Abb. C-8: Typ II: Profilstärke 4, Antwortfehler $s=4$.Abb. C-9: Typ III: Profilstärke 2, Antwortfehler $s=0$.Abb. C-10: Typ III: Profilstärke 4, Antwortfehler $s=0$.Abb. C-11: Typ III: Profilstärke 2, Antwortfehler $s=4$.Abb. C-12: Typ III: Profilstärke 4, Antwortfehler $s=4$.

Grundsätzlich ist aus allen durchgeführten Simulationen deutlich eine Gemeinsamkeit beider Paarvergleichsdesigns erkennbar: Die Wiedererkennung der wahren Nutzenstruktur nimmt mit dem Antwortfehler ab und mit der Anzahl der Paarvergleiche und der Profilstärke zu.

Beim PSM des Typs I ist erkennbar, daß ACA bereits nach dem ersten Paarvergleich eine bessere Strukturwiedererkennungsleistung zeigt. Der Effekt ist um so deutlicher, je stärker die Objektprofile sind. Offensichtlich hat ACA bei der Wahl der ersten Frage eine besonders 'glückliche Hand'. Dies ist deswegen verwunderlich, weil zu diesem Zeitpunkt keinerlei Hinweise darüber zur Verfügung stehen, welche Attributstufen in der Paarvergleichsfrage günstig sind, um möglichst viel Information über das wahre PSM zu gewinnen. ALASCA zieht daraus die logische Konsequenz und wählt die erste Frage zufällig aus. ACA dagegen macht an dieser Stelle offensichtlich mehrere Annahmen darüber, wie wahre Nutzenstrukturen in aller Regel beschaffen sind. Es würde hier zu weit führen, die Annahmen und die Vorgehensweise des ACA-Paarvergleichsdesigns beim ersten Paarvergleich herauszuarbeiten. Sie finden im Programmhandbuch (Sawtooth Software, 1994) keinerlei Erwähnung und sie zu erraten, bedürfte einer eigenen umfangreichen Arbeit. Wir möchten an dieser Stelle aber vermuten, daß z.B. die Annahme von monotonen Nutzenverläufen innerhalb der Attribute eine Rolle spielt. Wie die Abbildungen C-1 bis C-12 zeigen, verschwindet dieser Vorteil des ACA-Designs vollständig, wenn das wahre PSM keine monotonen Nutzenverläufe innerhalb der Attribute besitzt.

Das Ausnutzen der Monotonitätsannahme bei der Designbildung besitzt eine gewisse Berechtigung, wenn man davon ausgeht, daß die meisten Anwender von ACA intuitiv die Reihenfolge der Attributstufen so wählen, daß der Nutzen bei der Mehrzahl der Apn monoton verlaufen wird. Das ALASCA-Design macht diese Annahme nicht und erstellt die Paarvergleiche nur nach dem bislang geschätzten PSM. Sind wie in diesem Fall die TNW der Initialschätzung identisch, wird die Auswahl per Zufall gesteuert. Wir halten diese Vorgehensweise für theoretisch angemessener. Die Beibehaltung scheint insbesondere deswegen unproblematisch, da der 'Vorsprung' von ACA innerhalb weniger Paarvergleiche aufgeholt wird.

Vernachlässigt man die minimale Überlegenheit des ACA-Designs für den Spezialfall von monotonen und homogenen wahren Nutzenverläufen, zeigt das Verhalten der beiden adaptiven Designs und die darauf beruhende Strukturwiedererkennung keinen Unterschied.

Diskussion

Wie aus allen vergleichenden Simulationen deutlich wird, verhalten sich ACA und ALASCA bei der Wiedererkennung einer wahren Nutzenstruktur sehr ähnlich. Die Wiedererkennungsleistung nimmt erwartungsgemäß mit der Anzahl der Paarvergleiche und der Profilstärke zu und mit der Höhe des Antwortfehlers ab. Bei einigen speziellen wahren Nutzenstrukturen finden sich kleinere Unterschiede bei der ersten Paarvergleichsfrage. Dies spricht jedoch nicht gegen den Einsatz von ALASCA zur Untersuchung der gestellten Forschungsfrage. Demzufolge ziehen wir den Schluß, daß die Ergebnisse einer empirischen Untersuchung zur Anzahl der Paarvergleiche und deren Profilstärke mit ALASCA auch für ACA 4.0 oder eine modifizierte Version von ALASCA gelten.

Es wurden zum Vergleich von ACA 4.0 und ALASCA nur Simulationen mit einer bestimmten Anzahl von Attributen und Attributstufen, einer bestimmten Initialschätzung und drei mehr oder weniger willkürlich festgelegten wahren Nutzenstrukturen durchgeführt. Durch diese Einschränkungen sind auch die Grenzen der Aussagekraft dieser vergleichenden Simulationsstudie bestimmt. Es ist zu vermuten, daß sich das Verhalten der beiden adaptiven Paarvergleichsdesigns geringfügig in Abhängigkeit o.g. Variablen ändert. Aus diesem Grund wäre eine Verallgemeinerung der gefundenen Unterschiede zwischen den Designs von ACA und ALASCA verfrüht. Es ist jedoch zu vermuten, daß die ungefähren Gemeinsamkeiten beider Verfahren generell zu finden sind.

Zusammenfassend kann festgehalten werden, daß das adaptive Paarvergleichsdesign von ALASCA sich nur geringfügig von ACA unterscheidet. Die Untersuchungsergebnisse zur gestellten Forschungsfrage, die mit der Forschungssoftware ALASCA gewonnen werden, sollten sich auf ACA 4.0 übertragen lassen.

Anhang D Die möglichen Objekte und deren Einsatz

Tabelle D: Die möglichen Objekte und deren Verwendung in verschiedenen Phasen der empirischen Untersuchung II.

Lfd. Nr.	Code	Entfernung	Größe	Art des Hauses	Zustand	Miete	Nutzen	Objekt der Lernphase	Objekt des Validierungssets 1	Objekt des Validierungssets 2
1	303	1	1	1	1	1	390		X	X
2	786	1	1	1	2	1	369			
3	952	1	1	1	1	2	365			
4	919	1	2	1	1	1	345			
5	905	1	1	1	2	2	344			
6	524	1	1	1	1	3	340			
7	935	2	1	1	1	1	340			
8	544	1	2	1	2	1	324			
9	141	1	1	1	3	1	320			
10	349	1	2	1	1	2	320			
11	372	1	1	1	2	3	319			
12	543	2	1	1	2	1	319			
13	261	1	1	2	1	1	318			
14	159	2	1	1	1	2	315			
15	871	1	1	3	1	1	310			
16	858	1	3	1	1	1	300			
17	477	1	2	1	2	2	299			
18	857	1	1	2	2	1	297			
19	285	1	1	1	3	2	295			
20	469	1	2	1	1	3	295			
21	498	2	2	1	1	1	295			
22	199	2	1	1	2	2	294			
23	337	1	1	2	1	2	293	A		
24	599	2	1	1	1	3	290		X	X
25	643	3	1	1	1	1	290			
26	187	1	1	3	2	1	289			X
27	532	1	1	3	1	2	285			
28	638	1	3	1	2	1	279			
29	181	1	2	1	3	1	275			
30	995	1	3	1	1	2	275			
31	154	1	2	1	2	3	274			
32	706	2	2	1	2	1	274			
33	925	1	2	2	1	1	273	B		
34	754	1	1	2	2	2	272			
35	391	1	1	1	3	3	270			
36	548	2	1	1	3	1	270			
37	879	2	2	1	1	2	270			
38	302	2	1	1	2	3	269	C		
39	801	3	1	1	2	1	269			
40	107	1	1	2	1	3	268			
41	297	2	1	2	1	1	268			
42	325	1	2	3	1	1	265		X	
43	378	3	1	1	1	2	265	D	X	X
44	116	1	1	3	2	2	264		X	
45	158	1	1	3	1	3	260			
46	808	2	1	3	1	1	260			
47	259	1	3	1	2	2	254			
48	675	1	2	2	2	1	252	A		
49	423	1	2	1	3	2	250			
50	767	1	3	1	1	3	250			
51	998	2	3	1	1	1	250			
52	208	2	2	1	2	2	249			
53	307	1	1	2	3	1	248			X
54	713	1	2	2	1	2	248			
55	557	1	1	2	2	3	247		X	
56	682	2	1	2	2	1	247			
57	157	2	1	1	3	2	245			
58	683	2	2	1	1	3	245	B		
59	761	3	2	1	1	1	245			
60	413	1	2	3	2	1	244			
61	418	3	1	1	2	2	244	D		
62	704	2	1	2	1	2	243			
63	228	1	1	3	3	1	240	C		
64	657	1	2	3	1	2	240			X
65	696	3	1	1	1	3	240			
66	255	1	1	3	2	3	239			
67	263	2	1	3	2	1	239			

Lfd. Nr.	Code	Ent- fernung	Größe	Art des Hauses	Zustand	Miete	Nutzen	Objekt der Lernphase	Objekt des Validierungssets 1	Objekt des Validierungssets 2
68	505	2	1	3	1	2	235			
69	364	1	3	1	3	1	230			
70	127	1	3	1	2	3	229			X
71	780	2	3	1	2	1	229		X	
72	924	1	3	2	1	1	228		X	
73	120	1	2	2	2	2	227	A		
74	137	1	2	1	3	3	225		X	
75	235	2	2	1	3	1	225			X
76	893	2	3	1	1	2	225			
77	955	2	2	1	2	3	224			
78	972	3	2	1	2	1	224		X	X
79	380	1	1	2	3	2	223	D		
80	559	1	2	2	1	3	223			X
81	782	2	2	2	1	1	223			
82	705	2	1	2	2	2	222			X
83	171	1	3	3	1	1	220	B		
84	251	2	1	1	3	3	220			
85	365	3	1	1	3	1	220			
86	971	3	2	1	1	2	220			
87	168	1	2	3	2	2	219			
88	726	3	1	1	2	3	219	C		
89	577	2	1	2	1	3	218			
90	751	3	1	2	1	1	218			
91	237	1	1	3	3	2	215			
92	930	1	2	3	1	3	215			
93	978	2	2	3	1	1	215			
94	484	2	1	3	2	2	214			
95	370	2	1	3	1	3	210			
96	615	3	1	3	1	1	210			
97	783	1	3	2	2	1	207	D		
98	110	1	3	1	3	2	205	A	X	X
99	371	2	3	1	2	2	204			
100	425	1	2	2	3	1	203			
101	965	1	3	2	1	2	203			
102	136	1	2	2	2	3	202			
103	432	2	2	2	2	1	202			
104	356	2	2	1	3	2	200			
105	565	2	3	1	1	3	200			
106	691	3	3	1	1	1	200			
107	398	1	3	3	2	1	199			
108	916	3	2	1	2	2	199	B		
109	149	1	1	2	3	3	198			
110	160	2	1	2	3	1	198			
111	397	2	2	2	1	2	198		X	
112	601	2	1	2	2	3	197			
113	752	3	1	2	2	1	197	C		
114	707	1	2	3	3	1	195			
115	708	1	3	3	1	2	195	D		
116	833	3	1	1	3	2	195			
117	939	3	2	1	1	3	195			
118	361	1	2	3	2	3	194			
119	424	2	2	3	2	1	194			
120	828	3	1	2	1	2	193			
121	387	1	1	3	3	3	190			
122	448	2	1	3	3	1	190		X	
123	892	2	2	3	1	2	190	A		
124	439	2	1	3	2	3	189			
125	578	3	1	3	2	1	189			
126	670	3	1	3	1	2	185			
127	284	1	3	2	2	2	182			
128	153	1	3	1	3	3	180			
129	587	2	3	1	3	1	180			
130	377	2	3	1	2	3	179			
131	394	3	3	1	2	1	179			
132	132	1	2	2	3	2	178			
133	351	1	3	2	1	3	178	B/D		
134	949	2	3	2	1	1	178			
135	272	2	2	2	2	2	177			
136	515	2	2	1	3	3	175			
137	838	3	2	1	3	1	175			
138	945	3	3	1	1	2	175	C		
139	148	1	3	3	2	2	174			
140	388	3	2	1	2	3	174			
141	367	2	1	2	3	2	173			
142	482	2	2	2	1	3	173			
143	948	3	2	2	1	1	173			

Lfd. Nr.	Code	Ent- fernung	Größe	Art des Hauses	Zustand	Miete	Nutzen	Objekt der Lernphase	Objekt des Validierungssets 1	Objekt des Validierungssets 2
144	395	3	1	2	2	2	172			
145	108	1	2	3	3	2	170			
146	422	1	3	3	1	3	170			
147	699	2	3	3	1	1	170			X
148	876	3	1	1	3	3	170	A		
149	333	2	2	3	2	2	169			
150	461	3	1	2	1	3	168			
151	818	2	1	3	3	2	165	D		
152	934	2	2	3	1	3	165			
153	961	3	2	3	1	1	165			
154	247	3	1	3	2	2	164			
155	740	3	1	3	1	3	160			
156	241	1	3	2	3	1	158			
157	497	1	3	2	2	3	157			
158	499	2	3	2	2	1	157	B		
159	273	2	3	1	3	2	155			
160	375	3	3	1	2	2	154			
161	113	1	2	2	3	3	153			
162	412	2	2	2	3	1	153			
163	471	2	3	2	1	2	153	C		
164	400	2	2	2	2	3	152			
165	822	3	2	2	2	1	152			
166	138	1	3	3	3	1	150			
167	386	3	2	1	3	2	150			
168	747	3	3	1	1	3	150			
169	376	1	3	3	2	3	149			
170	573	2	3	3	2	1	149			
171	165	2	1	2	3	3	148			
172	732	3	1	2	3	1	148		X	
173	877	3	2	2	1	2	148	A		
174	323	3	1	2	2	3	147			
175	249	1	2	3	3	3	145			
176	559	2	2	3	3	1	145			
177	618	2	3	3	1	2	145			
178	701	2	2	3	2	3	144			
179	891	3	2	3	2	1	144			
180	191	2	1	3	3	3	140			
181	473	3	1	3	3	1	140			
182	600	3	2	3	1	2	140			
183	717	3	1	3	2	3	139	B		
184	551	1	3	2	3	2	133			
185	556	2	3	2	2	2	132			
186	479	2	3	1	3	3	130			
187	785	3	3	1	3	1	130			
188	538	3	3	1	2	3	129	C		
189	458	2	2	2	3	2	128			
190	553	2	3	2	1	3	128			
191	970	3	3	2	1	1	128			X
192	922	3	2	2	2	2	127			
193	156	1	3	3	3	2	125			
194	712	3	2	1	3	3	125			
195	341	2	3	3	2	2	124			
196	575	3	1	2	3	2	123			
197	881	3	2	2	1	3	123			
198	227	2	2	3	3	2	120			
199	344	2	3	3	1	3	120			
200	829	3	3	3	1	1	120			
201	931	3	2	3	2	2	119			
202	253	3	1	3	3	2	115			
203	642	3	2	3	1	3	115			
204	115	1	3	2	3	3	108			
205	140	2	3	2	3	1	108			
206	295	2	3	2	2	3	107			
207	816	3	3	2	2	1	107			
208	770	3	3	1	3	2	105			
209	613	2	2	2	3	3	103			
210	779	3	2	2	3	1	103			
211	957	3	3	2	1	2	103			
212	346	3	2	2	2	3	102			
213	161	1	3	3	3	3	100			
214	246	2	3	3	3	1	100			
215	435	2	3	3	2	3	99			
216	869	3	3	3	2	1	99			
217	776	3	1	2	3	3	98			
218	216	2	2	3	3	3	95			
219	563	3	2	3	3	1	95			
220	711	3	3	3	1	2	95			

Lfd. Nr.	Code	Entfernung	Größe	Art des Hauses	Zustand	Miete	Nutzen	Objekt der Lernphase	Objekt des Validierungssets 1	Objekt des Validierungssets 2
221	980	3	2	3	2	3	94			
222	419	3	1	3	3	3	90			x
223	605	2	3	2	3	2	83			
224	441	3	3	2	2	2	82			
225	420	3	3	1	3	3	80			
226	169	3	2	2	3	2	78			
227	277	3	3	2	1	3	78			
228	379	2	3	3	3	2	75			
229	465	3	3	3	2	2	74			
230	170	3	2	3	3	2	70			
231	234	3	3	3	1	3	70		x	
232	466	2	3	2	3	3	58			
233	933	3	3	2	3	1	58			
234	724	3	3	2	2	3	57			
235	956	3	2	2	3	3	53			
236	114	2	3	3	3	3	50			
237	570	3	3	3	3	1	50			
238	609	3	3	3	2	3	49			
239	632	3	2	3	3	3	45			
240	411	3	3	2	3	2	33			
241	421	3	3	3	3	2	25			
242	299	3	3	2	3	3	8			
243	609	3	3	3	3	3	0			

Anhang E Schriftliche Instruktionen der Lernphase

Liebe Untersuchungsteilnehmerin, lieber Untersuchungsteilnehmer,

Du hast Dich bereit erklärt, an einer Studie zum Thema "Menschliche Wertvorstellungen" teilzunehmen. Durchgeführt wird diese Studie von Mitarbeitern und Diplomanden des Psychologischen Institut IV am Lehrstuhl von Prof. Dr. Holling. Wir möchten Dir vorab für Dein Interesse danken.

Deine Aufgabe wird sein, Dich in eine fremde Person hineinzusetzen, die Dir im folgenden Abschnitt beschrieben wird. Diese Person sucht ein Appartement, das ihren Wünschen und Vorstellungen am ehesten gerecht wird. Wir werden Dir diese Vorstellungen ausführlich darstellen. Anschließend wirst Du über die Präferenzen der Person befragt. Bei dieser Untersuchung sind Deine eigenen Wohnungspräferenzen unwichtig. Dennoch könnte es interessant für Dich sein, etwas über Präferenzstrukturen allgemein zu erfahren.

Nun versuche bitte, Dich in folgende Rolle hineinzusetzen:

Helga K. hat zum WS 95/96 mit dem Betriebswirtschaftstudium in Münster begonnen. Da sie nicht jeden Morgen von ihrer Heimatstadt Castrop-Rauxel nach Münster fahren möchte, sucht sie hier eine Wohnung. Es gibt fünf Kriterien, die für sie bei der Wahl einer Wohnung eine Rolle spielen. Diese sind:

- Die Entfernung zum Institut für Betriebswirtschaft
- Die Größe der Wohnung
- Die Art des Hauses (Wohnheim, Einfamilienhaus oder Mehrfamilienhaus)
- Der Zustand der Wohnung (Neubau, renovierter Altbau oder Altbau)
- Die Höhe der Miete pro Monat (inkl. Nebenkosten)

Diese Kriterien besitzen für Helga eine unterschiedlich große Bedeutung. Da sie eine Abneigung gegen das Fahrradfahren hat und es schwierig ist, in der Innenstadt (in der Nähe des Institutes für Betriebswirtschaft) einen Parkplatz zu bekommen, sollte die Wohnung so nah wie möglich beim Institut liegen. Eine Entfernung von ca. einem Kilometer betrachtet sie als fast ideal. Die maximale Entfernung, die sie noch in Kauf nehmen würde, ist 5 km.

Nur geringfügig weniger wichtig als die Entfernung zum Institut erscheint Helga die Größe der Wohnung. Um Freunden, die sie besuchen, den gewünschten Komfort bieten zu können, besitzt die Größe der zukünftigen Wohnung für Helga

dennoch einen hohen Stellenwert. Hinzu kommt, daß sie ein ausgesprochener Party-Fan ist und auch aus diesem Grund nicht unbedingt in einen Besenschrank ziehen möchte. Je größer die Wohnung, desto größer und besser die Party, lautet Helgas Faustformel. Helgas Eltern sind allerdings der Ansicht, daß eine „Studentenbude“ nicht größer als ca. 35 m² sein sollte und sind deshalb nicht bereit, eine größere Wohnung zu bezahlen. Um nicht mit den Eltern auf Konfrontationskurs zu gehen, könnte sich Helga im äußersten Notfall vorstellen, in eine 15 m² große Wohnung zu ziehen.

Ihre Vorlieben für Parties sind außerdem der Grund dafür, daß sie von einem Platz in einem Wohnheim träumt. In Mehrfamilienhäusern befürchtet Helga, daß es Ärger geben könnte mit älteren Nachbarn, die sich in ihrer Ruhe gestört fühlen. Aus diesem Grund würde sie einem Mehrfamilienhaus ein Einfamilienhaus „im Notfall“ noch vorziehen. Die Art des Hauses ist zwar etwas weniger wichtig als die Größe der Wohnung, spielt jedoch eine etwas wichtigere Rolle als der Zustand der Wohnung.

Am liebsten würde Helga in einen Neubau ziehen, da dort ihrer Meinung nach der geringste Renovierungsbedarf besteht. Der Zustand eines Altbaus ist oft mit entsprechenden Problemen verbunden, deshalb wäre eine Wohnung im Altbau weniger wünschenswert für sie. Mehr als eine Kompromißlösung stellt jedoch ein renovierter Altbau dar.

Weniger wichtig als der Zustand der Wohnung ist für Helga die Höhe der Miete pro Monat. Die Miete bereitet Helga kaum Kopfzerbrechen, da ihre Eltern wohlhabend und zahlungswillig sind. Dennoch möchte sie ihre Eltern nicht bis zum Exzeß schröpfen und sieht daher in 25 DM/m² die Obergrenze. Helga kennt allerdings auch das hohe Niveau der Mieten und hat wenig Hoffnung auf einen Mietpreis von weniger als 15 DM/m².

Melde Dich bitte bei dem Untersuchungsleiter, wenn Du den Text bis hierher gelesen hast. Warte bitte an dieser Stelle, bis alle Untersuchungsteilnehmer mit dem Lesen fertig sind. Blättere erst nach Aufforderung durch den Untersuchungsleiter um.

[Seitenwechse]

Bevor Helga einen Makler aufsucht, will sie zunächst einmal versuchen, ihre Vorstellungen und Wünsche aufzuschreiben und zu ordnen.

Nun versuche bitte, Dich in Helga hineinzusetzen und die Wichtigkeit der einzelnen Merkmale einzuschätzen. Bitte kreuze auf den folgenden Skalen die Wichtigkeit eines jeden Merkmals an. Hält Helga ein Merkmal für sehr wichtig, solltest Du z.B. ein Kreuz ganz links auf der Skala machen.

Entfernung:

wichtig |-----| unwichtig

Größe:

wichtig |-----| unwichtig

Art des Hauses:

wichtig |-----| unwichtig

Zustand:

wichtig |-----| unwichtig

Miete:

wichtig |-----| unwichtig

Warte bitte an dieser Stelle, bis alle Untersuchungsteilnehmer diese Aufgabe beendet haben. Blättere bitte erst nach Aufforderung durch den Untersuchungsleiter um.

[Seitenwechse]

Nachdem Helga die verschiedenen Wohnungsmerkmale nach ihrer Wichtigkeit geordnet hat, überlegt sie, wie angenehm die verschiedenen Ausprägungen der Merkmale für sie sind.

Versuche bitte zu jeder Merkmalsausprägung einzuschätzen, wie Helga sie beurteilt. Kreuze hierzu auf den folgenden Skalen an:

E n t f e r n u n g :

1km
schlecht |-----| gut

3km
schlecht |-----| gut

5km:
schlecht |-----| gut

G r ö ß e :

35m²:
schlecht |-----| gut

25m²:
schlecht |-----| gut

15m²:
schlecht |-----| gut

A r t d e s H a u s e s :

Wohnheim:
schlecht |-----| gut

Einfamilienhaus:
schlecht |-----| gut

Mehrfamilienhaus:
schlecht |-----| gut

Z u s t a n d :

Neubau:
schlecht |-----| gut

renovierter Altbau:
schlecht |-----| gut

Altbau:
schlecht |-----| gut

M i e t e :

15 DM/m²:
schlecht |-----| gut

20 DM/m²:
schlecht |-----| gut

25 DM/m²:
schlecht |-----| gut

Warte bitte an dieser Stelle, bis alle Untersuchungsteilnehmer diese Aufgabe beendet haben. Blättere bitte erst nach Aufforderung durch den Untersuchungsleiter um.

[Seitenwechse]

Wie Du ja bereits erfahren hast, sind die Eltern von Helga wohlhabend. Dadurch sind sie auch in der Lage, ihre Tochter bei der Wohnungssuche finanzkräftig zu unterstützen. Sie ziehen deshalb zahlreiche Informationsquellen zum aktuellen Wohnungsmarkt in Münster heran, geben Inserate in den verschiedenen Tageszeitungen auf und beauftragen mehrere Makler mit der Ermittlung von Wohnungsangeboten. Helga entschließt sich, die Wohnungsangebote zunächst einzeln für jede Quelle (z.B. Makler xy) nach ihren Vorlieben und Wünschen zu ordnen, um dann aus den jeweils besten Angeboten ihre zukünftige Wohnung auszuwählen.

Dir werden nun die Angebote der einzelnen Quellen gezeigt. Bitte versuche herauszufinden, wie Helga die einzelnen Wohnungsangebote bewertet hat.

Anhang F Die Aufgaben der dritten Stufe der Lernphase

Tabelle F: Die dritte Stufe der Lernphase 'vergleichende Beurteilung von Wohnungskonzepten': Die Aufgaben sind in die Gruppen a bis e eingeteilt, je nachdem, auf wie vielen Attributen kompensatorische Urteile gefordert sind. Unterschiedliche Attributausprägungen einer Aufgabe sind kursiv gedruckt. Die Lösung der Aufgaben ist kurz erklärt.

Aufgabe	Lfd. Nr.	Nutzen	Ent- fernung	Größe	Art des Hauses	Zustand	Miete	Erklärung der Präferenz
a1	22	294	3km	35m ²	<i>Wohnh.</i>	<i>ren. Altb.</i>	20	Dominanz auf allen Merkmalen
	235	53	5km	25m ²	<i>Einfam.</i>	<i>Altbau</i>	25	
a2	8	324	1km	25m ²	Wohnh.	<i>ren. Altb.</i>	15	Dominanz durch drei Merkmale, auf keinem Merkmal unterlegen
	167	150	5km	25m ²	Wohnh.	<i>Altbau</i>	20	
a3	113	197	3km	35m ²	Einfam.	ren. Altb.	25	Dominanz durch ein Merkmal, auf keinem Merkmal unterlegen
	207	107	3km	15m ²	3km	ren. Altb.	25	
b1	30	275	1km	15m ²	Wohnh.	Neubau	20	maximal überlegen hinsichtl. Der Entfernung, maximal unterlegen hins. Größe; Entfernung ist das wichtigere Merkmal
	43	265	5km	35m ²	Wohnh.	Neubau	20	
b2	238	49	5km	15m ²	<i>Mehrfam.</i>	<i>ren. Altb.</i>	25	Zustand ist zwar weniger wichtig als Art des Hauses, der renovierte Altbau bietet gegenüber dem Altbau aber mehr Vorzüge als das Einfamilienhaus gegenüber dem Mehrfamilienhaus
	242	8	5km	15m ²	<i>Einfam.</i>	<i>Altbau</i>	25	
b3	156	158	1km	15m ²	Einfam.	Altbau	15	der große Preisunterschied wird höher bewertet als der mittlere Größenunterschied
	163	153	1km	25m ²	Einfam.	Altbau	25	
Lerncheck								
b4	105	200	5km	15m ²	<i>Wohnh.</i>	Neubau	15	Die Vorzüge des Wohnheims überwiegen die Vorteile der (etwas) geringeren Entfernung
	134	178	3km	15m ²	<i>Einfam.</i>	Neubau	15	
b6	109	198	1km	35m ²	Einfam.	<i>Altbau</i>	25	Vorteil hinsichtlich Größe, kann nicht durch besseren Zustand einer Alternative ausgeglichen werden; der renovierte Altbau ist mehr als eine Kompromißlösung und bietet gegenüber dem Altbau mehr Vorzüge als ein mittlerer Größengewinn
	157	157	1km	15m ²	Einfam.	<i>ren. Altb.</i>	25	
	161	153	1km	25m ²	Einfam.	<i>Altbau</i>	25	
b8	85	220	5km	35m ²	<i>Wohnh.</i>	Altbau	15	Vorteil auf zwei Merkmalen, auf keinem Merkmal unterlegen Eine um 5 DM/m ² geringere Miete wird offenbar als besser beurteilt als ein Einfamilienhaus anstatt eines Mehrfamilienhauses
	202	115	5km	35m ²	<i>Mehrfam.</i>	Altbau	20	
	217	98	5km	35m ²	<i>Einfam.</i>	Altbau	25	
b9	184	133	1km	15m ²	Einfam.	Altbau	20	Vorteil auf dem wichtigsten Merkmal, geringe Nachteile hinsichtlich der Miete sind unwesentlich; dominiert die schlechteren Alternativen hinsichtlich der Entfernung und dem Mietpreis; der große Preisunterschied gleicht den mittleren Unterschied in der Entfernung aus
	205	108	3km	15m ²	Einfam.	Altbau	15	
	233	58	5km	15m ²	Einfam.	Altbau	15	
	232	58	3km	15m ²	Einfam.	Altbau	25	
c3	126	185	5km	35m ²	Mehrfam.	Neubau	20	der Entfernungsnachteil wird durch mehr Größe und geringere Miete mehr als wettgemacht
	152	165	3km	25m ²	Mehrfam.	Neubau	25	
Lerncheck								
c4	132	178	1km	25m ²	<i>Einfam.</i>	<i>Altbau</i>	20	der renovierte Altbau stellt zwar mehr als eine Kompromißlösung zum Altbau dar, kann aber nicht den Nachteil auf zwei anderen (wichtigeren) Merkmalen ausgleichen
	139	174	1km	15m ²	<i>Mehrfam.</i>	<i>ren. Altb.</i>	20	
c6	186	130	3km	15m ²	<i>Wohnh.</i>	<i>Altbau</i>	25	das Wohnheim bietet gegenüber dem Mehrfamilienhaus offenbar mehr Vorzüge als mittlere Vorteile hinsichtlich des Zustands und der Miete
	195	124	3km	15m ²	<i>Mehrfam.</i>	<i>ren. Altb.</i>	20	
c9	159	155	3km	15m ²	<i>Wohnh.</i>	<i>Altbau</i>	20	das Wohnheim überwiegt die Nachteile der große Unterschied im Mietpreis ist wichtiger als die geringen Nachteile in der Art des Hauses und im Zustand
	170	149	3km	15m ²	<i>Mehrfam.</i>	<i>ren. Altb.</i>	15	
	190	128	3km	15m ²	<i>Einfam.</i>	<i>Neubau</i>	25	
c10	54	248	1km	25m ²	<i>Einfam.</i>	Neubau	20	dominiert die zweite Alternative hins. der Art des Hauses; die übrigen Merkmalsausprägungen sind gleich; Vorteile gegenüber anderen Alternativen hins. Entfernung, diese werden durch die Nachteile auf anderen Merkmalen nicht ausgeglichen; verglichen mit Alternative 3 wirkt der Entfernungsvorteil gegenüber dem Größenvorteil, verglichen mit Alternative 4 gegenüber dem Vorteil in der Art des Hauses gewichtiger; die mittleren Vorteile in der Entfernung und Größe wirken sich offenbar stärker aus als die deutlich schlechtere Art des Hauses
	64	240	1km	25m ²	<i>Mehrfam.</i>	Neubau	20	
	68	235	3km	35m ²	<i>Mehrfam.</i>	Neubau	20	
	86	220	5km	25m ²	<i>Wohnh.</i>	Neubau	20	

d1	66 147	239 170	1km 3km	35m ² 15m ²	Mehrfam. Mehrfam.	ren. Altb. Neubau	25 15	bessere Alternative, da auf den beiden wichtigsten Merkmalen überlegen und nur auf zwei weniger wichtigen schlechter
Lerncheck								
d3	69 81	230 223	1km 3km	15m ² 25m ²	Wohnh. Einfam.	Altbau Neubau	15 15	die Überlegenheit auf dem ersten und dritten Merkmal überwiegt die Nachteile auf dem zweiten und vierten Merkmal; es wird deutlich, daß das Einfamilienhaus gegenüber dem Mehrfamilienhaus nur sehr geringe Vorteile bietet
d5	65 164 175	240 152 145	5km 3km 1km	35m ² 25m ² 25m ²	Wohnh. Einfam. Mehrfam.	Neubau ren. Altb. Altbau	25 25 25	die Vorteile hins. Größe, Art des Hauses und Zustand überwiegen die ungünstigere Entfernung; der bessere Zustand und die etwas bessere Art des Hauses machen offenbar den mittleren Entfernungsnachteil wett
d6	143 149 209	173 169 103	5km 3km 3km	25m ² 25m ² 25m ²	Einfam. Mehrfam. Einfam.	Neubau ren. Altb. Altbau	15 20 25	die zweite Alternative wird dominiert, da auf drei Merkmalen überlegen und hins. Entfernung nur etwas schlechter; auch gegenüber der dritten Alternative leichte Nachteile hins. Entfernung, aber deutlich überlegen im Zustand und Mietpreis; der bessere Zustand und der günstigere Mietpreis bieten den eindeutig größeren Vorteil als das Einfamilienhaus gegenüber dem Mehrfamilienhaus
d7	26 176 193 203	289 145 125 115	1km 3km 1km 5km	35m ² 25m ² 15m ² 25m ²	Mehrfam. Mehrfam. Mehrfam. Mehrfam.	ren. Altb. Altbau Altbau Neubau	15 15 20 25	optimal hins. Entfernung, Größe und Mietpreis; der schlechtere Zustand im Vergleich zur 4. Alternative wird durch die Vorteile in anderen Merkmalen mehr als aufgewogen; Nachteile hins. Entfernung werden durch Größen- und Preisvorteile gegenüber Alternative 3 wettgemacht; gegenüber Alternative 4 bestehen geringe Vorzüge in der Entfernung und massive Vorzüge im Preis, der schlechtere Zustand ist somit unwesentlich; der massive Entfernungsvorteil und mittlere Preisvorteil bieten mehr Vorzüge als der mittlere Größenvorteil und der deutlich bessere Zustand der vierten Alternative
e2	158 175	157 145	3km 1km	15m ² 25m ²	Einfam. Mehrfam.	ren. Altb. Altbau	15 25	die Überlegenheit auf dem vierten und fünften Merkmal ist größer als die Überlegenheit der Alternativwohnung auf den ersten beiden Merkmalen; hinzu kommt die "etwas bessere" Art des Hauses
Lerncheck								

Anhang G Mündliche Instruktionen der Lernphase

Mündliche Instruktion zur Lerncheck-Aufgabe

Bisher wurden Dir jeweils zwei (drei oder vier) Wohnungsalternativen gezeigt. Deine Aufgabe war, herauszufinden, für welche Alternative sich Helga entschieden hat.

Nun erhältst Du sieben Wohnungsalternativen, jeweils eine Wohnung auf einer Papierkarte. Deine Aufgabe wird sein, diese sieben Karten entsprechend der Präferenz zu ordnen, die Du bei Helga vermutest. Versuche dabei bitte, Dich so gut wie möglich in Helga hineinzusetzen und eine Rangordnung nach ihren Vorstellungen und Wünschen zu erstellen. Ordne bitte die Wohnung, die Helga am meisten bevorzugt, oben an. Die Karte an der zweiten Stelle sollte die als nächste bevorzugte Wohnung enthalten, usw., bis zur untersten Karte. Sie enthält die Wohnung, die Helga am wenigsten akzeptieren würde.

Diese Aufgabe ist natürlich weitaus schwieriger als lediglich zwei Alternativen zu beurteilen. Ich bitte Dich deshalb, Dich gut zu konzentrieren und genau zu überlegen. Du solltest allerdings nicht länger als zehn Minuten an der Aufgabe sitzen.

Versuche bitte, alleine zu arbeiten. Du hast es nicht nötig abzuschauen, da Deine Leistung lediglich gemessen und nicht bewertet wird. Wir wissen nachher nicht mehr, welches Ergebnis zu welcher Person gehört. Das einzige, das wir feststellen können ist, wie gut die Leistungen im Durchschnitt sind. Das ist ein Zeichen dafür, wie gut wir Euch Helgas Wertvorstellungen bereits vermittelt haben.

Falls unbedingt notwendig, kannst Du Dir auf einem Blatt Notizen machen. Schreibe aber bitte nicht auf die Papierkarten, da diese noch benötigt werden.

Wenn Du die deiner Meinung nach richtige Lösung gefunden hast, kannst Du die Nummern, die auf den Karten vermerkt sind, in der entsprechenden Reihenfolge in den Computer eingeben. Die Untersuchungsleiter zeigen Dir, wenn Du so

weit bist, wie die Eingabe erfolgen soll. Der Computer gibt anschließend eine Rückmeldung über die Qualität der Lösung. Außerdem zeigt er Dir die wahre Rangfolge, die Helga angegeben hat. Bitte vergleiche die Rangfolge von Helga mit Deiner Lösung. Nimm Dir dazu ca. 5 Minuten und schaue, wo Du Fehler gemacht hast. Dadurch kannst Du weitere Einsichten in die Wertvorstellungen von Helga gewinnen.

Mündliche Instruktion zur dritten Lerncheck-Aufgabe

Ihr erhaltet nun wieder sieben Wohnungsalternativen, jeweils eine Wohnung auf einer Papierkarte. Versucht Euch bitte wieder so gut wie möglich in Helga hineinzusetzen und eine Rangordnung nach ihren Vorstellungen und Wünschen zu erstellen. Ordnet bitte die Wohnung, die Helga am meisten bevorzugt, oben an. Die Karte an der zweiten Stelle sollte die als nächste bevorzugte Wohnung enthalten, usw., bis zur untersten Karte. Sie enthält die Wohnung, die Helga am wenigsten akzeptieren würde.

Wahrscheinlich haben die meisten von Euch schon eine Strategie entwickelt, wie sie die Aufgabe am besten lösen können. Unsere Erfahrung hat gezeigt, daß es am günstigsten ist, wenn man die Wohnungen jeweils paarweise miteinander vergleicht. Man fängt mit zwei Wohnungen an, vergleicht diese, legt die bessere nach oben und die schlechtere nach unten. Anschließend nimmt man eine dritte Karte und vergleicht diese mit der oberen Wohnung. Ist die Wohnung der dritten Karte besser, wird sie darüber angeordnet, ist sie schlechter, vergleicht man sie mit der unteren Karte. Ist sie besser, legt man sie darüber, also in die Mitte, ist sie schlechter, kommt die Karte nach unten. Genauso wird dann mit den weiteren vier Karten verfahren.

Wenn jemand die seiner Meinung nach richtige Lösung gefunden hat, kann er die Nummern, die auf den Karten vermerkt sind, in der entsprechenden Reihenfolge in den Computer eingeben. Der Computer gibt anschließend wieder eine Rückmeldung über die Qualität der Lösung.

Anhang H Mündliche Instruktionen der Untersuchungsphase

Mündliche Instruktion zur Vollprofil-CA

Anstatt der sieben Wohnungsprofile erhaltet Ihr nun 15 Wohnungsprofile, die in eine Rangordnung gebracht werden sollen.

Am günstigsten ist es, wenn man die Karten zunächst intuitiv in drei Gruppen à fünf Karten ordnet. Eine Gruppe mit Wohnungen, die Helga gut findet, eine Gruppe mit Wohnungen, die etwas schlechter sind und eine dritte mit den schlechtesten. Anschließend könnt Ihr dann wie in der Lernaufgabe innerhalb der Gruppen die Wohnungen paarweise miteinander vergleichen und so für jede Gruppe eine Rangordnung erstellen. Danach fügt Ihr die Gruppen zusammen. Zur Kontrolle bietet es sich an, die Übergänge zwischen den Gruppen noch einmal zu kontrollieren, d.h. die dort benachbarten Wohnungen miteinander zu vergleichen und gegebenenfalls die Rangordnung zu korrigieren.

Legt bitte wie bei der Lernaufgabe die Karte mit der von Helga am meisten bevorzugten Wohnung oben hin. Die zweite Wohnung darunter usw.

Nach 30 Minuten solltet Ihr damit fertig sein. Ich kann Euch nur bitten, auch hierbei alleine zu arbeiten und Euch gut zu konzentrieren, da die Aufgabe natürlich noch schwieriger als die Rangordnung mit sieben Karten ist.

Die Lösung könnt Ihr anschließend auf dem Tisch liegenlassen; Ihr braucht sie also nicht in den Computer einzugeben. Bitte meldet Euch per Handzeichen beim Untersuchungsleiter, wenn Ihr so weit seid.

Mündliche Instruktion zu den Paarvergleichen der adaptiven CA und den Fragebogenteilen

Die nächste Aufgabe wird am Computer stattfinden. Um die Aufgabe zu verstehen, setzt sich am besten jeder an seinen Computer.

Die Aufgabe besteht aus insgesamt drei Untersuchungsabschnitten. Davon ist jeder Abschnitt zweigeteilt. Der erste Teil trägt jeweils die Bezeichnung 'Paarvergleich', der zweite die Bezeichnung 'Fragen'. Zum Üben gibt es gleich zu den Paarvergleichen und den Fragen jeweils eine Beispielaufgabe.

Die Zahl hinter 'Paarvergleich' und 'Fragen' gibt den Untersuchungsabschnitt an. Wir beginnen also nachher mit dem Paarvergleich 1. Anschließend werden die Fragen 1 bearbeitet. Danach macht jeder individuell, also je nachdem, wann er fertig ist, zehn Minuten Pause. Nach der Pause folgen dann Paarvergleich 2 und Fragen 2. Dann hat wieder jeder zehn Minuten Zeit, sich zu erholen, bevor Paarvergleich 3 und Fragen 3 bearbeitet werden.

Ihr könnt die Aufgabe, die bearbeitet werden soll, mit der Maus auswählen. Durch einen „Doppelklick“ mit der linken Maustaste läßt sich die Aufgabe starten. Wenn die Aufgabe beendet ist, kehrt das Programm automatisch zum Ausgangsbildschirm zurück. Auf die gleiche Weise könnt Ihr dann die nächste Aufgabe auswählen.

Bitte klickt einmal alle den Beispiel-Paarvergleich an.

Die Paarvergleiche werden folgendermaßen aussehen (VISUALISIERT):

Wie zuvor auf den Folien, werden Euch nacheinander jeweils zwei Wohnungsalternativen gezeigt. Die Wohnungen werden nicht anhand aller fünf Merkmale beschrieben sein. Die jeweils fehlenden Merkmale sind als konstant anzusehen, d.h. ihre Ausprägung ist bei beiden Alternativen gleich.

Eure Aufgabe ist, einzuschätzen, welche Alternative Helga bevorzugt. Dafür steht Euch eine Skala mit einem Schieber zur Verfügung. Der Schieber befindet sich in der Bildmitte, also zwischen den beiden Alternativen, läßt sich jedoch nach rechts und links verschieben. Je stärker Helga eine Wohnungsalternative gegenüber der anderen bevorzugt, desto mehr solltet Ihr den Schieber in die eine Richtung verschieben. Bleibt der Schieber in der Mitte, bedeutet dies, daß Helga diesen beiden Wohnungen indifferent gegenübersteht.

Bedienen läßt sich das ganze Programm mit der linken Maustaste. Ihr könnt die Maus zum Schieber bewegen, die linke Maustaste drücken, festhalten und den Schieber mit der Maus dorthin ziehen, wo Ihr ihn hinhaben möchtet. Wer damit Probleme hat, kann auch den Schieber anklicken und dann die grauen Pfeiltasten verwenden, um den Schieber zu bewegen. Ihr könnt das jetzt einmal ausprobieren.

Wenn Ihr der Meinung seid, daß Ihr fertig seid und der Schieber an der richtigen Stelle ist, klickt Ihr mit der Maus das Feld „WEITER“ an; dann erscheint der nächste Paarvergleich. Mit dem Feld „Letzte Frage“ könnt Ihr notfalls zu einem früheren Paarvergleich zurückkehren, wenn Ihr glaubt, einen Fehler gemacht zu haben oder versehentlich „WEITER“ gedrückt habt, obwohl Ihr noch nicht so weit wart.

Ihr könnt die Beispiele jetzt bearbeiten. Es ist sehr wichtig, daß Ihr Euch bei den Aufgaben richtig anstrengt, da die Ergebnisse sonst wertlos für uns sind. Das wäre ärgerlich, da Ihr Eure wertvolle Zeit in die Aufgabe investiert habt. Ihr habt letztendlich nichts davon und wir noch weniger. Es ist sicher nicht ganz einfach, sich die ganze Zeit 100-prozentig zu konzentrieren. Für diese Mühe bekommt Ihr allerdings eine angemessene Anzahl Versuchspersonenstunden angerechnet.

Hinzu kommt ein weiterer Vorteil: Wenn das Computerprogramm eine bestimmte Informationsmenge erhalten hat, ist die Aufgabe beendet. Das bedeutet zweierlei:

1. Je genauer Ihr die Paarvergleiche beantwortet, desto weniger Paarvergleiche verlangt das Computerprogramm von Euch.
2. Je mehr Paarvergleiche Ihr in einer bestimmten Zeit schafft, desto schneller seid Ihr fertig.

Es lohnt sich also auch aus diesem Grund, wenn ihr Euch sehr konzentriert. Ergebnisse aus ähnlichen Untersuchungen haben gezeigt, daß man nicht mehr als eine halbe bis eine Minute mit einem Paarvergleich zubringen sollte, da man sonst nicht mehr genügend geistige Reserven für die letzten Paarvergleiche hat.

Neben dem Beispiel-Paarvergleich befinden sich die Beispiel-Fragen. Ich bitte Euch, diese zu starten.

Die Fragen werden folgendermaßen aussehen (VISUALISIERT):

Im zweiten Teil eines jeden Abschnitts werden Euch einige Fragen zum ersten Teil, also zu den zuletzt bearbeiteten Paarvergleichen, gestellt. Dabei geht es z.B. um die Schwierigkeit der Aufgaben und ob es Euch leicht oder schwer gefallen ist, Euch zu konzentrieren. Bedient wird das Fragebogenprogramm durch den „Nummern-Block“, das ist der Zahlenblock rechts auf der Tastatur. Es gibt drei Typen von Fragen:

1. Einfache Frage: Ihr werdet z.B. gefragt, wie schwierig die Paarvergleiche waren. Als mögliche Antwort sind verschiedene Alternativen vorgegeben, aus denen Ihr eine auswählt. Das könnt Ihr machen, indem Ihr die Zahl, die vor der Alternative steht, durch den „Nummern-Block“ in den Computer eingibt. Das Programm springt dann automatisch weiter zur nächsten Frage. Drückt also bitte nicht die ENTER-Taste.

Ihr könnt das ausprobieren, indem Ihr jetzt eine Zahl zwischen eins und vier eingibt.

2. Numerische Frage: Ihr werdet z.B. gefragt, wie lange der Aufgabenabschnitt Eurer Meinung nach gedauert hat. Dabei kommt es nicht darauf an, daß Ihr eine gute Schätzung abgebt, sondern, daß wir erfahren, ob Ihr die Aufgabe als lang oder eher kurz empfunden habt. Schaut also bitte nicht zur Uhr. Wenn Ihr die Minutenzahl eingegeben habt, könnt Ihr die ENTER-Taste drücken, um zur nächsten Frage zu gelangen.

3. Rangordnungsfrage: Beispielsweise werdet Ihr in Fragen 3, danach befragt, welche der drei Paarvergleichsaufgaben am schwierigsten war, welche am zweitschwierigsten und welche am einfachsten. Ihr gebt dann zuerst die Zahl der Alternative ein, die Ihr am schwierigsten empfindet, als zweites die Zahl der zweitschwierigsten und zuletzt die Zahl der einfachsten. Wenn Ihr die letzte Zahl eingegeben habt, erscheint automatisch die nächste Frage. Ich bitte Euch, also auch in diesem Fall nicht die ENTER-Taste zu drücken. Versucht bitte einmal, eine Rangfolge zu erstellen.

Es ist sehr wichtig, daß Ihr die Fragen offen und ehrlich beantwortet. Wenn Ihr also beispielsweise merkt, daß die Konzentration doch einmal nicht optimal war, antwortet bitte entsprechend. Wir messen die Anstrengungsleistung etc. zwar auch noch mit anderen Methoden, wenn es dabei Widersprüche gibt, wissen wir aber nicht, ob Ihr Euch nun tatsächlich angestrengt habt oder nicht bzw. ob es mit den Aufgaben irgendwelche Probleme gab. Überlegt Euch bitte gut, welche Antwort Ihr geben wollt, da es nicht möglich ist, zu einer früheren Frage zurückzukehren.

Wie schon erwähnt, machen wir nach jedem Abschnitt eine kurze Pause. Wenn ihr mit einem Abschnitt fertig seid, könnt ihr eure Pause machen. Wenn alle mit den drei Abschnitten fertig sind, fangen wir gemeinsam mit der nächsten, d.h. letzten Aufgabe, an.

Anhang I Items der Zwischenfragebögen und des Abschlußfragebogens

Tabelle I-1: Die Items der (identischen) Zwischenfragebögen.

Nr.	Frage	Antwortmöglichkeiten	Code
1	Die Paarvergleichsaufgaben waren ...	schwierig eher schwierig eher einfach einfach	1 2 3 4
2	Die Paarvergleichsaufgaben fand ich insgesamt ...	interessant eher interessant eher langweilig langweilig	1 2 3 4
3	Die Paarvergleichsaufgaben fand ich im ersten Drittel der Teilaufgabe ...	interessant eher interessant eher langweilig langweilig	1 2 3 4
4	Die Paarvergleichsaufgaben fand ich im zweiten Drittel der Teilaufgabe ...	interessant eher interessant eher langweilig langweilig	1 2 3 4
5	Die Paarvergleichsaufgaben fand ich im letzten Drittel der Teilaufgabe ...	interessant eher interessant eher langweilig langweilig	1 2 3 4
6	Wie schwer ist es Dir gefallen, Dich zu konzentrieren?	schwer eher schwer eher leicht leicht	1 2 3 4
7	Wie schwer ist es Dir im ersten Drittel der Wohnungsvergleichsaufgabe gefallen, Dich zu konzentrieren?	schwer eher schwer eher leicht leicht	1 2 3 4
8	Wie schwer ist es Dir im zweiten Drittel der Wohnungsvergleichsaufgabe gefallen, Dich zu konzentrieren?	schwer eher schwer eher leicht leicht	1 2 3 4
9	Wie schwer ist es Dir im letzten Drittel der Wohnungsvergleichsaufgabe gefallen, Dich zu konzentrieren?	schwer eher schwer eher leicht leicht	1 2 3 4
10	Eine Antwort zu finden fiel mir insgesamt ...	schwer eher schwer eher leicht leicht	1 2 3 4
11	Eine Antwort zu finden fiel mir im ersten Drittel ...	schwer eher schwer eher leicht leicht	1 2 3 4
12	Eine Antwort zu finden fiel mir im zweiten Drittel ...	schwer eher schwer eher leicht leicht	1 2 3 4

13	Eine Antwort zu finden fiel mir im letzten Drittel ...	schwer eher schwer eher leicht leicht	1 2 3 4
14	Wie viele Minuten haben die Wohnungsvergleichsaufgaben etwa gedauert?		
15	Wie viele Minuten entfielen von der Gesamtzeit auf das erste Drittel?		
16	Wie viele Minuten entfielen von der Gesamtzeit auf das zweite Drittel?		
17	Wie viele Minuten entfielen von der Gesamtzeit auf das letzte Drittel?		
18	Wie beurteilst Du Deine Leistung bei den Paarvergleichen, gemessen an Deiner Leistung in der letzten 7-Karten-Rangordnungsaufgabe in der Lernphase?	besser eher besser gleich eher schlechter schlechter	1 2 3 4 5
19	Wie viele falsche Antworten schätzt Du gegeben zu haben?	weniger als 2 2 bis 4 5 bis 7 mehr als 7	1 2 3 4

Tabelle I-2: Die Items des Abschlußfragebogens.

Nr.	Frage	Antwortmöglichkeiten	Code
20	Bitte ordne die drei Aufgabenabschnitte (Wohnungen mit zwei, drei und vier Merkmalen) hinsichtlich ihrer Schwierigkeit. Beginne mit dem Schwierigsten.	zwei Merkmale drei Merkmale vier Merkmale	
21	Bitte ordne die drei Aufgabenabschnitte (Wohnungen mit zwei, drei und vier Merkmalen) hinsichtlich ihrer Interessantheit. Beginne mit dem Interessantesten.	zwei Merkmale drei Merkmale vier Merkmale	
22	Bitte ordne die drei Aufgabenabschnitte (Wohnungen mit zwei, drei und vier Merkmalen). In welchem Abschnitt fiel es Dir am schwersten (am zweitschwersten ...), Dich zu konzentrieren?	zwei Merkmale drei Merkmale vier Merkmale	
23	Bitte ordne die drei Aufgabenabschnitte (Wohnungen mit zwei, drei und vier Merkmalen). In welchem Abschnitt fiel es Dir am schwersten (am zweitschwersten ...), eine Antwort zu finden?	zwei Merkmale drei Merkmale vier Merkmale	
24	Ist Dein erstes Studienfach Psychologie oder ein anderes?	Psychologie anderes Fach	1 2
25	In welchem Semester studierst Du dieses Fach?		
26	Wie alt bist Du?		
27	Gib bitte Dein Geschlecht an.	männlich weiblich	1 2
28	Welche Gesamtnote hattest Du im Abitur?		
29	Wie schwierig fandest Du die Bedienung des Computerprogramms (z.B. Schieber)?	schwierig eher schwierig eher einfach einfach	1 2 3 4

Anhang J LPS-Skala 12 zur Erfassung der allgemeinen Intelligenz

Mündliche Instruktion

Bitte blättert das erste Blatt nicht um, bevor ich Euch dazu auffordere. Seht Euch die Aufgabengruppe 12 bitte an. Es geht nur um diese Spalte, die anderen Aufgaben interessieren uns nicht. Ich werde Euch gleich erklären, was Eure Aufgabe ist und danach zum Umblättern auffordern. Die zweite Seite sieht genau so aus wie die erste, außer daß dort statt der großen weißen Fläche die Aufgaben zu sehen sind. Wenn ich nachher also zum Umblättern auffordere, beginnt bitte sofort die Aufgabe 12 zügig zu bearbeiten.

Was Ihr dann zu tun habt, erkläre ich Euch jetzt. Seht Euch bitte unter der 12 das erste Wort aus verstümmelten Buchstaben an.

[ORIGINALTEXT HANDANWEISUNG LPS]

Wie heißt das erste Wort aus verstümmelten Buchstaben? - Es heißt Lappem. Einen Lappem gibt es nicht, sondern es müßte Lappen mit n am Ende heißen. Das m ist offensichtlich ein Druckfehler und wurde deshalb durchgestrichen. - Das zweitoberste Wort aus verstümmelten Buchstaben heißt Valter mit V. Es gibt ein Falter mit F oder ein Walter mit W am Anfang. Der erste Buchstabe ist also in jedem Fall ein Druckfehler und wurde deshalb durchgestrichen. Unter den beiden Wörtern stehen *auf der zweiten Seite* [nicht original] 40 einfache Hauptwörter in der Einzahl. Es versucht jetzt jeder, diese Wörter aus verstümmelten Buchstaben möglichst schnell zu erraten. Der eine falsche Buchstabe in jedem Wort wird kräftig durchgestrichen.

Es sind zwei Minuten Zeit. Wer noch nicht verstanden hat, was er tun soll, hebt seinen Arm hoch!

Bitte umblättern und sofort anfangen.

[Nach genau zwei Minuten] Halt! Bitte Stift hinlegen.

Fragebogen

siehe Horn (1962)

Anhang K MARPS-Skala zur Erfassung der Leistungsmotivation

Mündliche Instruktion

In dem Fragebogen vor Euch findet Ihr 20 Satzanfänge mit jeweils zwei Satzenden. Lest bitte immer einen Satzanfang und beide Satzenden. Kreuzt danach die Alternative an, die eher auf Euch zutrifft. Es gibt keine richtigen und falschen Antworten, wir möchten von jedem wissen, was auf ihn ganz persönlich eher zutrifft. Laßt bitte keine Frage aus. Sollte keine der beiden Alternativen so recht zutreffen, so wählt bitte diejenige aus, die noch am ehesten paßt. Wenn Ihr Euch einmal vertan haben solltet, so kreuzt das andere Kästchen bitte deutlich stärker an als das fälschlich angestrichene. Bitte nehmt Euch jetzt ein paar Minuten Zeit, den Fragebogen auszufüllen. Wenn noch etwas unklar ist, so fragt bitte.

Fragebogen

Im folgenden findest Du 20 Satzanfänge, zu denen jeweils zwei Satzvervollständigungen angegeben sind. Lies bitte immer einen Satzanfang und beide Satzenden. Kreuze danach die Alternative an, die eher auf Dich zutrifft. Lasse bitte keine Frage aus, also kreuze, auch wenn Du keiner der beiden Alternativen völlig zustimmst, diejenige an, die noch eher zutreffend ist.

1. Wenn ich von anderen im Gespräch unterbrochen werde
 - a) setze ich mich dagegen zur Wehr
 - b) macht mir das nur wenig aus
2. Wenn ich mit Freunden eine Party veranstalte
 - a) organisiere ich das am liebsten selbst (x)
 - b) überlasse ich die Organisation den anderen
3. Wenn ich zwei Arbeiten nicht erledigt habe und ich habe nur für eine Zeit, dann
 - a) entscheide ich mich für die leichtere Aufgabe
 - b) entscheide ich mich für die schwerere Aufgabe (x)
4. Wenn ich in eine neue Gesellschaft komme
 - a) halte ich mich eher im Hintergrund
 - b) bemühe ich mich, viele Leute kennenzulernen
5. In meiner Freizeit
 - a) lese ich lieber Bücher, aus denen ich etwas lernen kann (x)
 - b) gehe ich lieber ins Kino

6. Ich verwende meine Zeit lieber
 - a) für mehrere Sachen, die schnell getan sind
 - b) für eine einzige Sache, die genau so lange dauert (x)
7. Ich mache mir im allgemeinen mehr Gedanken über Menschen
 - a) die ich gern habe
 - b) die ich nicht gern mag
8. Mir ist eine Arbeit lieber
 - a) bei der ich selbst entscheide, was zu tun ist (x)
 - b) bei der mir genau vorgeschrieben wird, was ich zu tun habe
9. Wenn mich jemand auf einen Fehler aufmerksam macht
 - a) bin ich ihm dankbar
 - b) ist mir das unangenehm
10. Wenn ich etwas nicht gut kann
 - a) strengere ich mich an, um es doch zu schaffen (x)
 - b) lasse ich es sein und mache etwas anderes
11. Mir sind Kartenspiele lieber
 - a) bei denen man viel nachdenken muß (x)
 - b) bei denen es lustig zugeht
12. Wenn mir jemand seine Probleme erzählt
 - a) freue ich mich über sein Vertrauen
 - b) ist mir das peinlich
13. In meiner Freizeit
 - a) lerne ich gern etwas Neues dazu (x)
 - b) will ich mich einfach erholen
14. Ich arbeite lieber an einer Aufgabe
 - a) für die ich allein verantwortlich bin (x)
 - b) für die neben mir auch andere die Verantwortung tragen
15. Ich würde eher etwas tun
 - a) was mir Freude macht, aber nicht so schwierig ist
 - b) was wichtig ist, mir aber weniger Freude macht (x)
16. Wenn mir jemand vorgestellt wird, beurteile ich ihn eher
 - a) nach seinem Äußeren
 - b) nach seinem Verhalten
17. Ich bevorzuge Tätigkeiten
 - a) die anspruchslos sind und mir leicht von der Hand gehen
 - b) die anspruchsvoll sind und wo ich mich ganz einsetzen muß (x)
18. Wenn ich Besuch bekomme
 - a) freue ich mich über die Abwechslung
 - b) fühle ich mich in meiner Ruhe gestört
19. Wenn ich gleich viel verdiene, würde ich lieber
 - a) als Selbständiger arbeiten (x)
 - b) als Angestellter arbeiten
20. Ich arbeite lieber
 - a) an schweren Aufgaben (x)
 - b) an leichten Aufgaben

Die Antworten, die eine hohe Leistungsmotivation indizieren, sind hier mit (x) gekennzeichnet. Der Fragebogen beinhaltet 7 Füllitems.

Anhang L Ausführliche Ergebnisdarstellung der zentralen Varianzanalysen und Kovarianzanalysen

Tabelle L-1: Die Ergebnisse der (2X4X3)-ANOVA aus Abschnitt 6.5.2.6.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	27.23	23	1.18		
CONSTANT	706.47	1	706.47	596.64	.000
WITHIN+RESIDUAL	6.49	23	.28		
HOLD-SET	.01	1	.01	.05	.826
WITHIN+RESIDUAL	3.76	69	.05		
PAARVGL	15.25	3	5.08	93.30	.000
WITHIN+RESIDUAL	10.21	46	.22		
PROFIL	.98	2	.49	2.20	.123
WITHIN+RESIDUAL	1.44	69	.02		
HOLD-SET BY PAARVGL	.39	3	.13	6.23	.001
WITHIN+RESIDUAL	3.39	46	.07		
HOLD-SET BY PROFIL	.59	2	.29	4.00	.025
WITHIN+RESIDUAL	4.93	138	.04		
PAARVGL BY PROFIL	.40	6	.07	1.87	.090
WITHIN+RESIDUAL	2.24	138	.02		
HOLD-SET BY PAARVGL BY PROFIL	.40	6	.07	4.06	.001

Tabelle L-2: Die Ergebnisse der (4X3)- ANOVA aus Abschnitt 6.5.2.6 (kriterienorient. Validität zu Holdout-Set 1).

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	16.45	23	.72		
CONSTANT	356.39	1	356.39	498.18	.000
WITHIN+RESIDUAL	2.59	69	.04		
PAARVGL	10.24	3	3.41	91.05	.000
WITHIN+RESIDUAL	7.04	46	.15		
PROFIL	.05	2	.02	.15	.862
WITHIN+RESIDUAL	3.79	138	.03		
PAARVGL BY PROFIL	.37	6	.06	2.27	.040

Tabelle L-3: Die Ergebnisse der (4X3)-ANOVA aus Abschnitt 6.5.2.6 (kriterienorient. Validität zu Holdout-Set 2).

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	17.27	23	.75		
CONSTANT	350.10	1	350.10	466.13	.000
WITHIN+RESIDUAL	2.61	69	.04		
PAARVGL	5.40	3	1.80	47.56	.000
WITHIN+RESIDUAL	6.56	46	.14		
PROFIL	1.52	2	.76	5.33	.008
WITHIN+RESIDUAL	3.38	138	.02		
PAARVGL BY PROFIL	.42	6	.07	2.88	.011

Tabelle L-4: Die Ergebnisse der (4X3)-ANOVA aus Abschnitt 6.5.2.7 (Strukturwiedererkennung).

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	494127	23	21483		
CONSTANT	15549889	1	15549890	723.80	.000
WITHIN+RESIDUAL	150913	69	2187		
PAARVGL	1404356	3	468118	214.03	.000
WITHIN+RESIDUAL	495212	46	10765.49		
PROFIL	84696	2	42348	3.93	.026
WITHIN+RESIDUAL	220275	138	1596		
PAARVGL BY PROFIL	38630	6	6438	4.03	.001

Tabelle L-5: Die Ergebnisse der (3X7)-ANOVA aus Abschnitt 6.5.2.9 (Strukturwiedererkennung/ Antwortzeit).

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	965788	23	41990		
CONSTANT	49633940	1	49633940	1182	.000
WITHIN+RESIDUAL	1064648	46	23144		
PROFIL	1406659	2	703329	30.39	.000
WITHIN+RESIDUAL	297070	138	2152		
ZEITPKT	9310642	6	1551773	720.86	.000
WITHIN+RESIDUAL	526684	276	1908		
PROFIL BY ZEITPKT	527704	12	43975	23.04	.000

Tabelle L-6: Die Ergebnisse der (4X3)-ANCOVA aus Abschnitt 6.5.2.12 mit der Kovariate 'LPS-Wert'.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	486659	22	22120		
REGRESSION	7467	1	7467	.34	.567
CONSTANT	243895	1	243895	11.03	.003
WITHIN+RESIDUAL	150913	69	2187		
PAARVGL	1404356	3	468118	214.03	.000
WITHIN+RESIDUAL	495212	46	10765		
PROFIL	84696	2	42348	3.93	.026
WITHIN+RESIDUAL	220275	138	1596		
PAARVGL BY PROFIL	38630	6	6438	4.03	.001
COVARIATE	B	Beta	Std.Err.	t-Value	Sig. of t
LPS	1.32714	.12294	2.284	.581	.567

Tabelle L-7: Die Ergebnisse der (4X3)-ANCOVA aus Abschnitt 6.5.2.12 mit der Kovariate 'Abiturnote'.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	466929	21	22234		
REGRESSION	26376	1	26376	1.19	.288
CONSTANT	998369	1	998369	44.90	.000
WITHIN+RESIDUAL	148498	66	2249		
PAARVGL	1326943	3	442314	196.59	.000
WITHIN+RESIDUAL	489734	44	11130		
PROFIL	89562	2	44781	4.02	.025
WITHIN+RESIDUAL	214643	132	1626		
PAARVGL BY PROFIL	33876	6	5646	3.47	.003
COVARIATE	B	Beta	Std.Err.	t-Value	Sig. of t
ABINOTE	15.22901	.23123	13.982	1.089	.288

Tabelle L-8: Die Ergebnisse der (4X3)-ANCOVA aus Abschnitt 6.5.2.12 mit der Kovariate 'MARPS'.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	493167	22	22416		
REGRESSION	960	1	960	.04	.838
CONSTANT	1093615	1	1093615	48.79	.000
WITHIN+RESIDUAL	150913	69	2187		
PAARVGL	1404356	3	468118	214.03	.000
WITHIN+RESIDUAL	495212	46	10765		
PROFIL	84696	2	42348	3.93	.026
WITHIN+RESIDUAL	220275	138	1596		
PAARVGL BY PROFIL	38630	6	6438	4.03	.001
COVARIATE	B	Beta	Std.Err.	t-Value	Sig. of t
MARPS	-.92337	-.04408	4.462	-.207	.838

Literatur

- Acito, F. (1977). An investigation of some data collection issues in conjoint measurement. In B.A. Greenberg & D.N. Bellenger (Eds.), *Contemporary marketing thought* (82-85). Chicago: Proceedings. American Marketing Association.
- Acito, F. (1979). An investigation of the reliability of conjoint measurement for various orthogonal designs. In R.S. Franz, R.M. Hopkins, & A. Toma (Eds.), *Proceedings, Southern Marketing Association 1979 Conference* (175-178). Lafayette: University of Southwestern Louisiana.
- Acito, F. & Jain, A.K. (1980). Evaluation of conjoint analysis results: A comparison of methods. *Journal of Marketing Research*, 17, 106-112.
- Acito, F. & Olshavsky, R.W. (1980). Limits to accuracy in conjoint analysis. *Advances in Consumer Research*, 8, 313-316.
- Addelman, S. (1962a). Orthogonal main-effect plans for asymmetrical factorial experiments. *Technometrics*, 4, 21-46.
- Addelman, S. (1962b). Symmetrical and asymmetrical fractional factorial designs. *Technometrics*, 4, 47-58.
- Agarwal, M.K. (1983). A Monte Carlo study of ties in configuration recovery using conjoint analysis. *Proceedings of the 1983 AMA Educators Conference*, 447-451.
- Agarwal, M.K. (1988a). A monte carlo study investigating configuration recovery in adaptive conjoint analysis. In *1988 AMA educator's proceedings* (series 54, 292-296). Chicago, IL: American Marketing Association.
- Agarwal, M.K. (1988b). Comparison of conjoint methods. *Proceedings of the Sawtooth Software Conference of Perceptual Mapping, Conjoint Analysis, and Computer Interviewing* (51-57). Ketchum, ID: Sawtooth Software, Inc.
- Agarwal, M. (1988c). An empirical comparison of traditional conjoint and adaptive conjoint analysis. Working Paper No. 88-140, School of Management, State University of New York at Binghamton.
- Agarwal, M.K. (1989). How many pairs should we use in adaptive conjoint analysis? An empirical analysis. In *AMA Winter Educators' Conference Proceedings* (7-11). Chicago: American Marketing Association.
- Agarwal, M.K. & Green, P.E. (1989). Adaptive conjoint analysis versus self-explicated models: Some empirical results. Working paper, State University of New York at Binghamton.
- Ajzen, I. & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, 84, 888-918.

- Akaah, I.P. & Korgaonkar, P.K. (1983). An empirical comparison of the predictive validity of self-explicated, huber-hybrid, traditional conjoint, and hybrid conjoint models. *Journal of Marketing Research*, 20, 187-197.
- Alpert, M.I. (1971). Identification of determinant attributes: A comparison of methods. *Journal of Marketing Research*, 7, 184-191.
- Alpert, M.I. (1980). Unresolved issues in identification of determinant attributes. In J.C. Olson (Ed.), *Advances in consumer research* (Vol. 7, 83-88). Ann Arbor: Association for Consumer Research.
- Alpert, M.I., Betak, J.F. & Golden, L.L. (1978). Data gathering issues in conjoint measurement. Working Paper, Graduate School of Business, The University of Texas, Austin.
- Anderson, J.C. & Donthu, N. (1988). A proximate assessment of the external validity of conjoint analysis. In 1988 AMA educator's proceedings (series 54, 287-291). Chicago, IL: American Marketing Association.
- Anderson, N.H. (1970). Functional measurement and psychological judgment. *Psychological Review*, 77, 153-170.
- Anderson, N.H. (1971). Integration theory and attitude change. *Psychological Review*, 78, 177-206.
- Anderson, N.H. (1981). *Foundations of information integration theory*. New York: Academic Press.
- Anderson, N.H. (1982). *Methods of information integration theory*. New York: Academic Press.
- Anttila, M, van den Heuvel, R.R. & Möller, K. (1980). Conjoint measurement for marketing management. *European Journal of Marketing*, 14, 397-408.
- Atkinson, J.W. (1964). *An introduction to motivation*. New York: Van Nostrand.
- Backhaus, K., Erichson, B., Plinke, W. & Weiber R. (1994). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung* (7. Aufl., 433-497). Berlin: Springer.
- Banks, S. (1950). The relationship between preference and purchase of brands. *Journal of Marketing*, 15, 145-157.
- Bateson, J.E.G., Reibstein, D.J. & Boulding, W. (1987). Conjoint analysis reliability and validity: A framework for future research. In M.J. Houston (Ed.), *Review of Marketing* (451-481). Chicago: American Marketing Association.
- Benbenisty, R.L. (1983). Attitude research, conjoint analysis guided Ma Bell's entry into data terminal market. *Marketing News*, 13 (May), 12.
- Bettman, J.R. (1979). *An information processing theory of consumer choice*. Reading, MA: Addison-Wesley.

- Bock, R.O. & Jones, L.V. (1968). *The Measurement and prediction of judgment and choice*. San Francisco: Holden-Day.
- Böcker, F. (1986). Präferenzforschung als Mittel marktorientierter Unternehmensführung. *Zeitschrift für betriebswirtschaftliche Forschung*, 7/8, 543-574.
- Borcherding, K. (1983). Entscheidungstheorie und Entscheidungshilfverfahren für komplexe Entscheidungssituationen. In M. Irle & W. Bussmann (Hrsg.), *Methoden und Anwendungen in der Marktpsychologie*. Enzyklopädie der Psychologie D/III/5 (64-173). Göttingen: Hogrefe.
- Bradley, R.A. & Terry, M.E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparison. *Biometrika*, 39, 324-345.
- Bringmann, M.W. (1992). Computer-based methods for the analysis and interpretation of personal construct systems. In R.A. Neimeyer & G.J. Neimeyer (Eds.), *Advances in Personal Construct Psychology*, (Vol. 2, 57-90). London: JAI.
- Brown, H.L. & Kirsner, K. (1980). A within-subject analysis of the relationship between memory span and processing rate in short-term memory. *Cognitive Psychology*, 12, 177-187.
- Bucklin, R.E. & Srinivasan, V. (1991). Determining interbrand substitutability through survey measurement of consumer preference structures. *Journal of Marketing Research*, 28, 58-71.
- Carmone, F. (1987). ACA system for adaptive conjoint analysis. *Journal of Marketing Research*, 24, 325-327.
- Carmone, F.J. & Green, P.E. (1981). Model misspecification in multiattribute parameter estimation. *Journal of Marketing Research*, 18, 87-93.
- Carmone, F.J., Green, P.E. & Jain, A.K. (1978). Robustness of conjoint analysis: Some Monte Carlo results. *Journal of Marketing Research*, 15, 300-303.
- Carroll, J.D. (1969). *Categorical conjoint measurement*. Ann Arbor: Meeting of Mathematical Psychology.
- Carroll, J.D. (1972). Individual differences and multidimensional scaling. In R.N. Shepard, A.K. Romney & S. Nerlove (Eds.), *Multidimensional scaling: Theory and application in the behavioral sciences* (Vol. 1). New York: Seminar Press.
- Carroll, J.D. (1973). Models and algorithms for multidimensional scaling, conjoint measurement, and related techniques. In P.E. Green & Y. Wind (Eds.). *Multivariate decisions in marketing: A measurement approach*. (299-371). Hinsdale: The Dryden Press.
- Cattin, P. & Bliemel, F. (1978). Metric vs. nonmetric procedures for multiattribute modeling: Some simulation results. *Decision Sciences*, 9, 474-480.
- Cattin, P., Hermet, G. & Pioche, A. (1982). Alternative hybrid models for conjoint analysis: Some empirical results. In *Analytical Approaches to Product and*

- Market Planning: The Second Conference (44-53). Cambridge, MA: Marketing Science Institute.
- Cattin, P. & Weinberger, M. (1980). Some validity and reliability issues in the measurement of attribute utilities. In J. Olson (Ed.), *Advances in consumer research*, (Vol. 7, 780-783). Ann Arbor, MI: Association for Consumer Research.
- Cattin, P. & Wittink, D.R. (1976). A Monte Carlo study of metric and nonmetric estimation methods for multiattribute models. Research Paper No. 341. Graduate School of Business, Stanford University.
- Cattin, P. & Wittink, D.R. (1981). Commercial use of conjoint analysis: A survey. In J.W. Keon (Ed.), *Marketing: Measurement and analysis 1981*, Proceedings of the Third ORSA/ TIMS special interest conference on market measurement and analysis (21-31).
- Cattin, P. & Wittink, D.R. (1982). Commercial use of conjoint analysis: A survey. *Journal of Marketing*, 46 (Summer), 44-53.
- Chapman, R.G. & Bolton, R.N. (1985). Attribute presentation order bias and nonstationarity in full profile conjoint analysis tasks. In R.F. Lusch et al. (Eds.), *1985 AMA Educators' Proceedings* (373-379). Chicago: American Marketing Association.
- Chrzan, K. & Grisaffe, D.B. (1992). A comparison of telephone conjoint analysis with full profile conjoint analysis and adaptive conjoint analysis. Proceedings of the Sawtooth Software Conference (225-242). Sun Valley, ID: Sawtooth Software, Inc.
- Clancy, K.J. & Garsen, R. (1970). Why some scales predict better. *Journal of Advertising Research*, 10(5), 33-38.
- Cohen, J. (1968). Multiple Regression as a general data-analytic system. *Psychological Bulletin*, 70, 426-443.
- Colberg, R.T. (1978). A Monte Carlo evaluation of metric recovery of conjoint measurement algorithms. In S.C. Jain (Ed.), *Research frontiers in marketing: Dialogues and directions* (Series 43, 22-26). Chicago: American Marketing Association.
- Cook, T.D. & Campbell, D.T., (1979). *Quasi-experimentation. Design & analysis issues for field settings*. Chicago: Rand McNally.
- Coombs, C.H. (1950). Psychological scaling without a unit of measurement. *Psychological Review*, 57, 145-158.
- Corstjens, M.L. & Gautschi, D.A. (1983). Conjoint analysis: A comparative analysis of specification test for the utility function. *Management Science*, 29(12), 1393-1413.

- Currin, I.S. (1981). Using segmentation approaches for better prediction and understanding from consumer mode choice models. *Journal of Marketing Research*, 18, 301-309.
- Currin, I. (1982). Predictive testing of consumer choice models not subject to independence of irrelevant alternatives. *Journal of Marketing Research*, 19, 208-222.
- Currin, I.S., Weinberg, C.B. & Wittink, D.R. (1981). The design of subscription programs for a performing arts series. *Journal of Consumer Research*, 8, 67-75.
- Curry, D.J., Menasco, M.B. & van Ark, J.W. (1991). Multiattribute dyadic choice: Models and tests. *Journal of Marketing Research*, 28, 259-267.
- David, H.A. (1988). *The method of paired comparison*, 2nd ed. London: Griffin.
- Dawes, R.M. (1982). The robust beauty of improper linear models in decision making. In D. Kahneman, P. Slovic, & A. Tversky, A.(Eds.), *Judgement under uncertainty. Heuristics and biases* (391-407). New York.
- Dawes, R.M. & Corrigan, B. (1974). Linear models in decision-making. *Psychological Bulletin*, 81, 95-106.
- Debreu, G. (1960). Typological methods in cardinal utility theory. In K.J. Arrow, S. Karlin, & P. Suppes (Eds.), *Mathematical methods in the social sciences* (16-26). Stanford, CA: Stanford University Press.
- DeSarbo, W.S., Oliver, R.S. & Rangaswamy, A. (1989). A simulated annealing methodology for clusterwise linear regression. *Psychometrika*, 54, 707-736.
- Dorsch, M.J. & Teas, R.K. (1992). A test of the convergent validity of self-explicated and decompositional conjoint measurement. *Journal of the Academy of Marketing Science*, 20, 37-48.
- Edwards, W. (1977). How to use multiattribute utility measurement for social decision making. *IEEE Transactions on Systems, Man, and Cybernetics*. SMC-7, 326-340.
- Elrod, T., Louviere, J.J. & Davey, K.S. (1992). An empirical comparison of ratings-based and choice-based conjoint models. *Journal of Marketing Research*, 29, 368-377.
- Faverge, J.-M. (1980). *Mathematisch-statistische Methoden in der Psychologie*. Stuttgart: Huber.
- Fiedler, J.A. (1988). Conjoint predictions: 15 years later. *Proceedings of the Sawtooth Software Conference of Perceptual Mapping, Conjoint Analysis, and Computer Interviewing* (25-35). Ketchum, ID: Sawtooth Software, Inc.
- Finkbeiner, C.T. (1988). Comparison of conjoint choice simulators. *Proceedings of the Sawtooth Software Conference of Perceptual Mapping, Conjoint Analysis, and Computer Interviewing* (75-103). Ketchum, ID: Sawtooth Software, Inc.

- Finkbeiner, C.T. & Platz, P.J. (1986). Computerized versus paper and pencil methods: A comparison study. Paper presented at the Association for Consumer Research Conference. Toronto (October).
- Finn, A. (1985). A theory of the consumer evaluation process for new product concepts. In J.N. Sheth (Ed.), *Research in Consumer Behavior* (Vol. 1, 35-65). Greenwich, CT: JAI.
- Fishbein, M. (1967). A behavior theory approach to the relations between beliefs about an object and the attitude toward the object. In M. Fishbein (Ed.), *Readings in attitude theory and measurement* (389-399). New York: Wiley.
- Fishbein, M. & Ajzen, I. (1972). Attitudes and opinions. *Annual Review of Psychology*, 23, 487-544.
- Fishbein, M. & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.
- Fishburn, P.C. (1967). Methods of estimating additive utilities. *Management Science*, 13, 435-453.
- Glaser W.R. (1978). *Varianzanalyse*. Stuttgart: Fischer.
- Glazer, R. (1984). Multiattribute perceptual bias as revealing of preference structure. *Journal of Consumer Research*, 11, 510-521.
- Glover, J., Ronning, R. & Reynolds, C. (1989). *Handbook of creativity*. New York: Plenum Press.
- Green, P.E. (1973). On the analysis of interactions in marketing research data. *Journal of Marketing Research*, 10, 410-420.
- Green, P.E. (1974). On the design of choice experiments involving multifactor alternatives. *Journal of Consumer Research*, 1(9), 61-68.
- Green, P.E. (1975). On the robustness of multidimensional scaling techniques. *Journal of Marketing Research*, 12, 73-81.
- Green, P.E. (1977). A new approach to market segmentation. *Business Horizons*, 20, 61-73.
- Green, P.E. (1984). Hybrid models for conjoint analysis: An expository review. *Journal of Marketing Research*, 21, 155-169.
- Green, P.E., Carmone, F.J. & Wind, Y. (1972). Subjective evaluation models and conjoint measurement. *Behavioral Science*, 17, 288-299.
- Green, P.E., Carroll, J.D. & Carmone, F.J. (1978). Some new types of fractional factorial designs for marketing experiments. In J.N. Sheth (Ed.), *Research in marketing*, (Vol. 1, 99-122). Greenwich: JAI.
- Green, P.E., Carroll, J.D. & Goldberg, S.M. (1981). A general approach to product design optimization via conjoint analysis. *Journal of Marketing*, 45(3), 17-37.

- Green, P.E. & DeSarbo, W.S. (1978). Additive decomposition of perceptions data via conjoint analysis. *Journal of Consumer Research*, 1, 56-67.
- Green, P.E. & DeSarbo, W.S. (1979). Componential segmentation in the analysis of consumer trade-offs. *Journal of Marketing*, 43, 83-91.
- Green, P.E. & Devita, M.T. (1974). A complementarity model of consumer utility for item collections. *Journal of Consumer Research*, 1(12), 56-67.
- Green, P.E. & Devita, M.T. (1975). An interaction model of consumer utility. *Journal of Consumer Research*, 2, 146-153.
- Green, P.E., Goldberg, S.M. & Montemayor, M. (1981). A hybrid utility estimation model for conjoint analysis. *Journal of Marketing*, 45(1), 33-41.
- Green, P.E., Goldberg, S.E. & Wiley, J.B. (1983). A cross validation test of hybrid conjoint models. In R.P. Bagozzi & A.M. Tybout (Eds.), *Advances in consumer research*, (Vol. 10, 147-150). Ann Arbor, MI: Association for Consumer Research.
- Green, P.E. & Helsen, K. (1989). Cross-validation assessment of alternatives to individual level conjoint analysis: A case study. *Journal of Marketing Research*, 26, 346-350.
- Green, P.E., Helsen, K. & Shandler, B. (1988). Conjoint internal validity under alternative profile presentations. *Journal of Consumer Research*, 15, 392-397.
- Green, P.E. & Krieger, A.M. (1988). Choice rules and sensitivity analysis in conjoint simulators. *Journal of the Academy of Marketing Science*, 16, 114-127.
- Green, P.E. & Krieger, A.M. (1989). Recent contributions to optimal product positioning and buyer segmentation. *European Journal of Operational Research*, 41, 127-141.
- Green, P.E. & Krieger, A.M. (1991). Segmenting markets with conjoint analysis. *Journal of Marketing*, 55(4), 20-31
- Green, P.E., Krieger, A.M. & Agarwal, M.K. (1991). Adaptive conjoint analysis: Some caveats and suggestions. *Journal of Marketing Research*, 28, 215-222.
- Green, P.E., Krieger, A.M. & Bansal, P. (1988). Completely unacceptable levels in conjoint analysis: A cautionary note. *Journal of Marketing Research*, 25, 293-300.
- Green, P.E., Krieger, A.M. & Schaffer, C.M. (1993). An empirical test of optimal respondent weighting in conjoint analysis. *Journal of the Academy of Marketing Science*, 21, 345-351.
- Green, P.E. & Rao, V.R. (1971). Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research*, 8, 355-363.
- Green, P.E., Rao, V.R. & DeSarbo, W.S. (1978). Incorporating group-level similarity judgments in conjoint analysis. *Journal of Consumer Research*, 5, 187-193.

- Green, P.E., Schaffer, C.M. & Patterson, K.M. (1991). A validation study of Sawtooth Software's adaptive conjoint analysis. In 1991 Sawtooth Software Conference Proceedings, (303-314). Ketchum, ID: Sawtooth Software Inc.
- Green, P.E. & Srinivasan, V. (1978). Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research*, 5, 103-123.
- Green, P.E. & Srinivasan, V. (1990). Conjoint analysis in marketing: New developments with implications for research and practice. *Journal of Marketing*, 54 (Oct), 3-19.
- Green, P.E. & Wind, Y. (1973). *Multivariate decisions in marketing: A measurement approach*. Hinsdale: The Dryden Press.
- Green, P.E. & Wind, Y. (1975). New way to measure consumers' judgments. *Harvard Business Review*, 53 (Jul/Aug), 107-117.
- Grunert, K.G. (1985). Methoden zur Messung der Bedeutung von Produktmerkmalen: Ein Vergleich. In GfK-Nürnberg (Hrsg.), *Jahrbuch der Absatz- und Verbrauchsforschung*. Berlin.
- Hagerty, M.R. (1985). Improving predictive power of conjoint analysis: The use of factor analysis and cluster analysis. *Journal of Marketing Research*, 22, 168-184.
- Hagerty, M.R. (1986). The cost of simplifying preference models. *Marketing Science*, 5, 298-319.
- Hagerty, M.R. (1993). Can segmentation improve predictive accuracy in conjoint analysis? *Journal of the Academy of Marketing Science*, 21, 353-355.
- Hahn, G.J. & Shapiro, S.S. (1966). A catalog and computer program for the design and analysis of orthogonal symmetric and asymmetric fractional factorial designs. GE-Report, 66-C-165.
- Hauser, J.R. & Shugan, S.M. (1980). Intensity measures of consumer preference. *Operations Research*, 28, 278-320.
- Hausruckingner, G. & Herker, A. (1992). Die Konstruktion von Schätzdesigns für Conjoint-Analysen auf der Basis von Paarvergleichen. *Marketing*, 2(2), 99-110.
- Hoepfl, R.T. & Huber, G.P. (1970). A study of self-explicated utility models. *Behavioral Science*, 15, 408-414.
- Hoffman, P.J. (1960). The paramorphic representation of clinical judgment. *Psychological Bulletin*, 57, 116-131.
- Hoffman, P.J. (1968). Cue-consistency and configurality in human judgments. In B. Kleinmuntz (Ed.), *Formal representation of human judgment*. New York: John Wiley.
- Holbrook, M.B. (1981). Integrating compositional and decompositional analysis to represent the intervening role of perceptions in evaluative judgments. *Journal of Marketing Research*, 18, 13-28.

- Holbrook, M.B. & Moore, W.L. (1981). Feature interactions in consumer judgments of verbal versus pictorial presentations. *Journal of Consumer Behavior*, 8, 103-113.
- Holt, J.O. & Wallsten, T.S. (1974). A user's manual for CONJOINT: A computer program for evaluating certain conjoint-measurement axioms. (Tech. Rep. No. 42). Chapel Hill: University of North Carolina, Thurstone Psychometric Laboratory.
- Horn, W. (1962). Leistungsprüfsystem LPS. Handanweisung (2. Aufl.). Göttingen: Hogrefe.
- Huber, G.P. (1974a). Methods for quantifying subjective probabilities and multi-attribute utilities. *Decision Science*, 5, 430-458.
- Huber, G.P. (1974b). Multiattribute utility models: A review of field and field-like studies. *Management Science*, 20, 1393-1402.
- Huber, G.P., Daneshgar, R. & Ford, D.L. (1971). An empirical comparison of five utility models for predicting job preferences. *Organizational Behavior and Human Performance*, 6, 267-282.
- Huber, G.P., Sahney, V.K. & Ford, D.L. (1969). A study of subjective evaluation models. *Behavioral Science*. 14, 483-489.
- Huber, J. & Hansen, D. (1986). Testing the impact of dimensional complexity and affective differences of paired concepts in adaptive conjoint analysis. In M. Wallendorf & P. Anderson (Eds.), *Advances in consumer research* (Vol. 14, 159-163). Provo, UT: Association for Consumer Research.
- Huber, J. & McCann, J. (1982). The impact of inferential beliefs on product evaluations. *Journal of Marketing Research*, 19, 324-333.
- Huber, J. & Moore, W. (1979). A comparison of alternative ways to aggregate individual conjoint analysis. In N.E. Beckwith et al. (Eds.), *1979 educators' conference proceedings* (64-68). Chicago: American Marketing Association.
- Huber, J., Wittink, D.R., Fiedler, J.A. & Miller, R.L. (1991). An empirical comparison of ACA and full profile judgments. In *1991 Sawtooth Software Conference Proceedings*, (189-202). Ketchum, ID: Sawtooth Software Inc.
- Huber, J., Wittink, D.R., Fiedler, J.A. & Miller, R. (1993). The effectiveness of alternative preference elicitation procedures in predicting choice. *Journal of Marketing Research*, 30, 105-114.
- Jacoby, J. (1977). Information load and decision quality. Some contested issues. *Journal of Marketing Research*, 14, 569-573.
- Jain, A.K., Acito, F., Malhotra, N.K. & Mahajan, V. (1979). A comparison of the internal validity of alternative parameter estimation methods in decompositional multiattribute preference models. *Journal of Marketing Research*, 16, 313-322.

- Jain, A.K., Mahajan, V. & Malhotra, N.K. (1979). Multiattribute preference models for consumer research: A synthesis. In: Wilkie, W.L. (ed.), *Advances in consumer research* (Vol. VI, 248-252). Ann Arbor: Association for Consumer Research.
- Jankowicz, A.D. (1990). Applications of personal construct psychology in business practice. In G.J. Neimeyer & R.A. Neimeyer (Eds.), *Advances in Personal Construct Psychology*, (Vol. 1, 257-287). London: JAI.
- Johnson, E.J., Meyer, R.J. (1984). Compensatory choice models of noncompensatory processes: The effect of varying context. *Journal of Consumer Research*, 11, 528-541.
- Johnson, E.J., Meyer, R.J. & Ghose, S. (1989). When choice models fail: Compensatory models in negatively correlated environments. *Journal of Marketing Research*, 26, 255-270.
- Johnson, R.D. (1987). Making judgments when information is missing: Inferences, biases, and framing effects. *Acta Psychologica*, 66, 69-82.
- Johnson, R.M. (1973). Pairwise nonmetric multidimensional scaling. *Psychometrika*, 38, 11-18.
- Johnson, R.M. (1974). Trade-off analysis of consumer values. *Journal of Marketing Research*, 11, 121-127.
- Johnson, R.M. (1975). A simple method for pairwise monotone regression. *Psychometrika*, 40, 163-168.
- Johnson, R.M. (1981). Problems in applying conjoint analysis. In R.K. Srivastava & A.D. Shocker (Eds.), *Analytic approaches to product and marketing planning: The second conference* (154-165). Cambridge, MA: Marketing Science Institute.
- Johnson, R.M. (1987a). Adaptive conjoint analysis. Sawtooth Software Conference on perceptual mapping, conjoint analysis, and computer interviewing (253-265). Ketchum, ID: Sawtooth Software.
- Johnson, R.M. (1987b). Accuracy of utility estimation in ACA. Working Paper. Ketchum, ID: Sawtooth Software Inc.
- Johnson, R.M. (1991). Comment on "Adaptive conjoint analysis: Some caveats and suggestions". *Journal of Marketing Research*, 28, 223-225.
- Judd, C.M. & McClelland, G.H. (1989). *Data analysis. A model-comparison approach*. San Diego: Harcourt Brace Jovanovich.
- Jütting, A. & Holling, H. (1996). ALASCA. Ein Softwarepaket zur computergestützten Durchführung von Nutzenschätzungen. Psychologisches Institut IV der Westfälischen Wilhelms-Universität Münster.
- Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237-251.
- Kamakura, W. (1988). A least squares procedure for benefit segmentation for conjoint experiments. *Journal of Marketing Research*, 25, 157-167.

- Kelly, G.A. (1955). *The psychology of personal constructs* (Vol. 1 & 2). New York: Norton.
- Klauer, K.Ch. (1988). Eine Untersuchung zu einem linearen Modell dichotomer Komponenten. *Zeitschrift für Sozialpsychologie*, 118-130.
- Klein, N.M. (1986). Assessing unacceptable attribute levels in conjoint analysis. In M. Wallendorf & P. Anderson (Eds.), *Advances in Consumer Research* (Vol. 14, 154-158). Provo, UT: Association for Consumer Research.
- Krantz, D.H. (1964). Conjoint measurement: The Luce-Tukey axiomatization and some extensions. *Journal of Mathematical Psychology*, 1, 248-277.
- Krantz, D.H., Luce, R.D., Suppes, P. & Tversky, A. (1971). *Foundations of measurement. Vol. I: Additive and polynomial representations*. New York: Academic Press.
- Krantz, D.H. & Tversky, A. (1971). Conjoint-measurement analysis of composition rules in psychology. *Psychological Review*, 78, 151-169.
- Krishnamurthi, L. (1988). Conjoint models of family decision making. *International Journal of Research in Marketing*, 5, 185-198.
- Krishnamurthi, L. & Wittink, D.R. (1991). Idiosyncratic functional forms in conjoint analysis. *International Journal of Research in Marketing*, 8, 301-313.
- Kruskal, J.B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29, 1-27.
- Kruskal, J.B. (1964b). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 29, 115-129.
- Kruskal, J.B. (1965). Analysis of factorial experiments by estimating a monotone transformation of the data. *Journal of the Royal Statistical Society, Series B*, 27, 251-263.
- Kruskal, J.B. & Carmone, F.J. (1969). MONANOVA: A FORTRAN IV program for monotone analysis of variance. *Behavioral Science*, 14, 165-166.
- Kuehn, A.A. & Day, R.L. (1962). Strategy of product quality. *Harvard Business Review*, 40, 100-110.
- Lancaster, K. (1971). *Consumer demand. A new approach*. New York.
- Lefkoff-Hagius, R. & Mason, C.H. (1993). Characteristic, beneficial, and image attributes in consumer judgments of similarity and preference. *Journal of Consumer Research*, 20, 100-110.
- Leigh, T.W., MacKay, D.B. & Summers, J.O. (1984). Reliability and validity of conjoint analysis and self-explicated weights: A comparison. *Journal of Marketing Research*, 21, 456-462.

- Lerman, S.R. & Louviere, J.J. (1978). Using functional measurement to identify the form of utility functions in travel demand models. *Transportation Research Record*, 673, 78-86.
- Lim, J., Olshavsky, R.W. & Kim, J. (1988). The impact of inferences on product evaluations: Replication and extension. *Journal of Marketing Research*, 25, 308-316.
- Loftus, E.F. (1974). Activation of semantic memory. *American Journal of Psychology*, 86, 331-337.
- Lord, F.M. & Novick, M.R. (1974). *Statistical theories of mental test scores* (2nd ed.). Reading, Mass.: Addison-Wesley.
- Louviere, J.J. (1984). Hierarchical information integration: A new method for the design and analysis of complex multiattribute judgment problems. In T.C. Kinnear (Ed.), *Advances in consumer research*, (148-155). Provo, UT: Association of Consumer Research.
- Louviere J.J. & Gaeth, G.J. (1987). Decomposing the determinants of retail facility choice using the method of hierarchical information integration: A supermarket illustration. *Journal of Retailing*, 63, 25-48.
- Louviere, J.J. & Timmermans, H.J. (1992). Testing the external validity of hierarchical conjoint analysis models of recreational destination choice. *Leisure Sciences*, 14, 179-194.
- Louviere, J.J. & Woodworth, G. (1983). Design and analysis of simulated consumer choice of allocation experiments. An approach based on aggregate data. *Journal of Marketing Research*, 20, 350-367.
- Luce, R.D. (1959). *Individual choice behavior*. New York: Wiley.
- Luce, R.D. (1966). Two extensions of conjoint measurement. *Journal of Mathematical Psychology*, 3, 348-370.
- Luce, R.D., Bush, R.R. & Galanter, E. (1965). *Handbook of Mathematical Psychology* (Vol. III). New York: Wiley & Sons.
- Luce, R.D. & Tukey, J.W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, 1, 1-27.
- Lukas, J. (1985). COMESCAL: A microcomputer program for testing axioms and finding scale values for conjoint measurement data. *Behavior Research Methods, Instruments, & Computers*, 17, 129-130.
- Lussier, D.A. & Olshavsky, R.W. (1979). Task complexity and contingent processing in brand choice. *Journal of Consumer Research*, 6, 154-165.
- Lynch, J.G. Jr. (1985). Uniqueness issues in the decompositional modeling of multiattribute overall evaluations: An information integration perspective. *Journal of Marketing Research*, 22, 1-19.

- MacBride, J.N. & Johnson, R.M. (1979). Respondent reaction to computer interactive interviewing techniques. Paper presented at the ESOMAR Conference.
- MacLachlan, D.L., Mulhern, M.G. & Shocker, A.D. (1988). Attribute selection and representation in conjoint analysis: Reliability and validity issues. Proceedings of the Sawtooth Software Conference of Perceptual Mapping, Conjoint Analysis, and Computer Interviewing (25-35). Ketchum, ID: Sawtooth Software, Inc.
- McCullough, J. & Best, R. (1979). Conjoint measurement: Temporal stability and structural reliability. *Journal of Marketing Research*, 16, 26-31.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (105-142). New York: Academic Press.
- McFadden, D. (1976). Quantal choice analysis: A survey. *Annals of Economic and Social Measurement*, 5, 363-390.
- McLauchlan, W.G. (1991). Scaling prior utilities in Sawtooth Software's Adaptive Conjoint Analysis. In 1991 Sawtooth Software Conference Proceedings, (251-268). Ketchum, ID: Sawtooth Software Inc.
- Meehl, P.E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.
- Mehrabian, A. (1968). Male and female scales of the tendency to achieve. *Educational and Psychological Measurement*, 28, 493-502.
- Mehrabian, A. (1969) Measures of achieving tendency. *Educational and Psychological Measurement*, 29, 445-451.
- Mikula, G., Schwinger, T. & Uray, H. (1976). Die Entwicklung einer deutschsprachigen Fassung der Mehrabian Achievement Risk Preference Scale. *Diagnostica*, 22, 87-97.
- Mitterer, J. & Adams-Webber, J.R. (1988). OMNIGRID: A program for the construction, administration, and analysis of repertory grids. In J.C. Mancuso & M.L.G. Shaw (Eds.), *Cognition and personal structure*, (91-106). New York: Praeger Publishers.
- Mohn, N.C. (1990). Simulated purchase 'chip' testing vs. tradeoff (conjoint) analysis - Coca Cola's experience. *Marketing Research*, 2 (March), 49-54.
- Montgomery, D.B. & Wittink, D.R. (1980). The predictive validity of conjoint analysis for alternative aggregation schemes. In D.B. Montgomery & D.R. Wittink (Eds.), *Proceedings of the first ORSA/ TIMS special interest conference on market measurement and analysis* (298-309). Cambridge, MA: Marketing Science Institute.
- Montgomery, D.B., Wittink, D.R. & Glaze, T. (1977). A predictive test of individual level concept evaluation and trade-off analysis. Research Paper No. 415, Graduate School of Business, Stanford University.

- Moore, R. (1992). Doing conjoint analysis on the telephone. Proceedings of the Sawtooth Software Conference (245-251). Sun Valley, ID: Sawtooth Software, Inc.
- Moore, W.L. (1980). Levels of aggregation in conjoint analysis: An empirical comparison. *Journal of Marketing Research*, 17, 516-523.
- Moore, W.L. & Holbrook, M.B. (1990). Conjoint analysis on objects with environmentally correlated attributes: The questionable importance of representative design. *Journal of Consumer Research*, 16, 490-497.
- Moore, W.L. & Semenik, R.J. (1988). Measuring preferences with hybrid conjoint analysis: The impact of a different number of attributes in the master design. *Journal of Business Research*, 16, 261-274.
- Müller-Hagedorn, L. & Vornberger, E. (1979). Die Eignung der Grid-Methode für die Suche nach einstellungsrelevanten Dimensionen. In H. Meffert, H. Steffenhagen & H. Freter (Hrsg.), *Konsumentenverhalten und Information* (185-207). Wiesbaden.
- Myers, J.H. & Alpert, M.I. (1968). Determinant buying attributes: Meaning and measurement. *Journal of Marketing*, 32, 13-20.
- Myers, J.H. & Alpert, M. (1977). Semantic confusion in attitude research: Saliency vs. importance vs. determinance. In W.D. Perreault Jr. (Ed.), *Advances in consumer research* (Vol. 4, 106-110). Atlanta: Association for Consumer Research.
- Myers, J.H. & Shocker, A.D. (1981). The nature of product-related attributes. *Research in Marketing*, 5, 211-236.
- Neslin, S.A. (1981). Linking product features to perceptions: Self stated versus statistically revealed importance weights. *Journal of Marketing Research*, 18, 80-86.
- Nickerson, C.A., McClelland, G.H. & Petersen, D.M. (1990). Solutions to some problems in the implementation of conjoint analysis. *Behavior Research Methods, Instruments, and Computers*, 22, 360-374.
- Nygren, T.E. (1986). A two-stage algorithm for assessing violations of additivity via axiomatic and numerical conjoint analysis. *Psychometrika*, 51, 483-491.
- Ogawa, K. (1987). An approach to simultaneous estimation and segmentation in conjoint analysis. *Marketing Science*, 6 (Winter), 66-81.
- Olshavsky, R.W. & Acito, F. (1980). An information processing probe into conjoint analysis. *Decision Sciences*, 11, 451-470.
- Oppedijk van Veen, W.M. & Beazley, D. (1977). An investigation of alternative methods of applying the trade-off model. *Journal of the Marketing Research Society*, 19, 2-11.

- Oppewal, H., Louviere, J.J. & Timmermans, H.J.P. (1994). Modeling hierarchical conjoint processes with integrated choice experiments. *Journal of Marketing Research*, 26, 92-105.
- Orne, M.T. (1962). On the social psychology of the psychological experiments: With particular reference to demand characteristics and their implications. *American Psychologist*, 17, 776-783.
- Page, A.L. & Rosenbaum, H.F. (1987). Redesigning product lines with conjoint analysis: How sunbeam does it. *Journal of Product Innovation Management*, 4, 120-137.
- Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart and Winston.
- Parker, B.R. & Srinivasan V. (1976). A consumer preference approach to the planning of rural primary health-care facilities. *Operations Research*, 24, 991-1025.
- Pawlik, K. (1976). Modell- und Praxisdimensionen psychologischer Diagnostik. In K. Pawlik (Hrsg.), *Diagnose der Diagnostik: Beiträge zur Diskussion der psychologischen Diagnostik in der Verhaltensmodifikation* (13-43). Stuttgart: Klett.
- Payne, J.W., Bettman, J.R. & Johnson, E.J. (1992). Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology*, 43, 87-131.
- Pekelman, D. & Sen, S.K. (1979a). Measurement and estimation of conjoint utility functions. *Journal of Consumer Research*, 5, 263-271.
- Pekelman, D. & Sen, S.K. (1979b). Improving prediction in conjoint analysis. *Journal of Marketing Research*, 16, 211-220.
- Petty, R.E. & Cacioppo, J.T. (1981). *Attitudes and persuasion: Classic and contemporary approaches*. Dubuque: Brown.
- Pras, B. (1975). Issues in identifying determinant attributes. *Der Markt*, 55, 72-77.
- Quack, H. (1980). Die Einstellungsrelevanz von Produktmerkmalen. *Multidimensionale Messung und Bedeutung für die Wirkung von Werbebotschaften*. Frankfurt.
- Raeithel, A. (1993). Auswertungsmethoden für Repertory Grids. In J.W. Scheer & A. Catina (Hrsg.), *Einführung in die Repertory Grid-Technik. Band 1: Grundlagen und Methoden* (41-67). Bern, Göttingen: Huber.
- Rao, V.R. (1977). Conjoint measurement in marketing analysis. In J.N. Sheth (Ed.), *Multivariate methods for market and survey research* (257-286). Chicago: American Marketing Association.
- Reibstein, D., Bateson, J.E.G. & Boulding, W. (1988). Conjoint analysis reliability: Empirical findings. *Marketing Science*, 7, 271-286.
- Reynolds, T.J., Fiedler, J.A. & Gutman, J. (1985). Understanding consumers' cognitive structures: The relationship of levels of abstraction to judgments of

- psychological distance and preference. *Psychological Processes and Advertising Effects*, 261-272.
- Riemann, R. (1991). *Repertory Grid Technik*. Handanweisung. Göttingen: Hogrefe.
- Rosenberg, M.J. (1956). Cognitive structure attitudinal affect. *Journal of Abnormal and Social Psychology*, 53, 367-372.
- Rosenberg, M.J. & Hovland, C.I. (1960). Cognitive, affective, and behavioral components of attitudes. In C.I. Hovland & M.J. Rosenberg (Eds.), *Attitude organization and change*. New Haven: Yale University Press.
- Safizadeh, M.H. (1989). The internal validity of the trade-off method of conjoint analysis. *Decision Sciences*, 20, 451-461.
- Sawtooth Software (1985). *Adaptive conjoint analysis: User's manual*. Ketchum, ID: Sawtooth Software.
- Sawtooth Software (1994). *ACA system*. Version 4.0. Evanston.
- Scheer, J.W. (1993). Planung und Durchführung von Repertory Grid-Untersuchungen. In J.W. Scheer & A. Catina (Hrsg.), *Einführung in die Repertory Grid-Technik*. Band 1: Grundlagen und Methoden (24-40). Bern, Göttingen: Verlag Hans Huber.
- Schlicksupp, H. (1985). *Ideenfindung - Management Wissen* (2. Aufl.). Würzburg: Vogel Buchverlag.
- Schubert, B. (1991). Entwicklung von Konzepten für Produktinnovationen mittels Conjointanalyse. Stuttgart: Poeschel.
- Schweikl, H. (1985). *Computergestützte Präferenzanalyse mit individuell wichtigen Produktmerkmalen*. Berlin: Duncker & Humblot.
- Scott, J.E. & Wright, P. (1976). Modeling an organizational buyer's product evaluation strategy: Validity and procedural considerations. *Journal of Marketing Research*, 13, 211-224.
- Segal, M.N. (1982). Reliability of conjoint analysis: Contrasting data collection procedures. *Journal of Marketing Research*, 19, 139-143.
- Sheluga, D.A., Jaccard, J. & Jacoby, J. (1979). Preference, search and choice: An integrative approach. *Journal of Consumer Research*, 5, 166-176.
- Shocker, A.D. & Srinivasan, V. (1974). A consumer-based methodology for the identification of new product ideas. *Management Science*, 20, 921-937.
- Sixtl, F. (1967). *Meßmethoden der Psychologie*. Theoretische Grundlagen und Probleme. Weinheim: Beltz.
- Slovic, P., Fischhoff, B. & Lichtenstein, S. (1977). Behavioral decision theory. *Annual Review of Psychology*, 28, 1-39.
- Slovic, P., Lichtenstein, S. & Fischhoff, B. (1988). Decision making. In R.D. Atkinson, R.J. Herrnstein, G. Lindzey & R.D. Luce (ed.), *Steven's handbook of*

- experimental psychology. Vol. 2: Learning and cognition. (673-738). New York: Wiley.
- Springer, S.P. & Deutsch, G. (1988). Linkes - rechtes Gehirn: Funktionelle Asymmetrien (2. Aufl.). Heidelberg: Spektrum der Wissenschaft.
- SPSS (1994). SPSS 6.1 Categories. Handbuch. Chicago, IL: SPSS Inc.
- Srinivasan, V. (1988). A conjunctive-compensatory approach to the self-explication of multiattributed preferences. *Decision Sciences*, 19, 295-305.
- Srinivasan, V. & Shocker, A.D. (1973a). Linear programming techniques for multidimensional analysis of preferences. *Psychometrika*, 38, 337-369.
- Srinivasan, V. & Shocker, A.D. (1973b). Estimating the weights for multiple attributes in a composite criterion using pairwise judgments. *Psychometrika*, 38, 473-493.
- Srinivasan, V. & Shocker, A.D. (1973c). Estimating the weights for multiple attributes in a composite criterion using pairwise judgments. *Psychometrika*, 38, 473-493.
- Srinivasan, V. & Shocker, A.D. (1981). LINMAP version-IV - Users manual. Nashville, TN: Vanderbilt University.
- Srinivasan, V., Jain, A.K. & Malhotra, N.K. (1983). Improving predictive power of conjoint analysis by constrained parameter estimation. *Journal of Marketing Research*, 20, 433-438.
- Stahl, B. (1988). Conjoint analysis by telephone. Proceedings of the Sawtooth Software Conference of Perceptual Mapping, Conjoint Analysis, and Computer Interviewing (131-138). Ketchum, ID: Sawtooth Software, Inc.
- Tantiwong, D. & Wilton, P.C. (1985). Understanding food store preferences among the elderly using hybrid conjoint measurement models. *Journal of Retailing*, 61(4), 35-64.
- Teas, R.K. (1987). Magnitude scaling of the dependent variable in decompositional multiattribute preference models. *Journal of the Academy of Marketing Science*, 15, 64-73.
- Teas, R.K. & Perr, A.L. (1989). A test of a decompositional method of multiattribute perceptions measurement. *Journal of Consumer Research*, 16, 384-391.
- Teichert, T. (1994). Zur Validität der in Conjoint-Analysen ermittelten Nutzenwerte. *Zeitschrift für betriebswirtschaftliche Forschung*, 7/8, 610-629.
- Thomas, L. (1979). Conjoint Measurement als Instrument der Absatzforschung. *Marketing. Zeitschrift für Forschung und Praxis*, 1, 199-211.
- Thomas, L. (1983). Der Einfluß von Kindern auf die Produktpräferenzen ihrer Mütter. Berlin: Duncker-Humblot.
- Thurstone, L.L. (1945). The prediction of choice. *Psychometrika*, 10, 237-253.
- Timmermans, H.J.P. (1982). Consumer choice of shopping centre: An information integration approach. *Regional Studies*, 16, 171-182.

- Timmermans, H. (1987). Hybrid and non-hybrid evaluation models for predicting outdoor recreation behavior: A test of predictive ability. *Leisure Sciences*, 9(2), 67-76.
- Toy, D., Rager, R. & Guadagnolo, F. (1989). Strategic marketing for recreational facilities: A hybrid conjoint analysis approach. *Journal of Leisure Research*, 21, 276-296.
- Trost, A. (1995). Messung der Kundenzufriedenheit im Beratungsprojekt. Diplomarbeit an der Universität Mannheim.
- Troutman, C.M. & Shanteau, J.C. (1976). Do consumers evaluate products by adding and averaging attribute information? *Journal of Consumer Research*, 3, 101-106.
- Tucker, L.R. & Messick, S. (1963). An individual differences model for multidimensional scaling. *Psychometrika*, 28, 333-367.
- Tukey, J.W. (1949). One degree of freedom for additivity. *Biometrics*, 5, 232-242.
- Tversky, A. (1967). A general theory of polynomial conjoint measurement. *Journal of Mathematical Psychology*, 4, 1-20.
- Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, 76, 31-98.
- Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, 79, 281-299.
- Ullrich, J.R. & Cummins, D.E. (1973). PCJM: A program for conjoint measurement analysis of polynomial composition rules. *Behavioral Science*, 18, 226-227.
- Ullrich, J.R., Cummins, D.E. & Walkenbach, J. (1978). PCJM2: A program for the axiomatic conjoint measurement analysis of polynomial composition rules. *Behavior Research Methods & Instrumentation*, 10, 89-90.
- Ullrich, J.R. & Wilson, R.E. (1990). CPCJM - A set of programs for checking polynomial conjoint measurement and additivity axioms of threedimensional matrices. *Applied Psychological Measurement*, 14, 433-434.
- van der Lans, I.A. & Heiser, W.J. (1992). Constrained part-worth estimation in conjoint analysis using the self-explicated utility model. *International Journal of Research in Marketing*, 9, 325-344.
- van der Lans, I.A., Wittink, D.R., Huber, J. & Vriens, M. (1992). Within- and across-attribute constraints in ACA and full profile conjoint analysis. In 1992 Sawtooth Software Conference Proceedings, (365-379). Ketchum, ID: Sawtooth Software Inc.
- Wallsten, T.S. (1977). Measurement and interpretation of beliefs: A review. In H. Jungermann & G. de Zeeuw (Eds.), *Decision making and change in human affairs* (369-393). Dordrecht.
- Wedel, M. & Kistemaker, C. (1989). Consumer benefit segmentation using clusterwise linear regression. *International Journal of Research in Marketing*, 6, 45-59.

- Wedel, M. & Steenkamp, J.-B.E.M. (1989). A fuzzy clusterwise regression approach to benefit segmentation. *International Journal of Research in Marketing*, 6, 241-258.
- Wedel, M. & Steenkamp, J.-B.E.M. (1991). A clusterwise regression method for simultaneous fuzzy market structuring and benefit segmentation. *Journal of Marketing Research*, 28, 385-396.
- Westenberg, M.R.M. & Koele, P. (1992). Response modes, decision processes and decision outcomes. *Acta Psychologica*. 80, 169-184.
- Westwood, D., Lunn, T. & Beazley, D. (1974). The trade-off model and its extensions. *Journal of the Market Research Society*, 16, 227-241.
- Wiley, J.B. & Low, J.T. (1983). A Monte Carlo simulation study of two approaches for aggregating conjoint data. *Journal of Marketing Research*, 20, 405-416.
- Wiley, J.B., MacLachlan, D.L. & Moinpour, R. (1977). Comparison of stated and inferred parameter values in additive models: An illustration of a paradigm. In W.D. Perreault Jr. (Ed.), *Advances in consumer research* (Vol. 4, 98-105). Atlanta: Association for Consumer Research.
- Wilkie, W.L. & Pessemier, E.A. (1973). Issues in marketing's use of multi-attribute attitude models. *Journal of Marketing Research*, 10, 428-441.
- Willutzki, U. & Raeithel, A. (1993). Software für Repertory Grids. In J.W. Scheer & A. Catina (Hrsg.), *Einführung in die Repertory Grid-Technik. Band 1: Grundlagen und Methoden* (80-91). Bern, Göttingen: Huber.
- Wind, Y. (1978). Issues and advances in segmentation research. *Journal of Marketing Research*, 15, 317-337.
- Wittink, D.R. & Cattin, P. (1981). Alternative estimation methods for conjoint analysis: A Monte Carlo study. *Journal of Marketing Research*, 18, 101-106.
- Wittink, D.R. & Cattin, P. (1989). Commercial use of conjoint analysis: An update. *Journal of Marketing*, 53, 91-96.
- Wittink, D.R. & Montgomery, D.T. (1979). Predictive validity of trade-off analysis for alternative segmentation schemes. In N. Beckwith et al. (Eds.), *1979 Educators' Conference Proceedings* (series 44, 69-73). Chicago: American Marketing Association.
- Wittink, D.R., Huber, J., Fiedler, J.A. & Miller, R. (1991). The magnitude of and an explanation/ solution for the number of levels effect in conjoint analysis. Research Paper, Cornell University at Ithaca, NY.
- Wittink, D.R., Huber, J., Zandan, P. & Johnson, R.M. (1992). The number of levels effect in conjoint: Where does it come from, and can it be eliminated? *Proceedings of the Sawtooth Software Conference* (355-364). Sun Valley, ID: Sawtooth Software, Inc.

-
- Wittink, D.R., Krishnamurthi, L. & Nutter, J.B. (1982). Comparing derived importance weights across attributes. *Journal of Consumer Research*, 8, 471-474.
- Wittink, D.R., Krishnamurthi, L. & Reibstein, D.J. (1989). The effect of differences in the number of attribute levels on conjoint results. *Marketing Letters*, 1, 113-123.
- Wittink, D.R., Vriens, M. & Burhenne, W. (1994). Commercial use of conjoint analysis in europe: Results and critical reflections. *International Journal of Research in Marketing*, 11, 41-52.
- Wright, P., & Kriewall, M.A. (1980). State-of-mind effects on the accuracy with which utility functions predict marketplace choice. *Journal of Marketing Research*, 17, 277-293.
- Yntema, D.B. & Torgerson, W.S. (1961). Man-computer cooperation in decisions requiring common sense. *IRE Transactions on Human Factors in Electronics*, HFE-2, 20-26.
- Young, F.W. (1972). A model for polynomial conjoint analysis algorithms. In R.N. Shepard, A.K. Romney & S.B. Nerlove (Eds.), *Multidimensional scaling. Theory and applications in the behavioral sciences* (Vol. I: Theory, 69-104). New York.

