

UNIVERSITÄT MANNHEIM

Assoziatives Gedächtnis im Alter

Ein multinomiales Modell zur Trennung von
assoziativen Enkodierungs- und Abrufprozessen
sowie dessen gerontologische Anwendung

Inauguraldissertation zur Erlangung des akademischen
Grades einer Doktorin der Sozialwissenschaften der
Universität Mannheim

vorgelegt von

Dipl.-Psych. Tina Sarah Auer

Dekan: Prof. Dr. Josef Brüderl

Gutachter: Prof. Dr. Edgar Erdfelder

Prof. Dr. Rüdiger Pohl

Tag der Disputation: 04. Juni 2009

Zusammenfassung

Assoziative Gedächtnisprozesse sind für die erfolgreiche Bewältigung verschiedenster kognitiver Aufgaben sowohl im Labor als auch im Alltag notwendig. Die von Naveh-Benjamin aufgestellte Assoziationsdefizit-Hypothese (Naveh-Benjamin, 2000) postuliert, dass ein bei älteren Menschen auftretendes Defizit in assoziativen Gedächtnisprozessen als Erklärung für deren schlechtere episodische Gedächtnisleistung herangezogen werden kann.

Ein assoziatives Defizit konnte in zahlreichen Studien mit unterschiedlichen Materialien repliziert werden und viele moderierende Faktoren wurden bereits untersucht (zusammenfassend siehe Old & Naveh-Benjamin, 2008a). Experimente konnten zeigen, dass es sich bei dem Phänomen um ein multikausales Altersdefizit handelt, an dem neben strategischen Anteilen Gedächtnisprozesse eine zentrale Rolle spielen. Es wird auch diskutiert, die defizitären assoziativen Gedächtnisprozesse als Basismechanismus des kognitiven Alterns einzuordnen. Trotz all dieser Erkenntnisse ist bisher die zentrale Frage der Lokalisation unbeantwortet. Es konnte nicht geklärt werden, inwieweit assoziative Gedächtnisprozesse bereits während der Enkodierung scheitern oder erst beim Abruf.

Zur Lösung dieser Lokalisationsfrage wurde in der vorliegenden Arbeit zunächst ein neues Paradigma entwickelt, welches es ermöglicht, assoziative Rekognitionsprozesse und Itemrekognitionsprozesse in einem Test zu erfassen. Des Weiteren wurde ein mathematisches Modell formuliert, in dem die beiden Prozesse der assoziativen Enkodierung und des assoziativen Abrufs dekomponiert werden. Auch die Itemrekognitionsleistung sowie Rateprozesse werden in dem Modell berücksichtigt und können zwischen den Altersgruppen verglichen werden.

In vier Experimenten zeigte das Modell eine gute Datenanpassung und es gelang, die Modellparameter zu validieren, sowohl im Hinblick auf die konvergente als auch diskriminante Validität. Das durchgeführte Altersexperiment trug zunächst zur Generalisierung des assoziativen Defizits bei: auch in dem neuen Paradigma zeigte sich der typische Alterseffekt. Der zentrale und innovative Befund des Altersexperiments bestand im Nachweis eines Enkodierungsdefizits für assoziative Informationen. In beiden Testlisten wiesen ältere Menschen gegenüber jüngeren Menschen ein signifikantes Enkodierungsdefizit auf.

Dieses Modell erlaubt erstmals die getrennte Erfassung von Enkodierungs- und Abrufprozessen in einem Rekognitionsparadigma und ist damit nicht nur in der Lage, eine Antwort auf die Lokalisationsfrage des assoziativen Defizits zu ermöglichen, sondern kann für zahlreiche Forschungsfragen, unter anderem in der Entwicklungspsychologie und Allgemeinen Psychologie, gewinnbringend eingesetzt werden.

Inhaltsverzeichnis

1	Einleitung	1
I	Theoretischer Hintergrund	5
2	Theoretische Perspektiven der kognitiven Altersforschung	7
3	Defizitäre assoziative Basismechanismen	17
3.1	Kontextdefizit-Hypothese	18
3.2	Die Assoziationsdefizit-Hypothese	31
3.2.1	Die Assoziationsdefizit-Hypothese: Eine Definition	33
3.2.2	Empirische Befunde zur Assoziationsdefizit-Hypothese	36
3.2.3	Qualitatives Defizit oder Basismechanismus	59
3.2.4	Das assoziative Defizit in Verbindung zu anderen Basismechanismen	68
3.2.5	Das assoziative Defizit aus der Perspektive von Lokalisationsansätzen	79
4	Methodische Aspekte zur Trennung von Enkodierung und Abruf	85
4.1	Die Lokalisationsfrage als Dekomponierungsproblem	86
4.2	Mathematische Modelle	89
4.2.1	Das Enkodierung-Abruf-Modell	94
4.2.2	Quellengedächtnis-Modelle	102
4.2.3	Das Modell zur Assoziationsdefizit-Hypothese	107
II	Empirische Modellvalidierung	127
5	Methodisches Vorgehen der Modellvalidierung	129
6	Experiment 1	133
6.1	Experiment 1a	133
6.1.1	Methode	140
6.1.2	Ergebnisse	146

6.1.3	Diskussion	152
6.2	Experiment 1b	153
6.2.1	Methode	155
6.2.2	Ergebnisse	158
6.2.3	Diskussion	161
7	Experiment 2	165
7.1	Methode	172
7.2	Ergebnisse	177
7.3	Diskussion	187
8	Experiment 3 und 4	193
8.1	Experiment 3	196
8.1.1	Methode	196
8.1.2	Ergebnisse	200
8.1.3	Diskussion	203
8.2	Experiment 4	205
8.2.1	Methode	205
8.2.2	Ergebnisse	208
8.2.3	Diskussion	212
9	Zusammenfassung und Diskussion der Validierungsexperimente	213
III	Gerontologische Modellanwendung	221
10	Experiment 5	223
10.1	Methode	223
10.2	Ergebnisse	228
10.3	Diskussion	237
IV	Diskussion	241
11	Abschließende Diskussion und Ausblick	243
11.1	Die Lokalisationsfrage des assoziativen Defizits	243
11.2	Das ADH-Modell: Eine kritische Evaluation	244
11.3	Einsatzmöglichkeiten des ADH-Modells	248
11.4	Entwicklungspsychologische Implikationen	250
11.5	Spezifizierung der Assoziationsdefizit-Hypothese	252
11.6	Ausblick	255
11.7	Fazit	257

V	Literatur	259
	Literaturverzeichnis	261
VI	Anhang	285
A	Verarbeitungsbaummodelle des AHD-Modells	287
B	eqn-Datei	293
C	Wortmaterial	297
D	Wortmaterial Experiment 2	307
E	Wortmaterial Experiment 5	311
	E.1 Wortpaare der ersten und zweiten Lernliste	311
	E.2 Testversionen A und B für die erste Lernliste	313
	E.3 Testversionen A und B für die zweite Lernliste	315
F	mdt-Dateien	317
	F.1 Experiment 1a	317
	F.2 Experiment 1b	318
	F.3 Experiment 2	320
	F.4 Experiment 3	323
	F.5 Experiment 4	324
	F.6 Experiment 5	326
	Danksagung	331

Tabellenverzeichnis

4.1	Datenmatrix des modifizierten Paradigmas	110
4.2	Beschreibung der Parameter des ADH-Modells	119
5.1	Ausblick auf die Validierungsexperimente	131
6.1	Kennwerte der Wortlisten	145
6.2	Experiment 1a: Relative Antworthäufigkeiten	147
6.3	Parameterschätzungen Experiment 1a	151
6.4	Experiment 1b: Relative Antworthäufigkeiten	159
6.5	Parameterschätzungen Experiment 1b	162
7.1	Satzbeispiele von Experiment 2	174
7.2	Experiment 2: Relative Antworthäufigkeiten der Enkodierungsbedingung	179
7.3	Parameterschätzungen Experiment 2 (Enkodierungsbedingung)	179
7.4	Experiment 2: Relative Antworthäufigkeiten der Abrufbedingung	183
7.5	Parameterschätzungen Experiment 2 (Abrufbedingung)	183
8.1	Anzahl der unterschiedlichen Wortkombinationen in Experiment 3	197
8.2	Experiment 3: Relative Antworthäufigkeiten	201
8.3	Parameterschätzungen Experiment 3	204
8.4	Experiment 4: Relative Antworthäufigkeiten	209
8.5	Parameterschätzungen Experiment 4	211
9.1	Zusammenfassung der Validierungsergebnisse	217
10.1	Experiment 5: Relative Antworthäufigkeiten Liste 1 und Liste 2	232
10.2	Parameterschätzungen Experiment 5	236

Abbildungsverzeichnis

2.1	Einordnung der Forschungsfrage in die kognitive Altersforschung	15
3.1	Lernphase des Experiments von Chalfonte und Johnson (1996)	29
3.2	Testphase des Experiments von Chalfonte und Johnson (1996)	30
3.3	Ergebnisse der Meta-Analyse von Old und Naveh-Benjamin (2008a)	39
3.4	Integrationspotential unterschiedlicher Materialien	56
3.5	Faktoren mit moderierendem Einfluss auf das Assoziationsdefizit	58
3.6	Ergebnisse der Lebensspannenstudie von Shing et al. (2008)	67
3.7	Das Assoziationsdefizit aus der Perspektive reduzierter Basismechanismen	68
3.8	Auswirkungen einer Zweitaufgabe bei Item- und assoziative Tests	75
4.1	Das Enkodierung-Abruf-Modell	97
4.2	Quellengedächtnismodelle	105
4.3	Zweidimensionales Quellengedächtnismodelle	106
4.4	Vereinfachte Darstellung des MVB-Modells für alte intakte Wortpaare	112
4.5	Vereinfachte Darstellung des MVB-Modells für neu kombinierte Wortpaare	114
4.6	Vereinfachte Darstellung des MVB-Modells für Target-Distraktor-Paare	116
4.7	Darstellung der Rateparameter	117
4.8	Verarbeitungsbaummodell für die Kombination von zwei Distraktoren	118
4.9	Verarbeitungsbaummodell für Target-Distraktor-Kombinationen	119
4.10	Verarbeitungsbaummodell für alte, intakte Wortpaare	120
4.11	Verarbeitungsbaummodell für neu kombinierte Wortpaare	121
4.12	Modellhierarchie des ADH-Modells	125
6.1	Darstellung des sequentiellen Ja-nein-Formats	134
6.2	Darstellung des 5AFC-Formats	135
6.3	Bildschirmausschnitt aus der Lernphase	142
6.4	Darstellung der beiden Testformate	143
6.5	Experiment 1a: Gedächtnisparameter des ADH-Modells	149
6.6	Experiment 1a: Rateparameter des ADH-Modells	150
6.7	Abbildung zur gemeinsamen Verarbeitungszeit	154

6.8	Manipulation der Darbietungszeit	156
6.9	Experiment 1b: Darstellung der Gedächtnisparameter	160
6.10	Experiment 1b: Rateparameter des ADH-Modells	161
7.1	Enkodierungs-Abruf-Paradigma nach Tulving (1983)	166
7.2	Schematische Darstellung der Hypothesen von Experiments 2	167
7.3	Operationalisierung der tiefen Enkodierungsbedingung	175
7.4	Operationalisierung der flachen Enkodierungsbedingung	176
7.5	Experiment 2: Gedächtnisparameter der Enkodierungsbedingung	180
7.6	Experiment 2: Rateparameterschätzungen der Enkodierungsbedingung . . .	181
7.7	Experiment 2: Distraktorentdeckungsparameter der Enkodierungsbedingung	181
7.8	Experiment 2: Gedächtnisparameter der Abrufbedingung	184
7.9	Experiment 2: Rateparameterschätzungen Abrufbedingung	185
7.10	Experiment 2: Enkodierungs- und Abrufparameter	185
7.11	Überblick über die Ergebnisse von Experiment 2	186
8.1	Leistungsrückmeldung an die Teilnehmer	199
8.2	Experiment 3: Rateparameter des ADH-Modells	203
8.3	Experiment 3: Parameterschätzungen der Gedächtnisparameter	204
8.4	Darstellung der Testphase	207
8.5	Erklärung der Auszahlungsmatrix	208
8.6	Experiment 4: Darstellung der Rateparameter	210
8.7	Experiment 4: Parameterschätzungen der Gedächtnisparameter	211
9.1	Zwischenexperimenteller Parametervergleich	218
10.1	Auswertung der korrigierten Rekognitionsleistung für die erste Testliste . .	229
10.2	Auswertung der korrigierten Rekognitionsleistung für die zweite Testliste . .	229
10.3	Experiment 5: Parameterschätzungen der Gedächtnisparameter (Liste 1) . .	233
10.4	Experiment 5: Rateparameter des ADH-Modells (Liste 1)	233
10.5	Experiment 5: Parameterschätzungen der Gedächtnisparameter (Liste 2) . .	234
10.6	Ratparameter des ADH-Modells (Liste 2)	235
11.1	Einordnung der Assoziationsdefizit-Hypothese	253
A.1	Verarbeitungsbaummodell für alte, intakte Wortpaare	288
A.2	Verarbeitungsbaummodell für neu kombinierte Wortpaare	289
A.3	Verarbeitungsbaummodell für Target-Distraktor Kombinationen	290
A.4	Verarbeitungsbaummodell für Distraktor-Target Kombinationen	291
A.5	Verarbeitungsbaummodell für zwei Distraktoren	292

Kapitel 1

Einleitung

Zu den Standardbefunden der kognitiven Altersforschung gehört die schlechtere Leistung älterer Menschen in episodischen Gedächtnisaufgaben (für einen Überblick siehe Zacks, Hasher & Li, 2000). Dieses episodische Defizit ist abzugrenzen gegenüber der Leistung in semantischen Gedächtnisaufgaben, denn in diesen finden sich tendenziell keine Altersunterschiede oder höchstens derart, dass ältere Menschen jüngeren überlegen sind (für einen Überblick siehe Salthouse, 1991). Zur Erklärung dieses episodischen Defizits werden verschiedene Prozesse herangezogen. Neben der Verlangsamung der Verarbeitungsgeschwindigkeit (Salthouse, 1996), der herabgesetzten Inhibitionsfähigkeit (Hasher & Zacks, 1988; Hasher, Zacks & May, 1999) und dem Nachlassen selbstinitiiertter Prozesse (Craik & Byrd, 1982) wird auch diskutiert, inwieweit defizitäre assoziative Mechanismen diesem Phänomen zugrunde liegen könnten (Bayen, Phelps & Spaniol, 2000; Chalfonte & Johnson, 1996; Naveh-Benjamin, 2000). Episodische Gedächtnisaufgaben zeichnen sich ja gerade dadurch aus, dass eine Assoziation zu dem zeitlichen räumlichen Kontext geschaffen werden muss. In Laboraufgaben ist die Verbindung zwischen den zu lernenden Wörtern und dem aktuellen experimentellen Kontext relevant, denn in der Testphase muss entschieden werden, ob ein Wort in diesem spezifischen Kontext der Lernliste gelesen wurde, oder ob das Wort in einem ganz anderen Kontext, z.B. beim Zeitungslesen verarbeitet wurde. Auch in alltäglichen Situationen ist es von zentraler Bedeutung, Assoziationen zu erinnern. Es kann nur dann korrekt darüber geurteilt werden, zu welchem Zeitpunkt ein Medikament eingenommen wurde, wenn die Verbindung zwischen ausgeführter Handlung und dem genauen Zeitpunkt der Handlungsausführung erinnert wird. Die Assoziation von einer Aussage und der dazugehörigen Quelle muss auch abrufbar sein, um zu entscheiden, ob eine Information von einer seriösen Quelle (Arzt) oder einer zweifelhaften Quelle (Werbung) stammt. Assoziative Gedächtnisprozesse sind also nicht nur im Labor, sondern ebenso im Alltag relevant.

Die zentrale Rolle assoziativer Gedächtnisprozesse zur Erklärung von Altersunterschieden in episodischen Gedächtnisaufgaben wurde in zahlreichen Experimenten empi-

risch nachgewiesen (Bayen, Phelps & Spaniol, 2000; Chalfonte & Johnson, 1996; Mitchell, Johnson, Raye, Mather & D'Esposito, 2000; Naveh-Benjamin, 2000). Im Jahr 2000 fasste Naveh-Benjamin diese empirischen Befunde zur Assoziationsdefizit-Hypothese zusammen. Die Kernaussage dieser Hypothese gibt folgendes Zitat wieder: „This hypothesis claimed that an important component of older adults' poorer episodic memory is their deficiency in creating and retrieving links between single units of information. The basic units can be two items; an item and its context; two contextual elements; or, more generally, the representation of two mental codes.“ (Naveh-Benjamin, 2000, S. 1170).

Dieses assoziative Defizit, tritt besonders deutlich zu Tage, wenn die Kombination zweier Wörter erinnert werden soll und gleichzeitig die Rekognitionsleistung an die einzelnen Elemente erfasst wird. Die erstgenannte *assoziative* Rekognitionsaufgabe kann nur korrekt beantwortet werden, wenn zum einen die Assoziation von beiden Wörtern zum anderen die Assoziation zu dem Kontext der Episode abgerufen werden kann. Die Verbindung zwischen episodischem Kontext und der Iteminformation ist ausreichend für die erfolgreiche Bearbeitung der *Itemrekognitionsaufgabe*. Die beiden Tests beruhen also unterschiedlich stark auf assoziativen Gedächtnisprozessen und eignen sich daher zur Erfassung eines *assoziativen* Gedächtnisdefizits, welches über ein allgemeines Gedächtnisdefizit der Itemrekognition hinausgeht. In diesem Testvergleich findet sich bei älteren Menschen ein spezifisches Defizit für die Assoziation der beiden Elemente, welches das Rekognitionsdefizit der Einzelelemente übersteigt.

Das assoziative Defizit konnte mit diesem Verfahren für unterschiedliche Materialien (z.B. zwei Wörter, Wort und Schriftart, Gesicht und Name) nachgewiesen werden (für einen Überblick siehe Old & Naveh-Benjamin, 2008a). Auch die Rolle von strategischen Prozessen an diesen Aufgaben (Naveh-Benjamin, Brav & Levy, 2007; Shing, Werkle-Bergner, Li & Lindenberger, 2008) sowie moderierende Faktoren der Lern- und Testsituation wurden untersucht (Chalfonte & Johnson, 1996; Naveh-Benjamin, 2000). Unbeantwortet blieb bisher, inwieweit dieses Defizit auf Enkodierungs- oder Abrufprozesse zurückzuführen ist, d.h. das Lokalisationsproblem ist bisher ungelöst. Es wurde in einzelnen Experimenten versucht, über ein experimentelles Vorgehen (d.h. designbasierte Methoden) eine Antwort zu finden. Allerdings handelt es sich hierbei um ein Dekomponierungsproblem zweier kognitiver Prozesse, daher sind diese Verfahren ungeeignet und adäquate Lösungsansätze sind in mathematischen Modellen zu suchen (Brainerd, 1985; Erdfelder, 2000). Modellbasierte Lösungsansätze für diese zentrale Fragestellung lagen bisher allerdings nicht vor.

Diese Arbeit wendet sich so der Lokalisationsfrage des assoziativen Defizits zu und setzt sich zum Ziel, ein neues mathematisches Modell zu entwickeln, welches es ermöglicht, die Prozesse der assoziativen Enkodierung und des assoziativen Abrufs im Rekognitionsparadigma getrennt voneinander zu erfassen. Weiterhin werden auch die Prozesse der Itemrekognition und Distraktorentdeckung sowie Ratetendenzen modelliert. Das Modell

soll nach der sorgfältigen Validierung und empirischen Prüfung in einer Altersstudie angewandt werden und erste Ergebnisse im Hinblick auf das Lokalisationsproblem erbringen.

Die vorliegende Arbeit ist in vier Teile untergliedert: einen theoretischen Teil (I), einen empirischen Teil der Modellvalidierung (II), einen empirischen Teil der gerontologischen Modellanwendung (III) sowie einen Diskussionsteil (IV).

Im ersten Teil wird der theoretische Hintergrund der Forschungsfrage erläutert. Da das Ziel dieser Arbeit in der Weiterführung der kognitiven Altersforschung gesehen wird, wird sich der theoretische Teil mit diesem Forschungsfeld beschäftigen und die Forschungsfrage dieser Dissertation in das Gebiet der kognitiven Altersforschung einordnen. Zu diesem Zweck wird in Kapitel 2 zunächst ein Rahmenkonzept (Salthouse, 1991) vorgestellt, welches einen systematischen Zugang zu Theorien und Befunden der Altersforschung ermöglicht. Im Kapitel 3 werden empirische Befunde sowie theoretische Ansätze zu defizitären assoziativen Basismechanismen erläutert. Der Aufbau dieses Kapitels orientiert sich dabei an dem im vorangehenden Kapitel vorgestellten Rahmenkonzept. Es wird zum einen aufgezeigt, mit welchen Materialien das assoziative Defizit bisher repliziert wurde, zum anderen wird auf den Einfluss von Kontextfaktoren wie Lern- und Testsituation eingegangen (Kapitel 3.2.2). Im Hinblick auf Moderatoren werden nicht nur die bisher in der Literatur untersuchten Faktoren angesprochen. Auch auf Aspekte wie das Zuordnungsverhältnis oder das Integrationspotential wird eingegangen. Diese wurden bisher vernachlässigt, sind aber aus theoretischer Perspektive von Bedeutung. Um herauszuarbeiten, inwieweit in dem assoziativen Defizit ein Basismechanismus zu sehen ist, werden Arbeiten angeführt, die ausschließen, dass der Effekt allein auf Altersunterschieden in strategischen Prozessen beruht (Kapitel 3.2.3) oder durch andere Basismechanismen erklärt werden kann (Kapitel 3.2.4). Die wenigen bisher dazu vorliegenden Arbeiten rechtfertigen die Stellung des assoziativen Defizits als eigenständigen Basismechanismus.

Nachdem das Forschungsthema des assoziativen Defizits genauer spezifiziert und dessen wichtige Rolle in der kognitiven Altersforschung geklärt wurde, wird auf offene Fragen des Forschungsfeldes eingegangen. Ein wichtiger Aspekt ist dabei die Lokalisation des assoziativen Defizits (Kapitel 3.2.5). Ist dieses auf Probleme während der Enkodierungsphase oder während der Abrufphase zurückzuführen? Diese Frage konnte bisher noch nicht zufriedenstellend beantwortet werden und wird als zentrale Forschungsfrage in dieser Arbeit aufgegriffen.

In Kapitel 4 wird die Fragestellung zunächst spezifiziert und als Dekomponierungsproblem eingeordnet. Dann werden methodische Aspekte zur Separierung von Enkodierungs- und Abrufprozessen behandelt. Wie sich nach einem Überblick über die in diesem Kontext relevanten Modelle herausstellt, kann kein vorhandenes mathematisches Modell zur Beantwortung der Lokalisationsfrage genutzt werden (Kapitel 4.2). Da aber bereits ähnliche Prozesse und Paradigmen modelliert wurden, wird u.a. auf das Enkodierung-Abruf-

Modell (Batchelder & Riefer, 1980, 1986, Kapitel 4.2.1) sowie auf Quellengedächtnismodelle (Batchelder & Riefer, 1990; Bayen, Murnane & Erdfelder, 1996; Meiser & Bröder, 2002, Kapitel 4.2.2) genauer eingegangen, bevor das neue Modell entwickelt wird. In Kapitel 4.2.3 werden die Modellannahmen detailliert dargestellt und die Parameter des Modells werden erläutert.

Die Entwicklung eines neuen Modells ist mit der Spezifizierung von Modellgleichungen nicht abgeschlossen. Das Modell muss sich auch empirisch bewähren: Es muss identifizierbar sein, eine gute Datenanpassung aufweisen und die Interpretation der Parameter muss über Validierungsexperimente sichergestellt werden (Erdfelder, 2000). Diese Evaluation wird im empirischen Teil der Arbeit vorgenommen.

Der zweite Teil der Arbeit beginnt mit Kapitel 5, in dem zunächst ein Überblick über das Vorgehen der Modellevaluation und die zugrunde liegenden Kriterien gegeben wird. Dann folgen vier Experimente (Kapitel 6 bis 8), die sich gezielt der Parametervalidierung zuwenden. Im ersten Experiment (1a) werden dabei zwei alternative Antwortformate einander gegenübergestellt und die Paradigmenunabhängigkeit der Modellparameter soll gezeigt werden. In Experiment 1b sollen die Itemparameter über unterschiedliche Präsentationszeiten validiert werden. Im zweiten Experiment wird in einem Vier-Gruppen-Enkodierung-Abruf-Design die Validierung der zentralen Parameter der assoziativen Enkodierung und des assoziativen Abrufs angestrebt. Dabei sollte der Enkodierungsparameter selektiv auf eine tiefe vs. flache relationale Enkodierungsmanipulation reagieren, der Abrufparameter sollte durch ein Retentionsintervall von drei Minuten vs. 24 Stunden signifikant herabgesetzt werden. Im dritten und vierten Experiment werden unterschiedliche Basisraten (Experiment 3) sowie Auszahlungsmatrizen (Experiment 4) eingesetzt, um Effekte auf die Rateparameter zu erzielen. Kapitel 9 schließt diesen empirischen Teil ab. Die Validierungsexperimente werden zusammengefasst und die Ergebnisse diskutiert. Da sowohl die Identifizierbarkeit, die Modellanpassung als auch die Validierung der Parameter erfolgreich war, wird das Modell im nächsten Teil in einer gerontologischen Studie angewandt.

Im dritten Teil wird das Modell zur Analyse eines Altersexperiments eingesetzt. Sowohl über die traditionelle als auch über die modellbasierte Auswertung konnte das typische assoziative Defizit repliziert werden. Darüber hinaus erlaubt die multinomiale Analyse den spezifischen Altersvergleich in Enkodierungs- bzw. Abrufprozessen. Die Ergebnisse dieses Altersvergleichs werden erläutert und in Kapitel 10.3 diskutiert.

Im vierten Teil werden die wichtigsten Ergebnisse und Implikationen dieser Arbeit zusammengefasst. Dabei wird auch eine kritische Diskussion vorgenommen und ein Ausblick gegeben.

Teil I

Theoretischer Hintergrund

Kapitel 2

Theoretische Perspektiven der kognitiven Altersforschung

In den letzten 20 Jahren konnten unzählige empirische Befunde der kognitiven Altersforschung immer detaillierter spezifizieren, in welchen kognitiven Leistungsbereichen sich ältere und jüngere Menschen gar nicht oder nur kaum voneinander unterscheiden, und in welchen Gebieten sich große Altersunterschiede aufzeigen. Wie stark das Feld der kognitiven Altersforschung in dieser Zeit gewachsen ist, kann an der Zahl der Veröffentlichungen festgemacht werden. So finden sich in den psychologischen Datenbanken (PsycInfo und PsycArticles) zu dem Stichwort ‚kognitives Altern‘ (‚cognitive aging‘) im Zeitraum von 1976-1996 genau 335 Treffer, in den darauf folgenden 10 Jahren von 1997-2007 verdreifacht sich die Anzahl der relevanten Publikationen auf 1085. Ein verstärktes Interesse an der Thematik des kognitiven Alterns zeigt sich nicht nur in der Wissenschaft, sondern auch in der Öffentlichkeit. Nicht zuletzt aufgrund des demografischen Wandels in den Industrienationen und der Erhöhung des Renteneintrittsalters treten Fragen zur kognitiven Leistungsfähigkeit im Alter immer mehr in den Vordergrund.

In wissenschaftlichen Arbeiten zum kognitiven Altern werden zunehmend nicht nur Verhaltensdaten, sondern auch neurophysiologische Daten erhoben, die Altersunterschiede in kognitiven Leistungen mit neurophysiologischen Prozessen in Verbindung setzen (z.B. Raz, 2000). Insbesondere in Verbindung mit experimentellen Ansätzen eröffnen diese Methoden einen vielversprechenden Zugang zu offenen Fragen des kognitiven Alterns. Hinsichtlich der Forschungsinteressen sind neue Fragen zur intraindividuellen Variabilität von Leistungen (z.B. S.-C. Li & Baltes, 2006; Lövdén & Lindenberger, 2005; Lövdén, Li, Shing & Lindenberger, 2007), zur Entscheidungsforschung (z.B. Kim, Healey, Goldstein, Hasher & Wiprzycka, 2008; Mata, 2007; Mata, Schooler & Rieskamp, 2007), zur Interaktion von Emotion und Kognition (Carstensen & Mikels, 2005; Mather & Carstensen, 2005) sowie zu den Grenzen und Möglichkeiten von Interventionen (z.B. Ball et al., 2002; Brehmer et al., 2008; Salthouse, 2006) untersucht worden.

Während sich also die Forschungsmethoden und Forschungsfragen dieser Disziplin in den letzten zwei Jahrzehnten verändert haben und die große Anzahl an empirischen Studien ein immer genaueres Bild des alternden Menschen ermöglicht, so fällt auf, dass sich auf der Ebene der Alternstheorien ein eher stabiles Bild abzeichnet (siehe aber Carstensen, Fung & Charles, 2003; Carstensen & Mikels, 2005). Klassische Alternstheorien der kognitiven Verlangsamung oder der Inhibition - um nur zwei zu nennen - bilden weiterhin den theoretischen Ausgangspunkt vieler empirischer Studien (z.B. Kim, Hasher & Zacks, 2007; McCabe & Hartman, 2008). Neue Theorien wurden u.a. entwickelt, um Altersunterschiede in der Interaktion von Emotion und Kognition zu erklären (Carstensen et al., 2003); sie basieren daher auch nicht mehr ausschließlich auf kognitiven Konstrukten, sondern beziehen soziologische Konstrukte wie z.B. Lebensziele mit ein.

Gerade diese neueren Ansätze verdeutlichen, dass zur Erklärung von *kognitiven* Leistungsunterschieden im Alter unter Umständen nicht nur kognitive Mechanismen verantwortlich gemacht werden können. Die Altersforscherin Mather z.B. zeigte, dass die Erinnerung an Produkteigenschaften nicht nur von einer allgemeinen Gedächtnisschwäche älterer Menschen bestimmt wird, sondern dass die Wahl bzw. der Kauf eines Produkts ebenfalls die Erinnerung beeinflusst (Mather & Johnson, 2000, für alternative Erklärungen siehe Kapucu, Rotello, Ready & Seidl, 2008). Selbst wenn man sich also mit einer - auf den ersten Blick - rein kognitiven Forschungsfrage wie der Erinnerung an Produkteigenschaften beschäftigt, ist es wichtig, möglichst *alle* für diese Aufgabe relevanten Aspekte zu berücksichtigen.

Dies gilt auch für die Forschungsfrage, die dieser Arbeit zugrunde liegt. Zwar geht es in dieser Arbeit um eine kognitive Fragestellung, trotzdem muss abgeklärt werden, inwieweit z.B. Faktoren des Bildungshintergrunds oder der Motivation die kognitive Leistung beeinflussen. Aus diesem Grund wird zunächst ein umfassendes Konzept von theoretischen Perspektiven der kognitiven Altersforschung vorgestellt, bevor dann detaillierter auf den direkten theoretischen Hintergrund der Forschungsfrage eingegangen wird.

Salthouse hat 1991 ein Klassifikationssystem bestehend aus fünf theoretischen Perspektiven vorgeschlagen. Dieses System von disjunkten und exhaustiven theoretischen Perspektiven soll eine Klassifikation von sowohl Theorien als auch empirischen Befunden der kognitiven Altersforschung ermöglichen. Jede theoretische Perspektive beschäftigt sich dabei aus einem unterschiedlichen Blickwinkel mit der Ursache von kognitiven Altersunterschieden. Diese Blickwinkel sind: (1) Veränderungen der Umwelt (Kohorteneffekte), (2) nachlassende Nutzung von kognitiven Prozessen, (3) qualitative Unterschiede der kognitiven Verarbeitung, (4) reduzierte kognitive Basisprozesse, (5) analytische Ansätze der Lokalisation.

Der Begriff ‚theoretische Perspektive‘ impliziert dabei nicht, dass die unter diesem Gesichtspunkt zusammengefassten Erklärungsansätze notwendigerweise den Kriterien wissenschaftlicher Theorien (Westermann, 1987) genügen müssen. Salthouse unterscheidet

mehrere hierarchische Ebenen des theoretischen Diskurses, die sich von Weltansichten, Rahmenkonzepten, wissenschaftlichen Theorien und Modellen bis hin zu deskriptiven Generalisierungen erstrecken. All diese Ebenen des theoretischen Diskurses können in einer theoretischen Perspektive vertreten sein.

Auch empirische Beobachtungen sollen den theoretischen Perspektiven zugeordnet werden. Dies kann im einfachsten Fall dadurch geschehen, dass die empirische Fragestellung bereits explizit eine theoretische Motivierung aufweist und anhand dieser Motivierung in das Salthouse'sche Konzept eingeordnet werden kann. Allerdings ist es manchmal der Fall, dass Altersstudien einen sehr explorativen Charakter aufweisen und es ihnen an theoretischer Motivierung mangelt. Dies trifft insbesondere auf frühe Studien der Altersforschung zu. Wird eine theoretische Zuordnung nicht explizit getroffen, so kann die Forschungsarbeit in vielen Fällen trotzdem in das System eingeordnet werden, da die Anlehnung an eine der Salthouse'schen Perspektiven zu erkennen ist. Wenn sich die empirischen Forschungsergebnisse keiner Theorie bzw. keiner theoretischen Perspektive zuordnen lassen, so vertritt Salthouse den Standpunkt, dass diese auch keinen entscheidenden Beitrag zur Weiterentwicklung der Forschung leisten und damit nicht bedeutsam sind. Seiner Ansicht nach sollten Forschungsarbeiten immer theoriegeleitet, im Mindesten theoriebezogen sein; ist dies nicht der Fall, so ist es auch nicht notwendig, sich weiter mit diesen Ergebnissen zu beschäftigen bzw. diese in ein Klassifikationssystem aufzunehmen. Nach Salthouse stellt dieser Selektionsaspekt einen weiteren Vorteil seines Ansatzes dar.

Welche Bedeutung diesem Klassifikationssystem beizumessen ist, kann dadurch erahnt werden, dass die Arbeit in den ersten 15 Jahren ihres Erscheinens bereits mehr als 500 Mal zitiert wurde. Es ist also nicht übertrieben, dieses Konzept als eine klassische und zentrale Arbeit der kognitiven Altersforschung zu bezeichnen. Die Systematik zeichnet sich auch dadurch aus, dass es möglich ist, aktuelle Arbeitshypothesen der kognitiven Altersforschung und neue Theorien in das System zu integrieren und es damit problemlos auf den aktuellen Forschungsstand zu bringen.

Die fünf theoretischen Perspektiven werden nun kurz erläutert, damit dann die Forschungsfrage dieser Arbeit und deren theoretischer Hintergrund in das System eingeordnet werden können. Für eine ausführliche Beschreibung der Perspektiven und der umfassenden empirischen Befunde sei auf die Arbeit von Salthouse (1991) verwiesen. An dieser Stelle werden die Perspektiven nur in ihren Grundzügen mit selektiven Beispielen vorgestellt.

„It is frequently more plausible to argue that people of different ages differ on a given characteristic because they belong to a different generation than because they differ in age.“ (Schaie, 1975, S. 113). Dieses Zitat beschreibt den Ansatz, der die Ursache für Altersunterschiede in *Veränderungen der Umwelt (Kohorteneffekten)* sieht.

Positive Umweltveränderungen wie z.B. eine verbesserte Ernährungs- und Gesundheitssituation von Kindern wirken sich positiv auf die Intelligenzentwicklung aus; dies konnte

über die letzten Jahrzehnte in der immer fortwährenden Erhöhung der Intelligenz (Flynn-Effekt; Flynn, 1984) beobachtet werden. Ebenso kann der Zugang zu Bildungssystemen sowie die in diesen Systemen verbrachte Zeit zwischen Generationen variieren und Auswirkungen auf bestimmte kognitiven Fähigkeiten haben. Selbst wenn man nur Personengruppen vergleichen würde, die hinsichtlich der Dauer der Schulausbildung parallelisiert sind, können Unterschiede in der Art der schulischen Bildung und Förderung zwischen den Generationen existieren. Solche Effekte, die zusammenfassend als *Kohorteneffekte* bezeichnet werden, könnten Altersunterschiede vortäuschen, obwohl die Unterschiede zwischen den Stichproben nicht auf das Alter zurückgehen, sondern auf Veränderungen der Umwelt. Vor allem in Querschnittsuntersuchungen müssen diese Effekte daher unbedingt berücksichtigt werden (für einen Überblick siehe Erdfelder & Undorf, 2003). Über ein entsprechendes Untersuchungsdesign (Schaie, 1965, 1994) sowie die Aufnahme von Kohorte und Testzeitpunkt in die Datenanalyse (Erdfelder, 1987) kann der Einfluss von Kohorteneffekten geschätzt und damit kontrolliert werden. Erdfelder (1987) zeigte in seinen Analysen, dass der Kohorteneffekt nicht zu vernachlässigen ist. Dennoch kann über Kohorteneffekte nicht der gesamte Altersunterschied in kognitiven Leistungen erklärt werden (Salthouse, 1991). Unterschiede in der Umwelt, die sich in Kohorteneffekten widerspiegeln, tragen also zu Unterschieden zwischen Altersgruppen bei, sind aber nicht die alleinige Ursache von Altersunterschieden in kognitiven Aufgaben.

Der Einfluss der Umwelt als Erklärungsansatz von Altersunterschieden wurde m.W. nie in einer psychologischen Theorie festgehalten. In dieser Hinsicht handelt es sich hier eher um die Formulierung eines Rahmenkonzepts, in dem Hypothesen aufgestellt wurden. Empirisch konnte ein Einfluss der Umwelt quantifiziert werden, die Herausforderung besteht aber darin, diesen Einfluss zu erfassen, zu kontrollieren und gering zu halten, um andere potentielle Ursachen kognitiven Alterns untersuchen zu können. Die Berücksichtigung von Kohorteneffekten muss also bei der Interpretation von Querschnittsstudien immer bedacht werden und sollte durch eine Parallelisierung der Stichproben hinsichtlich Kontrollvariablen wie Bildungsstand und Gesundheit zumindest minimiert werden. Die Suche nach der Ursache von Altersunterschieden ist mit diesem Ansatz allerdings nicht erschöpfend aufgeklärt.

Unter der zweiten theoretischen Perspektive werden Ansätze subsumiert, die eine *nachlassende Nutzung kognitiver Ressourcen* („Disuse“) als Ursache der verringerten kognitiven Leistung älterer Menschen annehmen. Der Grundgedanke kann mit den Worten „Use it - or lose it“ bzw. „Wer rastet, der rostet“ beschrieben werden. Auch hier wurden die Hypothesen über den Einfluss von kognitiver Aktivität eher in losen Annahmen formuliert, als in einer wissenschaftlichen Theorie. Insbesondere zwei Hypothesen stehen dabei im Vordergrund:

1. Im Umkehrschluss wird aus „Wer rastet, der rostet“ häufig die Schlussfolgerung gezogen, dass Menschen, die ihr Leben lang, insbesondere bis ins hohe Alter hinein, geistig sehr aktiv sind, nicht rosten.¹ Diese Hypothese führte zu der Annahme, dass z.B. Universitätsprofessoren im Alter keine kognitiven Leistungseinbußen aufweisen sollten.

2. Darüber hinaus wird vermutet, dass sich Altersunterschiede durch ein Training der kognitiven Fähigkeiten aufheben lassen. Diese zweite Annahme ist zumindest insoweit gerechtfertigt, als dass die ‚Disuse-Hypothese‘ nicht spezifiziert, ob ein mangelnder Gebrauch kognitiver Ressourcen reversible oder irreversible Konsequenzen nach sich zieht.

Untersuchungen von Shimamura und Kollegen (Shimamura, Berry, Mangels, Rusting & Jurica, 1995) konnten zeigen, dass auch bei älteren Universitätsprofessoren ein Nachlassen von kognitiven Fähigkeiten festzustellen ist. Zwar waren in manchen kognitiven Maßen (Erinnerung an Prosatexte) keine Altersunterschiede festzustellen, aber das Gesamtergebnis der Studie lässt keinesfalls den Schluss zu, dass bei geistig sehr aktiven Menschen im Alter keine kognitiven Leistungseinbußen vorliegen. Die geistig aktiven Professoren dieser Studie wiesen durchaus eine bessere Leistung auf, als eine gleichaltrige bevölkerungsrepräsentative Stichprobe, aber es wurde sowohl ein Defizit gegenüber Studenten als auch gegenüber 30-40-jährigen Professoren verzeichnet. Ähnliche Ergebnisse wurden auch von anderen Studien mit akademischen Stichproben gefunden (Christensen & Henderson, 1991; Compton, Bachman, Brand & Avet, 2000; Sward, 1945). „Use it - or lose it“ konnte in diesen Untersuchungen nicht bestätigt werden, doch weisen die Befunde auf „Use it - and keep it longer“ hin (Schooler, 2007).

Die zweite Hypothese wurde mit Trainingsansätzen wie ‚Testing-the-Limits‘ (Baltes & Kliegl, 1992; Kliegl, Smith & Baltes, 1989) untersucht. Diese Studien konnten bestätigen, dass ältere Menschen von einem Gedächtnistraining profitieren. Allerdings wurde der Altersunterschied zwischen trainierten älteren und trainierten jüngeren Probanden nicht aufgehoben, sondern vergrößerte sich sogar. Eliminiert werden Altersunterschiede durch kognitive Trainings also nicht. Salthouse (2006) hat in einem Überblicksartikel zur ‚Use it - or lose it‘-Hypothese die empirische Evidenz zu Trainingsansätzen evaluiert und kommt zu einem ernüchternden Ergebnis: Seiner Ansicht nach konnte bisher keine Trainingsstudie unter experimentellen Standards ihren Erfolg nachweisen. In einer Reaktion auf diesen Artikel argumentierte Schooler (2007) gegen diese Schlussfolgerung. Er wirft Salthouse vor, zu strenge Kriterien anzulegen. Das Kriterium für einen Trainingserfolg erweist sich als kritischer Aspekt in der Diskussion. Salthouse (2007) vertritt den Standpunkt, dass ein Trainingserfolg erst dann vorliegt, wenn die trainierte Seniorengruppe einen anderen Entwicklungsverlauf zeigt, als die untrainierte Gruppe. Er postuliert also eine Wechselwirkung von Alter und Training in Bezug auf die kognitiven Leistungen. Dieses Kriterium wird in der Tat selten erreicht (Ball et al., 2002; Salthouse, 2006; Schaie, 2005). Schooler

¹Zwar ist dieser Umkehrschluss logisch nicht korrekt, trotzdem wurde er als Hypothese formuliert und empirisch untersucht.

(2007) dahingegen sieht einen einfachen Haupteffekt des Trainings bereits als ausreichend an, um von einem Trainingserfolg zu sprechen. Nach seiner Argumentation führt bereits ein einfacher Haupteffekt dazu, dass die trainierte Gruppe ihren kognitiven Status länger aufrechterhalten kann. Allein dies ist seiner Ansicht nach als erfolgreiches Trainingsergebnis zu würdigen. Dieses weniger strikte Kriterium wurde in vielen empirischen Studien erreicht und unterstützt - nach Schooler - den Nutzen von Trainingsmaßnahmen (siehe Schooler, 2007).

Zusammenfassend ist festzuhalten, dass man durch Trainingsinterventionen zwar Leistungen verbessern kann, es aber nicht gelingt, Altersunterschiede aufzuheben. Auch die mangelnde Nutzung von kognitiven Ressourcen kann nicht als Ursache der kognitiven Altersunterschiede angesehen werden.

„It is still not certain that some differences in the cognitive behavior of young and old adults are primarily a matter of deficit, since it is possible that adulthood and aging bring qualitative changes that may mimic decrements but in fact signal adaptive reorganization.“ (Labouvie-Vief, 1985, S. 519).

Spiegeln Altersunterschiede quantitative oder adaptive qualitative Verarbeitungsunterschiede des kognitiven Systems wider? Zunächst wird die Hypothese qualitativer Verarbeitungsunterschiede dargestellt, danach wird auf die Hypothese veränderte quantitativer Mechanismen eingegangen.

Am Beispiel der theoretischen Perspektive qualitativer Verarbeitungsunterschiede wird an dieser Stelle gezeigt, wie neue Theorien in das Salthouse'sche Klassifikationssystem eingeordnet werden können. Carstensen (Carstensen, 1992, 1995; Carstensen et al., 2003; Carstensen & Mikels, 2005) hat vor einigen Jahren die sozioemotionale Selektivitätstheorie formuliert, die seither häufig in Studien untersucht wurde (z.B. Mather & Johnson, 2000; Mather, Knight & McCaffrey, 2005). Die Grundannahme dieser Theorie besteht darin, dass ältere Menschen aufgrund ihrer begrenzt wahrgenommenen Lebenszeit andere Prioritäten haben und andere Ziele verfolgen als jüngere Menschen. Kurz- sowie langfristige emotionale Zufriedenheit ist ein wichtiges Ziel für ältere Menschen; für jüngere Menschen stehen Informationsgewinn und langfristige existenzielle Ziele im Vordergrund. Mit dieser Theorie lassen sich z.B. der Effekt der entscheidungsunterstützenden Erinnerung (Mather & Johnson, 2000) oder die Positivitätsverzerrung gegenüber emotionalen Stimuli (Charles, Mather & Carstensen, 2003) gut erklären. Nach einer getroffenen Entscheidung steht für ältere Menschen die Zufriedenheit mit der Entscheidung im Vordergrund. In einem Gedächtnistest ordnen sie daher positive Produkteigenschaften mit einer hohen Tendenz der gewählten Alternative zu (Mather & Carstensen, 2005; Mather & Johnson, 2000). Für jüngere Menschen dahingegen ist es wichtig, positive sowie negative Aspekte den Produktalternativen korrekt zuzuordnen, damit diese in zukünftigen Entscheidungen berücksichtigt werden können. Der Informationsgewinn ist bedeutender für die Ziele der jüngeren

Menschen als die emotionale Zufriedenheit; das Umgekehrte gilt für die älteren Menschen. Diese neue motivationale Theorie fokussiert auf qualitativ unterschiedliche Strategien und Ziele der beiden Altersgruppen. Sie ist also eindeutig der theoretischen Perspektive der qualitativen Unterschiede zuzuordnen und kann problemlos in das Klassifikationssystem eingeordnet werden.

Neben motivationalen Unterschieden werden innerhalb dieser Perspektive Unterschiede in der Strategieverwendung, der Einfluss von Metakognitionen sowie die Auswirkungen von Altersstereotypen auf die kognitive Leistung zusammengefasst. Der Einfluss von Altersstereotypen kann in Kulturen mit einem eher negativen Altersstereotyp untersucht werden. In den USA z.B. zeigte sich, dass ältere Menschen tatsächlich eine verschlechterte kognitive Leistung aufwiesen, wenn das negative Altersstereotyp durch Zeitungsartikel über Gedächtnisprobleme älterer Menschen aktiviert wurde. Wurde nach dem Lesen eines solchen Artikels ein Experiment zur *Gedächtnisleistung* angekündigt und durchgeführt, so zeigten Senioren eine verringerte kognitive Leistung im Vergleich zu einer Gruppe bei der das negative Stereotyp nicht durch die Lektüre aktiviert wurde, der *Gedächtnistest* nicht angekündigt wurde, aber trotzdem stattfand (Rahhal, Hasher & Colcombe, 2001). Auch hinsichtlich der Nutzung von Strategien konnte festgestellt werden, dass ältere Menschen spontan seltener Gedächtnisstrategien verwenden (Dunlosky & Hertzog, 2001). Natürlich stellt sich an dieser Stelle die Frage, ob dies ein Produktionsdefizit ist und damit prinzipiell aufgehoben werden kann, oder ob die mangelnde Verwendung von Strategien auf ein Ressourcendefizit zurückgeht und ältere Menschen aus diesem Grund Strategien nicht effizient anwenden können. Die bereits erwähnte Studie von Baltes und Kliegl (1992) zeigte, dass Altersunterschiede in der kognitiven Leistung bestehen bleiben, auch wenn beide Altersgruppen dieselbe Strategie anwenden und diese Strategie lange - über mehrere Monate hinweg - geübt wurde. Dieser Befund wurde bereits dahingehend interpretiert, dass Altersunterschiede in der Strategieverwendung die kognitiven Leistungsunterschiede zwar moderieren, nicht aber ursächlich erklären können. Auch die anderen erwähnten qualitativen Mechanismen haben einen moderierenden Einfluss auf die Altersunterschiede in kognitiven Leistungen, liegen diesen aber keinesfalls als alleinige Ursachen zugrunde (zusammenfassend siehe Salthouse, 1991).

Mangelnde Ressourcen und *reduzierte kognitive Basisprozesse* sind der Fokus einer weiteren theoretischen Perspektive im Salthouse'schen Klassifikationssystem. Die Grundrichtung dieser Perspektive kann an einem Zitat von Birren, Riegel und Morrison (1962, S. 10) verdeutlicht werden: „It would seem that the principal function of psychological research on aging is to reduce the great variety of changes in behavior associated with age to a smaller number of concepts.“ Als solche Konzepte kommen z.B. eine kognitive Verlangsamung (Salthouse, 1996), defizitäre Inhibitionsprozesse (Hasher & Zacks, 1988), reduzierte Aufmerksamkeitsressourcen (Craik & Byrd, 1982) oder auch defizitäre assoziative

Gedächtnisprozesse (Naveh-Benjamin, 2000) in Frage. Die grundlegende Idee all dieser Ansätze ist es, die Altersunterschiede in kognitiven Leistungen auf einen zentralen kognitiven Verarbeitungsmechanismus zurückzuführen, der im Laufe des Alterns in seiner Quantität abnimmt. Da diese kognitiven Basismechanismen für die vorliegende Arbeit von zentraler Bedeutung sind, werden sie im nächsten Kapitel genauer ausgeführt.

In der fünften und letzten theoretischen Perspektive wird nicht mehr die Frage der Ursache von Altersunterschieden behandelt, sondern der Fokus liegt auf der genauen *Lokalisationsanalyse* der Altersunterschiede. Ein Ergebnis solcher Lokalisationsfragen ist z.B., dass sich Altersunterschiede kaum in impliziten, sondern vor allem in expliziten Gedächtnisaufgaben finden (Light & Singh, 1987). Auch fallen die Altersunterschiede in episodischen Gedächtnistests zu Ungunsten von älteren Menschen aus, in semantischen Gedächtnisaufgaben hingegen kehrt sich dieses Muster um (für einen Überblick siehe Zacks et al., 2000). Durch die Erkenntnisse aus dieser Lokalisationsforschung können wiederum Hypothesen über die zugrunde liegenden Basismechanismen formuliert werden. In der Ursachenforschung erweist sich eine genaue Lokalisation als notwendig und rechtfertigt daher die Aufnahme der Lokalisationsperspektive in das Salthouse'sche Klassifikationssystem.

Der Klassifikationsansatz von Salthouse bietet Raum um empirische Befunde und theoretische Konzepte zum kognitiven Altern systematisch zu ordnen; ein klarer und strukturierter Zugang zu einem komplexen und umfangreichen Forschungsfeld wird durch diese Systematik erleichtert.

Um die Ursachen des kognitiven Alterns zu erforschen, scheinen manche Perspektiven, z.B. die der reduzierten kognitiven Prozesse, besonders vielversprechend zu sein. Forschungsergebnisse aus allen Perspektiven haben aber bedeutende Beiträge zur Erklärung von Altersunterschieden geliefert und es gilt, Erkenntnisse aus diesen Bereichen bei der weiteren Suche nach den Ursachen des kognitiven Alterns zu berücksichtigen. So hat die methodische Diskussion zu Kohorteneffekten und den Problemen der Querschnittsuntersuchung dazu geführt, dass in Altersexperimenten standardmäßig Kontrollvariablen erfasst werden, um die Altersgruppen (weitestgehend) zu parallelisieren. Hierzu gehören u.a. objektive Gesundheitsdaten und die Dauer der Schulbildung (Hertzog & Dixon, 1996, S. 86). Durch die Erfassung des Wortschatzes und der verbalen Fähigkeiten soll sichergestellt werden, dass beide Gruppen vergleichbar mit verbalem Material umgehen können und dies nicht zu einem scheinbaren Altersunterschied führt. Forschungsarbeiten über Auswirkungen der Aktivierung eines negativen Altersstereotyps durch Formulierungen wie ‚Gedächtnistests‘ (Rahhal et al., 2001) haben zur Folge, dass solche Ausdrücke in Instruktionen oder Ausschreibungen vermieden werden. Dies zeigt, dass Altersforschung unbedingt aus all diesen Perspektiven betrachtet werden muss. Es gilt, die Perspektiven zu verknüpfen, nicht zu isolieren, um ein umfassendes Verständnis des Forschungsgegenstands zu erlangen.

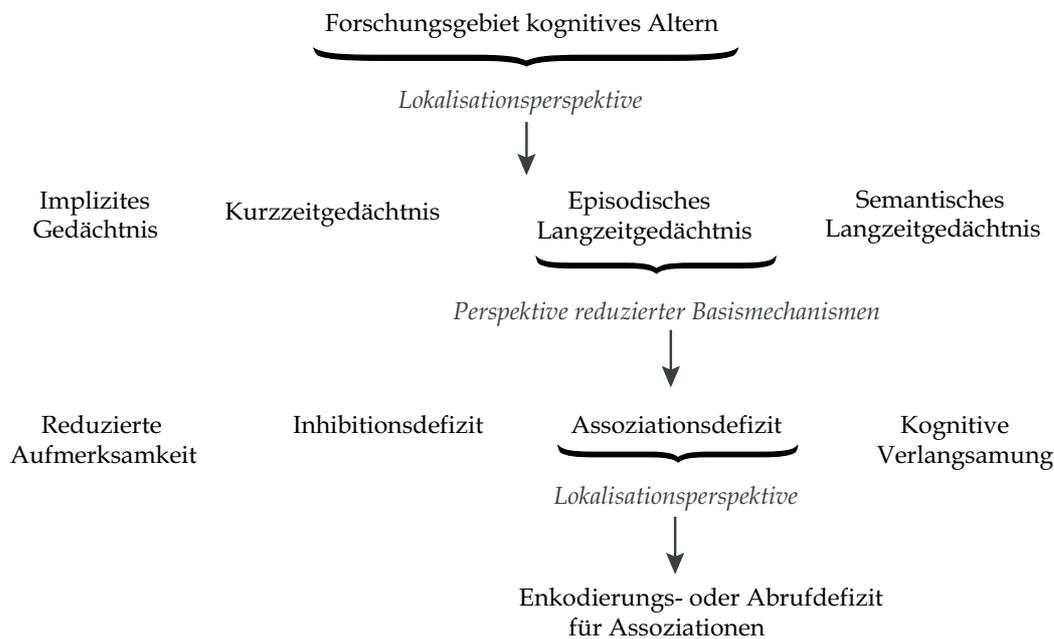


Abbildung 2.1: In dieser Abbildung wird veranschaulicht, wie die Forschungsfrage dieser Arbeit in das Feld der kognitiven Altersforschung eingeordnet werden kann. Zunächst wird das globale Forschungsfeld des kognitiven Alterns durch die Lokalisationsperspektive auf das Gebiet des episodischen Langzeitgedächtnisses eingeschränkt. Innerhalb des episodischen Langzeitgedächtnisses wird dann die Perspektive der reduzierten Verarbeitungsressourcen herangezogen und das Assoziationsdefizit als Erklärungsansatz fokussiert. Dieser Basismechanismus kann aus Sicht der Lokalisationsperspektive weiter unterteilt werden in ein assoziatives Defizit in der Enkodierungsphase und ein Defizit während des Abrufs.

Theoretische Perspektiven dieser Arbeit

In diesem Abschnitt soll der theoretische Rahmen der vorliegenden Arbeit spezifiziert und eingegrenzt werden. Zu diesem Zweck wird auf das System von Salthouse (1991) zurückgegriffen. Es wird herausgearbeitet, welche theoretischen Perspektiven für die vorliegende Forschungsfrage relevant sind. Dadurch soll die theoretische Verankerung der Fragestellung klar werden, und es soll damit auch der Aufbau des folgenden Theorieteils begründet werden. In Abbildung 2.1 wird die folgende Einteilung grafisch dargestellt.

Die erste Eingrenzung wird hinsichtlich der *analytischen Lokalisationsperspektive* getroffen: Diese Arbeit beschäftigt sich ausschließlich mit Altersunterschieden in *episodischen Langzeitgedächtnisaufgaben* (Tulving, 1983). Es werden also weder Altersunterschiede im Arbeitsgedächtnis oder in sensorischen Gedächtnissystemen thematisiert, noch sollen Alterseffekte in semantischen Langzeitgedächtnisaufgaben diskutiert werden.

Zur Erklärung von Altersunterschieden in episodischen Langzeitgedächtnisaufgaben können unterschiedliche *reduzierte kognitive Ressourcen* herangezogen werden: reduzierte Aufmerksamkeitsprozesse (Craik & Byrd, 1982), defizitäre Inhibitionsmechanismen (Hasher & Zacks, 1988), defizitäre assoziative Prozesse (Chalfonte & Johnson, 1996; Naveh-Benjamin,

2000) oder eine kognitive Verlangsamung (Salthouse, 1996). Diese verschiedenen kognitiven Prozesse werden im folgenden Theorieteil vorgestellt. Der Fokus wird dabei auf dem Ansatz des *assoziativen Gedächtnisdefizits* liegen.

Auf der Subebene dieses assoziativen Defizits können aus Sicht der *Lokalisationsperspektive* weitere Fragestellungen formuliert werden, z.B. ob sich dieses Defizit nur in expliziten oder auch in impliziten Gedächtnisaufgaben zeigt (Dew, Bayen & Giovanello, 2007) oder ob das Defizit ein Enkodierungsproblem oder eher ein Abrufproblem darstellt. Diese Frage, ob das assoziative Defizit auf Enkodierungs- oder Abrufprozesse zurückgeht, stellt die zentrale Forschungsfrage dieser Arbeit dar. Zur theoretischen Einbettung und Abgrenzung des Forschungsgegenstandes soll daher auf Alternstheorien reduzierter kognitiver Prozesse, insbesondere auf die Theorie des assoziativen Defizits unter dem Blickwinkel der Lokalisationsperspektive eingegangen werden.

Kapitel 3

Defizitäre assoziative Basismechanismen

Nachdem die vorliegende Forschungsfrage in das System von Salthouse (1991) eingeordnet wurde und die relevanten theoretischen Perspektiven spezifiziert wurden, soll sich der Aufbau des Theorieteils an diesem System anlehnen.

In diesem Kapitel werden defizitäre assoziative Gedächtnisprozesse als dem kognitiven Altern zugrunde liegende Basismechanismen diskutiert. In theoretischer Hinsicht wird der Fokus dabei auf der Assoziationsdefizit-Hypothese (Naveh-Benjamin, 2000) liegen. Dem Kapitel sei vorangestellt, dass der Gedanke, assoziative Gedächtnisprozesse als Basismechanismen zu betrachten, Gegenstand der aktuellen Forschung ist und daher als theoretisch umstritten und empirisch wenig gefestigt betrachtet werden muss. Zunächst wurden Defizite in assoziativen Gedächtnisprozessen eher als Folge anderer Basismechanismen wie z.B. kognitiver Verlangsamung oder reduzierter Aufmerksamkeitsprozesse interpretiert (Salthouse, 1991). Diese Einordnung wurde allein auf theoretischer Ebene vorgenommen und nicht durch empirische Befunde gestützt. Erst in den letzten Jahren fand diese Frage Eingang in die empirische Forschung. Die bisherigen Befunde sprechen dafür, in assoziativen Gedächtnisprozessen einen eigenständigen Basismechanismus zu sehen (Naveh-Benjamin, 2008a; Naveh-Benjamin, Kilb & Hara, 2007; Shing et al., 2008). Um die Einordnung als Basismechanismus zu rechtfertigen, muss gezeigt werden, dass assoziative Gedächtnisprozesse einen eigenständigen Beitrag an der Aufklärung der Altersvarianz leisten, der nicht durch andere Basisprozesse aufgeklärt werden kann. Um mit der vorliegenden Arbeit an den aktuellen Forschungsstand anzuknüpfen, wird das assoziative Gedächtnisdefizit hier als Basismechanismus parallel zu Ansätzen wie der kognitiven Verlangsamung aufgeführt. Zukünftige Forschungsergebnisse werden zeigen, inwieweit die vorgenommene Einteilung gerechtfertigt ist.

Die Hypothese des assoziativen Defizits hat sich historisch gesehen aus dem Ansatz der Kontextdefizit-Hypothese (Burke & Light, 1981) heraus entwickelt. Nicht nur aus his-

torischen Gründen, sondern auch aufgrund wichtiger empirischer Studien, methodischer Diskussionen und theoretischer Gemeinsamkeiten der beiden Ansätze soll daher dieses Kapitel mit der Kontextdefizit-Hypothese beginnen. Es werden zunächst die theoretischen Annahmen und empirischen Befunde zur Kontextdefizit-Hypothese dargestellt (Kap. 3.1), im Anschluss wird der Ansatz des assoziativen Defizits erläutert (Kap. 3.2). Dabei wird zunächst die Hypothese erklärt (Kap. 3.2.1) und relevante empirische Befunde werden diskutiert (Kap. 3.2.2). Abschließend wird auf das assoziative Defizit und dessen Abgrenzung zu anderen Basismechanismen (Kap. 3.2.4) eingegangen. Das Kapitel endet mit der Frage, ob das assoziative Defizit eher als Defizit in der Enkodierungs- oder in der Abrufphase zu verankern ist (Kap. 3.2.5).

3.1 Kontextdefizit-Hypothese

Die Kontextdefizit-Hypothese wurde systematisch zum ersten Mal von Burke und Light (1981) in ihrem Überblicksartikel zur Rolle von Abrufprozessen im Alter zusammengefasst. In diesem Artikel zeigen sie die Grenzen qualitativer Ansätze wie z.B. der Produktionsdefizit-Hypothese auf und weisen auf die Vorteile alternativer Ansätze wie u.a. der Kontextdefizit-Hypothese hin. Die zentrale Aussage der Kontextdefizit-Hypothese besteht darin, dass ältere Menschen Kontextinformationen nicht so effizient speichern und/oder abrufen können wie jüngere Menschen: „older adults may not store or retrieve certain types of contextual information as efficiently as younger adults“ (Burke & Light, 1981, S. 529). Als Ursachen für die weniger effiziente Verarbeitung von Kontextinformationen werden abnehmende Verarbeitungsressourcen sowie eine erhöhte Interferenzanfälligkeit älterer Menschen angenommen (Burke & Light, 1981). Das Defizit in der Verarbeitung von Kontextinformationen wird also klar als Folge von reduzierten Basismechanismen angesehen. Auch Salthouse (1991) schließt sich dieser Interpretation an; er sieht das Verarbeitungsdefizit für Kontextinformationen als Konsequenz verringerter Verarbeitungseffizienz während der Lern- oder Testphase an (Salthouse, 1991, S. 257).

Als Beispiele für Kontextinformationen haben Burke und Light (1981) zeitliche oder räumliche Elemente einer Gedächtnisepisode genannt. In einem klassischen Listenlernexperiment ist die Präsentationsreihenfolge der zu lernenden Wörter eine Kontextinformation, ebenso wie das Wissen darüber, welche Wörter in der oberen Hälfte des Bildschirms und welche in der unteren Hälfte dargeboten werden. Altersstudien im Rahmen der Kontextdefizit-Hypothese haben sich mit ganz unterschiedlichen Arten der Kontextinformation beschäftigt (zusammenfassend siehe Kausler, 1994). Es wurde u.a. untersucht, wie gut sich jüngere und ältere Menschen an die Darbietungsmodalität (Lehman & Mellinger, 1984), die Schriftfarbe (Park & Puglisi, 1985), die Schriftart (Kausler & Puckett, 1980), die zeitliche Reihenfolge (Kausler, Lichty & Davis, 1985; Kausler, Salthouse & Sauls, 1988; Wiley & Kausler, 1993), die räumliche Position (Denney, Dew & Kihlstrom, 1992; Park, Puglisi

& Lutz, 1982) oder auch an das Geschlecht der sprechenden Person (Kausler & Puckett, 1981b, 1981a) erinnern. In all diesen Studien ist intuitiv verständlich, welche Information als Kontextinformation und welche Information als zentraler Lerninhalt verwendet wird. Nichtsdestoweniger ist es für die Kontextdefizit-Hypothese wichtig, die Begriffe ‚Kontext‘ und ‚zentrale Information‘ zu definieren.

Von Burke und Light (1981) wurde diese Definitionsfrage nicht thematisiert. Post-Hoc schlugen Denney, Miller, Dew und Levav (1991) vor, den Begriff ‚zentrale Information‘ für die Information zu verwenden, die sich der Teilnehmer laut Anweisungen des Versuchsleiters einprägen soll. Alle weiteren Elemente der Episode, für die keine Lerninstruktion vorliegt, werden als Kontextinformation angesehen. Nach dieser Definition kann also potentiell jedes Element der Episode zur zentralen Information werden, indem der Versuchsleiter den Fokus auf dieses Element lenkt. Allerdings gibt es einige Studien, in denen die Teilnehmer aufgefordert wurden, sich mehrere Elemente einzuprägen, z.B. Wörter *und* deren Hintergrundfarbe. Nach der Definition von Denney et al. (1991) sollten in diesem Fall das Wort sowie dessen Hintergrundfarbe die zentrale Information darstellen. Die Präsentationsreihenfolge sei dahingegen eine Kontextinformation. Aber auch in solchen Studien interpretierten die Forscher das zu lernende Wort als zentrale Information und die Hintergrundfarbe als Kontext. Diese Experimente lassen sich also Post-Hoc nicht in die Definition von Denney et al. (1991) einordnen. Auch entwicklungspsychologische Studien an Kleinkindern (Rovee-Collier & Dufault, 1991) und Tierexperimente (Penick & Solomon, 1991; Phillips & LeDoux, 1992) stoßen bei der diskutierten Definition an Grenzen. In diesen Studien kann kaum durch verbale Instruktionen sichergestellt werden, welche Information als zentrale Information und welche Information als Kontextinformation angesehen werden soll. Daher kommen Chalfonte und Johnson (1996) zu dem Schluss, dass es nur eine Definition gibt, die sich über die Mehrheit der Studien hinweg konsistent anwenden lässt: „context is defined by what the experimenter views as the less salient or less predictive aspect of the experimental situation“ (Chalfonte & Johnson, 1996, S. 413). Als Kontextinformation soll also die Information angesehen werden, die der Experimentator als Kontext festlegt. Die Autoren räumen ein, dass dies eine wenig zufriedenstellende und äußerst pragmatische Definition ist, aber es scheint die einzige Definition zu sein, die sich Post-Hoc konsistent auf die empirischen Forschungsarbeiten zur Kontextdefizit-Hypothese anwenden lässt.

Empirische Befunde zur Kontextdefizit-Hypothese können nicht nur aus Studien gewonnen werden, die die direkte Überprüfung der Kontextdefizit-Hypothese zum Ziel haben. Auch Daten, die im Umfeld anderer Theorien bzw. Hypothesen erhoben wurden, lassen Rückschlüsse auf die Kontextdefizit-Theorie zu. Daher werden im Folgenden vier wichtige Untersuchungsparadigmen (zusammenfassend siehe Denney et al., 1991) eingeführt, aus denen empirische Evidenz für die Kontextdefizit-Hypothese ableitbar ist.

(1) Paradigma des intentionalen vs. inzidentellen Lernens. Arbeiten, die Gedächtnisleistungen in intentionalen vs. inzidentellen Lernsituationen vergleichen, beruhen auf einer Gegenposition zur Kontextdefizit-Hypothese. Hasher und Zacks (1979) postulierten, dass Kontextinformationen automatisch verarbeitet werden und daher nicht alterssensitiv sind. In vielen Studien nachgewiesen werden, dass bei älteren Menschen automatische Verarbeitungsprozesse gut erhalten bleiben, wohingegen vor allem Prozesse der bewussten Erinnerung abnehmen (z.B. Light, 1991; Light & Singh, 1987; Light, Singh & Capps, 1986). Neu war die Behauptung, dass Kontextinformationen automatisch verarbeitet werden und damit keinem Altersdefizit unterliegen sollten. Diese Hypothese (Hasher & Zacks, 1979) postuliert damit genau das Gegenteil der Kontextdefizit-Hypothese.

Zur empirischen Überprüfung dieses Ansatzes wurden vor allem Experimente² durchgeführt, die intentionale und inzidentelle Lernsituationen miteinander verglichen. Nach Hasher und Zacks (1979) sollten bei älteren Menschen Erinnerungen an Kontextinformationen unabhängig von der Lernbedingung vorhanden sein. Des Weiteren sollte unabhängig von der Lernbedingung *kein* Altersunterschied in der Kontexterinnerung vorliegen. Diese Hypothese konnte jedoch in zahlreichen Studien widerlegt werden, da ältere Menschen häufig ein Verarbeitungsdefizit für Kontextinformationen sowohl unter intentionalen als auch inzidentellen Lernbedingungen aufweisen (Kausler et al., 1985; Lehman & Mellinger, 1984; Light & Zelinski, 1983; Park et al., 1982; Park & Puglisi, 1985). Die Annahme, dass Kontextinformationen einer automatischen Verarbeitung unterliegen (Hasher & Zacks, 1979), wurde daher aufgegeben; die Ergebnisse wurden zunächst als unterstützend für die Kontextdefizit-Hypothese angesehen. Ob diese Befunde jedoch aufgrund ihrer Methodik zur Überprüfung der Kontextdefizit-Hypothese herangezogen werden können, wird später in diesem Kapitel diskutiert.

(2) Paradigma der Enkodierungsspezifität. Eine weitere Reihe von Studien wurde durch die Enkodierungsspezifitätshypothese (Tulving & Thomson, 1973) angestoßen. Diese trifft die Annahme, dass die Erinnerung an zentrale Informationen umso besser gelingt, je stärker der Kontext des Abrufs mit dem Kontext der Enkodierung übereinstimmt. Nimmt man im Sinne von Burke und Light (1981) an, dass ältere Menschen Kontextinformationen weniger gut enkodieren und/oder abrufen können, so sollten sie weniger stark von einer Wiederherstellung des Enkodierungskontexts beim Abruf profitieren. In diesem Sinne wurde eine der ersten systematischen Studien zur Kontextdefizit-Hypothese von Rabinowitz, Craik und Ackerman (1982) vorgenommen. Sie haben in mehreren Experimenten untersucht, inwieweit Altersunterschiede in der Nutzung von spezifischen vs. allgemeinen Abrufhilfen vorliegen. Als spezifische Abrufhilfen wurden dabei u.a. selbstgenerierte Abrufhilfen für die zu lernenden Wörter verwendet. Als allgemeine Abrufhilfen wurden Oberkategorien

²Mit Experimenten sind hier Quasi-Experimente gemeint, da das Alter naturgemäß nicht experimentell manipuliert werden kann. Auch im Folgenden wird häufig von Altersexperimenten oder Experimenten mit älteren und jüngeren Menschen gesprochen, auch hier sind Quasi-Experimente gemeint.

zu den Items vorgegeben. Die Hypothese war, dass enkodierungsspezifische Abrufhilfen bei jüngeren Menschen einen besonders förderlichen Effekt auf die Erinnerungsleistung haben sollten. Bei Älteren sollte dies nicht der Fall sein. Die Ergebnisse bestätigten die Hypothese und zeigten, dass ältere Menschen von spezifischen, selbstgenerierten Abrufhinweisen nicht profitieren, jüngere dahingegen schon. Um sicherzustellen, dass diese Befunde nicht auf die weniger effiziente *Generierung* von Abrufhilfen bei älteren Menschen zurückzuführen sind, wurden auch vorgegebene spezifische vs. vorgegebene unspezifische Abrufhilfen untersucht. Das Ergebnismuster konnte repliziert werden.

Auch weitere Studien konnten mit diesem Paradigma die Enkodierungsspezifitätshypothese bei jüngeren, nicht aber bei älteren Menschen bestätigen (Park, Puglisi & Sovacool, 1984). Sie stehen damit im Widerspruch zu der Hypothese der automatischen Kontextenkodierung (Hasher & Zacks, 1979) und unterstützen die Annahme eines Verarbeitungsdefizits für Kontextinformationen.

(3) Quellengedächtnisparadigma. Auch die Quelle, aus der eine Information stammt, kann als Kontextelement einer Episode angesehen werden. Daher liefern sowohl die zahlreichen Forschungsarbeiten zu Altersunterschieden in klassischen Quellengedächtnisaufgaben (z.B. Bayen & Murnane, 1996; Ferguson, Hashtroudi & Johnson, 1992; Hashtroudi, Johnson & Chrosniak, 1989; Schacter, Kaszniak, Kihlstrom & Valdiserri, 1991) sowie die Studien zur Erinnerungsleistung an das Ausführen vs. Vorstellen einer Handlung (Cohen & Faulkner, 1989; Guttentag & Hunt, 1988) oder an das laute Aussprechen vs. stumme Lesen eines Wortes (Hashtroudi et al., 1989) auf den ersten Blick stützende Evidenz für die Kontextdefizit-Hypothese. Inwieweit die Befunde aus der Quellengedächtnisforschung überhaupt zur Überprüfung der Kontextdefizit-Hypothese herangezogen werden können, wird im Folgenden noch aufgegriffen. Unabhängig von der Adäquatheit dieses Paradigmas zur Überprüfung der Hypothese wurden Quellengedächtnisstudien in diesem Sinne interpretiert (zusammenfassend siehe Spencer & Raz, 1995).

(4) Paradigma der direkten Testung von Kontextinformationen. Neben den bereits genannten Ansätzen haben Denney et al. (1991) ein Paradigma der direkten Testung von Kontextinformation und zentraler Information vorgeschlagen. Sie sehen drei voneinander unabhängige Gedächtnistests vor: Einen für die zentrale Information, einen für die Kontextinformation sowie einen dritten Test für die Verbindung beider Informationen. Diese Art des Gedächtnistests wurde vor allem entworfen, um methodische Probleme zu minimieren. Diese lagen unter anderem darin, dass unterschiedliche Gedächtnistests für zentrale Informationen und Kontextinformationen Verwendung fanden. So wurden häufig Rekognitionstests für die zentrale Information und geförderte oder freie Reproduktionstests für die Kontextinformation herangezogen. Da aber allein durch die unterschiedlichen Anforderungen dieser Testarten bereits Altersunterschiede hervorgerufen werden können

(Craik & Byrd, 1982), schlugen Denney et al. (1991) vor, möglichst identische Testarten für beide Elemente zu verwenden.

Für die Studien im Rahmen der Enkodierungsspezifitätshypothese stellt sich zudem das Problem, dass zwischen einem Defizit für Kontextinformationen und einem Defizit für die Verknüpfung von Kontext und zentraler Information nicht unterschieden werden kann. Sicherlich zeigen die Studien, dass ältere Menschen von der Wiederherstellung des Enkodierungskontexts beim Abruf weniger profitieren als jüngere Menschen (z.B. Park et al., 1984; Rabinowitz et al., 1982). Dieser empirische Befund *kann* durch eine verringerte Kontexterinnerung verursacht sein. Es könnte aber auch der Fall sein, dass ältere Menschen sich durchaus an den Lernkontext erinnern, da sie jedoch Kontext und zentrale Information nicht miteinander verbinden können, profitieren sie auch nicht von der Wiederherstellung des Lernkontexts. Schlussfolgerungen über die Ursachen des Defizits können anhand dieser Studien also nicht getroffen werden.

Darüber hinaus tritt vor allem in Quellengedächtnistests das Problem der Konfundierung von Item- und Quellengedächtnisleistung auf (siehe dazu Batchelder & Riefer, 1990; Bayen & Murnane, 1996; Bayen et al., 1996). Häufig wird die Quelleninformation nur in Abhängigkeit von der korrekten Erinnerung an die zentrale Information erfasst bzw. ausgewertet. Eine schlechte Erinnerungsleistung an Kontextinformationen kann bei einem solchen Vorgehen aber bereits aus einem allgemeinen Erinnerungsdefizit resultieren und muss nicht auf ein spezifisches Defizit für Kontextinformationen hinweisen. Als Ausweg kann daher ein experimentelles Vorgehen angewandt werden, in dem es drei getrennte Tests gibt: Einen Test für die Kontextinformation, einen für die zentrale Information und einen Test für die Verknüpfung der beiden Elemente (Denney et al., 1991). Alternativ können aber auch elaborierte Auswertungsmethoden wie z.B. multinomiale Modelle verwendet werden, um dieses Problem zu lösen (Batchelder & Riefer, 1990; Bayen & Murnane, 1996). Diese bieten zusätzlich die Möglichkeit, Antwortverzerrungen zu schätzen. Denn auch unterschiedliche Tendenzen zu liberalen oder konservativen Antworten in den verschiedenen Altersgruppen können die Ergebnisse oberflächlicher Auswertungen verzerren.

Nachdem nun die unterschiedlichen experimentellen Paradigmen zur Testung der Kontextdefizit-Hypothese vorgestellt wurden, soll auf die mit diesen Paradigmen erzielten Ergebnisse eingegangen werden. Die Ergebnisse stehen dabei allerdings zum größten Teil im Schatten der gerade erwähnten methodischen Probleme und werden daher auch eher überblicksartig geschildert. Lösungsansätze für die genannten Probleme werden anschließend aufgezeigt.

Empirische Befunde zur Kontextdefizit-Hypothese

Um ein zusammenfassendes Bild des Forschungsstandes zu geben, soll an dieser Stelle eine Meta-Analyse von Spencer und Raz (1995) vorgestellt werden. Spencer und Raz (1995) ha-

ben die Effektgrößen von Altersunterschieden in der Erinnerungsleistung zentraler vs. peripherer Merkmale über 46 Studien zusammengefasst. Als zentrale Informationen wurden verbale Materialien (Wörter oder einfache verbale Aussagen), visuelle Materialien (Objekte oder Strichzeichnungen) und Handlungen (Ausführen einfacher Handbewegungen oder Bearbeitung unterschiedlicher Tests) untersucht. Die Kontextinformationen bestanden aus perzeptuellen Merkmalen (Schriftfarbe, Schriftart, akustische Qualitäten), Modalitätsinformationen, räumlichen Positionsmerkmalen, zeitlichen Informationen und unterschiedlichen Personen als Quellen/Ausführende der zentralen Information. Die Autoren unterteilten die Studien weiter hinsichtlich des kognitiven Aufwands während Enkodierungs- und Abrufphase. Der Enkodierungsaufwand wurde hoch eingestuft bei intentionalen Lernbedingungen mit dem expliziten Hinweis darauf, welche Informationen später getestet werden. Ein niedriger Enkodierungsaufwand wurde angegeben, wenn die Studien inzidentelle Lernsituationen aufwiesen. Der Aufwand beim Abruf wurde als hoch angesetzt, wenn der Gedächtnistest geförderte oder freie Reproduktion verlangte, eine niedrige Einstufung erhielten Rekognitionstests.

Die Zusammenfassung der Studien zeigte zunächst eine Asymmetrie hinsichtlich der Erfassung der Gedächtnisleistung: Die Erinnerung an die Kontextinformation wurde in 58% der Studien mit geförderten oder freien Reproduktionstests erhoben, für die Erinnerungsleistung an zentrale Information verwendeten nur 42% der Studien dieses abrufintensive Testformat. Ein ähnliches Bild zeigte sich hinsichtlich des Enkodierungsaufwands. In der Mehrzahl der Studien war der Enkodierungsaufwand für zentrale Informationen hoch, für die Kontextinformation dahingegen niedrig. Dies spiegelt zunächst wider, dass zentrale Information und Kontextinformation selten unter vergleichbaren Bedingungen gelernt und getestet werden. Für die Meta-Analyse resultiert daraus eine verzerrte Gesamteffektstärke, da diese Konfundierungen nicht berücksichtigt wurden.

Die Analysen von Spencer und Raz (1995) beruhen auf den Effektstärken des Altersunterschieds. Eine positive Effektstärke steht für eine bessere Leistung der jüngeren Menschen, eine negative Effektstärke würde eine höhere Leistung der älteren Menschen signalisieren. Für die zentrale Information wurde eine Effektgröße von $d = 0.58$ errechnet (bzw. $d = 0.72$ unter Einbezug von Studien, die Ausreißerwerte aufwiesen). Die Effektgröße für die Kontextinformation betrug $d = 0.87$ (bzw. $d = 0.90$). Da sich die 95%-Konfidenzintervalle der Altersunterschiede für Kontextinformation und zentrale Information nicht überlappten, wurde dieser Unterschied als bedeutsam eingestuft. Die Ergebnisse der Meta-Analyse stehen also in Einklang mit der Kontextdefizit-Hypothese, da sie zeigen, dass für Kontextinformationen ein größerer Altersunterschied vorliegt als für zentrale Informationen. Hinsichtlich der Definition von Kontext und zentraler Information muss hier auf die pragmatische Definition von Chalfonte und Johnson (1996) zurückgegriffen werden. In Bezug auf die weiteren Klassifikationsmerkmale zeigte sich, dass der Altersunterschied zwischen Kontext- und zentraler Information in den Bedingungen mit niedrigem Enkodierungsaufwand und niedrigem Abrufaufwand jeweils größer ausfällt.

Allerdings weisen die in dieser Meta-Analyse zusammengefassten Studien die bereits diskutierten methodischen Probleme auf. Neben der Verwendung unterschiedlicher Testverfahren zu Erfassung der Gedächtnisleistungen liegt häufig eine Abhängigkeit der Erinnerung von Kontextinformationen an die Itemerinnerung vor. Die Autoren weisen daher selbst darauf hin, dass bei der Interpretation der Studien diese methodischen Vorbehalte beachtet werden müssen (Spencer & Raz, 1995, S. 534). Aufgrund dieser Kritikpunkte soll auch nicht genauer auf einzelne Studien eingegangen werden. Sinnvoller scheint es, sich den Lösungsansätzen für diese Probleme zuzuwenden.

Um die Kontextdefizit-Hypothese unabhängig von den methodischen Vorbehalten überprüfen zu können, haben verschiedene Autoren unterschiedliche experimentelle sowie modellbasierte Ansätze vorgeschlagen (z.B. Balota, Duchek & Paullin, 1989; Bayen & Murnane, 1996; Bayen, Phelps & Spaniol, 2000; G. D. Brown, Preece & Hulme, 2000; Chalfonte & Johnson, 1996; Howard & Kahana, 2002; Kliegl & Lindenberger, 1993; für einen Überblick siehe Spaniol & Bayen, 2004). Im Folgenden werden einige dieser Studien exemplarisch erläutert (Bayen & Murnane, 1996; Bayen, Phelps & Spaniol, 2000; Chalfonte & Johnson, 1996; Kliegl & Lindenberger, 1993).³ Anhand der Ergebnisse sollte es dann möglich sein, ein valides Fazit zur empirischen Überprüfung der Kontextdefizit-Hypothese zu ziehen.

Kliegl und Lindenberger (1993) verfolgten den Ansatz, durch individuell angepasste Lernzeiten den Altersunterschied auf Itemebene aufzuheben und damit zu vermeiden, dass ein allgemeines Erinnerungsdefizit die Interpretation der Ergebnisse erschwert. In ihrer Studie (Kliegl & Lindenberger, 1993) wurden die Probanden in der Mnemotechnik ‚Methode der Orte‘ trainiert (Bower, 1970; Kliegl et al., 1989). Sie mussten sich 30 Wörter merken, wobei verschiedene Berliner Stadtmerkmale als Erinnerungshinweise dienten. Die Lernzeiten wurden individuell so angepasst, dass ein bestimmtes Leistungskriterium von allen Probanden erreicht wurde. Der anschließende Test bestand aus einem geförderten Reproduktionstest, bei dem als Hinweisreize in zufälliger Reihenfolge die Berliner Stadtmerkmale vorgegeben wurden. Die zu lernenden Wörter mussten daraufhin reproduziert werden. Auf diese Art wurden vier verschiedene Listen sequentiell gelernt und getestet. Im Test sollten jeweils nur die Wörter der gerade präsentierten Liste erinnert werden. Als Maß für die Kontexterinnerung wurden Intrusionsfehler aus vorangegangenen Listen erfasst. Sobald ein Item also der falschen Liste d.h. dem falschen Kontext zugeteilt wurde, wurde dies als mangelnde Kontexterinnerung gewertet. Die Ergebnisse zeigten zunächst, dass ältere Menschen längere individuelle Lernzeiten als jüngere Menschen benötigten, um das Lernkriterium zu erreichen. Trotz der vergleichbaren Erinnerungsleistung für die zentrale Information traten bei den älteren Menschen signifikant mehr Intrusionsfehler aus vorangegangenen Listen auf, als bei den jüngeren. Da die Autoren zusätzlich ein formales

³Die Auswahl der Studien erfolgte im Hinblick auf die Relevanz für die vorliegende empirische Studie. Da in der vorliegenden Arbeit ein Rekognitionsparadigma verwendet wird, werden an dieser Stelle Auswertungsverfahren, Modelle und experimentelle Paradigmen vorgestellt, die sich auf Rekognitionsstudien anwenden lassen. Vergleichbare Analysemethoden für den freien Abruf werden nicht diskutiert.

Modell zur Datenauswertung verwendeten, konnte dieses Defizit älterer Menschen sogar noch weiter spezifiziert werden. Mit ihrem Zwei-Phasen-Markoff-Ketten-Modell (Kliegl & Lindenberger, 1993) wurden Parameter geschätzt, die zum einen für das Anlegen einer Gedächtnisspur mit dem Listenkontext als Hinweis stehen bzw. für die Wahrscheinlichkeit, dass der Listenkontext nicht erfolgreich enkodiert wird; zum anderen wird die Wahrscheinlichkeit der Aufrechterhaltung bzw. des Vergessens dieser Gedächtnisspuren geschätzt. Auf Ebene dieser Modellanalysen konnte ein Altersunterschied im Prozess der erfolgreichen Kontextenkodierung ausgemacht werden. Dies spricht dafür, dass während der Enkodierungsphase die Listeninformation nicht immer erfolgreich an die Iteminformation geknüpft wurde. Diese Studie konnte einen überzeugenden Weg aufzeigen, wie man das Problem der Abhängigkeit der beiden Gedächtnismaße lösen kann. Inwieweit diese Befunde die Kontextdefizit-Hypothese unterstützen, wird am Ende des nächsten Absatzes ausgeführt.

Um in Quellengedächtnisaufgaben die Abhängigkeit von Item (zentraler Information) und Quelle aufzuheben, schlugen Batchelder und Riefer (1990) die Anwendung von multinomialen Modellen zur Datenauswertung vor. Diese Methode ermöglicht es, unabhängige Parameter für Itemerkennung, Quellenidentifikation sowie für Rateprozesse zu schätzen. Bayen und Murnane (1996) weisen darauf hin, dass in Altersstudien die Konfundierung unterschiedlicher Gedächtnismaße dazu führen kann, dass Altersunterschiede im Quellengedächtnis beobachtet werden, die aber eigentlich auf Altersunterschiede in der Itemerinnerung zurückgehen. Um dieses Problem zu beheben, haben Batchelder und Riefer (1990) eine Klasse von multinomialen Modellen vorgestellt, mit deren Hilfe die Daten von Quellengedächtnisexperimenten analysiert werden können. Bayen et al. (1996) haben die vorgeschlagenen Quellengedächtnismodelle von Batchelder und Riefer (1990) sowie weitere Modellvarianten einer empirischen Evaluation unterzogen. Dabei konnte die Modellvariante, die von einer Zwei-Hochschwellen-Annahme⁴ ausgeht (Macmillan & Creelman, 1990), ihre Überlegenheit gegenüber anderen Modellvarianten unter Beweis stellen (zur genaueren Diskussion siehe Bayen et al., 1996). Bayen und Murnane (1996) haben dieses Modell verwendet, um die Daten älterer Probanden zu analysieren und damit u.a. der Frage nachzugehen, ob sich ein Altersunterschied sowohl in der Itemerkennung als auch in der Quellengedächtnisleistung finden würde. Die Ergebnisse zeigten, dass ältere Menschen neben der nachlassenden Leistung in der Itemerkennung auch schlechtere Leistungen in der Identifikation der Quellen aufwiesen.

Dieses Ergebnis kann aber streng genommen (genau wie das Ergebnis der Studie von Kliegl und Lindenberger (1993)) nicht als unterstützender Befund für die Kontextdefizit-Hypothese angesehen werden, da in Quellengedächtnistests kein reiner Test für die

⁴Zwei-Hochschwellen-Modelle gehen von einer Schwelle für die Erinnerung an alte Items und einer Schwelle für die bewusste Entdeckung von Distraktoren aus. Die Erinnerungsschwelle für alte Items wird dann überschritten, wenn ein altes Item als sehr vertraut eingestuft wird. Es wird angenommen, dass dieser Zustand nur für alte Items erreicht werden kann. Umgekehrt können nur neue Items einen Zustand von hoher Unvertrautheit hervorrufen und damit die Schwelle zu einem „Neu“-Urteil erreichen.

Quelleninformation stattfindet. Wenn z.B. verschiedene Sätze von drei unterschiedlichen Personen dargeboten werden, so wird im Quellengedächtnistest zunächst gefragt, ob eine bestimmte Aussage in der Lernphase gezeigt wurde oder ob es sich um eine Distraktorinformation handelt. Im ersten Fall wird dann nach der Quelle dieser Information, Person A, B oder C gefragt. Der Quellengedächtnistest bezieht sich immer auf eine bestimmte Aussage und deren Verbindung zu einer Quelle. Strenggenommen wird also getestet, inwieweit die Aussage mit einer der Quellen *verknüpft* wurde. Wäre man nur daran interessiert, ob Wissen über die Kontextinformation vorliegt, müsste man einen weiteren Test hinzunehmen, in dem Distraktorquellen von tatsächlich gelernten Quellen unterschieden werden müssen. Ein solcher Test wäre ein reiner Test für Kontextinformationen.⁵

An dieser Stelle ist es wichtig festzuhalten, dass sowohl die Studie von Kliegl und Lindenberger (1993) als auch die Studie von Bayen und Murnane (1996) Möglichkeiten aufweisen, wie die Abhängigkeit der Maße des Item- und Kontext-Gedächtnisses aufgehoben werden kann. Keine der Studien unterstützt aber notwendigerweise die Hypothese, dass ältere Menschen Kontextinformationen weniger gut erinnern können. Sie zeigen auf, dass ältere Menschen weniger gut erinnern können, welche Aussagen von welchen Personen dargeboten wurden bzw. welche Items in welcher Liste gelernt wurden. Dies weist aber eher auf ein Defizit für die Assoziation der Elementen hin (siehe Naveh-Benjamin, 2000) als auf ein reines Kontextdefizit.

Als Beispiel für Untersuchungen, die ebenfalls die Konfundierung von zentraler Information und Kontextinformation aufheben, zusätzlich aber einen Test für die reine Kontextinformation enthalten, können zum einen die Studie von Bayen, Phelps und Spaniol (2000) sowie die Studie von Chalfonte und Johnson (1996) genannt werden. Beide sollen im Folgenden kurz vorgestellt und ihre Ergebnisse im Hinblick auf die Gültigkeit der Kontextdefizit-Hypothese beurteilt werden.

Bayen, Phelps und Spaniol (2000) vertreten die Hypothese, dass ältere Menschen kein allgemeines Defizit in der Verarbeitung von Kontextinformationen haben, sondern dass ältere Menschen Kontextinformationen und zentrale Informationen schlechter integrieren können. Um diese Hypothese zu testen, haben sie ein globales Gedächtnismodell, das ICE-Modell (‘item-context-ensemble-model’; Murnane, Phelps & Malmberg, 1999) angewandt. Für einen Überblick zu globalen Gedächtnismodellen sei hier auf Clark und Gronlund (1996) und für eine detaillierte Einführung in das ICE-Modell auf Murnane et al. (1999) verwiesen. Kurz zusammengefasst unterscheidet das ICE-Modell Item-Information d.h. die zentrale Information (I), und zwei Arten von Kontextinformationen. Zum einen die rei-

⁵Da bei Quellengedächtnistests meist nur sehr wenige Quellen verwendet werden, kann allerdings davon ausgegangen werden, dass die Teilnehmer diese Quellen nach der Lernphase sehr gut gelernt haben und eine gute Erinnerungsleistung an den Quellenkontext vorliegt. Trotzdem ist diese konzeptuelle Trennung notwendig, um zwischen einem Kontextdefizit und einem Defizit der Verbindung von Kontext und zentraler Information zu trennen.

ne Information des Kontexts (C) und die Integration, d.h. die Verbindung von zentraler Information und Kontext, in diesem Modell Ensemble (E) genannt. Jede dieser drei Informationsquellen kann zur globalen Aktivierung beitragen. In Abhängigkeit davon, welche Informationen aktiviert werden, resultieren aus dem Modell unterschiedliche Vorhersagen für die Trefferraten und die Raten falscher Alarme sowie das daraus berechnete Diskriminationsmaß (d'). Das Modell wurde in der Studie von Bayen, Phelps und Spaniol (2000) auf Kontexteffekte im Alter angewandt. Dabei wurden zwei unterschiedliche Testsituationen erzeugt. Zentrale Informationen wurden entweder in dem Kontext getestet, in dem sie auch gelernt wurden, oder aber in einem neuen, unbekanntem Kontext (siehe Enkodierungsspezifitätshypothese; Kapitel 3.1). Das Ausmaß dieses Kontexteffekts wurde durch die Differenz zwischen Gedächtnismaßen bei übereinstimmendem Kontext und nicht übereinstimmendem Kontext berechnet. Zunächst werden nun die allgemeinen Modellvorhersagen erläutert, dann wird auf die spezifischen Altershypothesen eingegangen.

Folgende Vorhersagen ergeben sich, falls für den Gedächtnistest Kontextaktivierung und Itemaktivierung, aber keine Ensembleaktivierung zur Verfügung stehen:

- (1) Die globale Aktivierung ist bei gelernten Items, die mit ihrem ursprünglichen Kontext dargeboten werden, höher als bei gelernten Items, die in einem neuen Kontext getestet werden. Unter Bedingungen der Kontextübereinstimmung sollten also höhere Trefferraten resultieren.
- (2) Da auch ein bekannter Kontext eine Aktivierung hervorruft, sollten die Raten falscher Alarme für Distraktoren, die vor einem gelernten Kontext getestet werden, erhöht sein gegenüber der Bedingung, in der Distraktoren vor einem neuen Kontext getestet werden.
- (3) Eine Übereinstimmung des Kontexts zwischen Lern- und Testbedingung führt sowohl zu höheren Trefferraten als auch zu höheren Raten falscher Alarme. Beide Gedächtnismaße werden durch die Kontextaktivierung beeinflusst und sollten sich daher in ähnlichem Maße verändern. Daraus lässt sich folgern, dass sich für den Diskriminationsindex (d') keine Veränderung zwischen den beiden Kontextbedingungen ergeben sollte.

Falls jedoch zusätzlich zu der Kontext- und Itemaktivierung auch Ensembleaktivierung während des Gedächtnistests genutzt werden kann, verändern sich die Vorhersagen wie folgt:

- (1a) Analog zu der ersten Vorhersage resultieren bei gelernten Items unter der Bedingung von Kontextübereinstimmung höhere Trefferraten. Da nun aber zusätzlich die Aktivierung der Ensembleinformation hinzukommt, werden die Trefferraten stärker ansteigen, als dies alleine durch die Kontextübereinstimmung der Fall ist. Die Aktivierung von Kontextinformation *und* Ensembleinformation tragen zur Aktivierung bei.
- (2a) Für Distraktoren kann in der Bedingung der Kontextübereinstimmung eine Aktivierung nur durch den Kontext erfolgen, da für neue Items keine Ensembleinformation verfügbar ist. Es folgt also ein Anstieg in den Raten falscher Alarme für Distraktoren, wobei dieser Anstieg geringer ausfallen sollte, als der Anstieg in den Trefferraten (1a).

(3a) Da der Anstieg in den Trefferraten höher ausfallen sollte als der Anstieg in den Raten falscher Alarme (2a,b), ist ein Effekt auf den Diskriminationsindex zu erwarten.

Für den von Bayen und Kollegen postulierten Fall, dass ältere Menschen ein Defizit in der Integration der Informationen aufweisen, sollte für ältere Menschen *kein* Effekt im Diskriminationsmaß resultieren, bei jüngeren Menschen dahingegen sollte dieser Effekt vorhanden sein. Dieses Ergebnismuster würde im Widerspruch zur Kontextdefizit-Hypothese stehen. In dem experimentellen Paradigma von Bayen und Kollegen wurden Wortpaare vor einem komplexen Hintergrund wie z.B. einer Wohnzimmerszene dargeboten. Insgesamt wurden vier unterschiedliche, jeweils komplexe Hintergrundscenarien verwendet. Die Aufgabe war eine inzidentelle Lernaufgabe, bei der die Personen auf einer 5-Punkte-Skala angeben sollten, wie stark die zwei Wörter miteinander in Verbindung stehen. Nach einem Retentionsintervall wurden 32 Zielwörter und 32 Distraktoren entweder vor einem neuen Hintergrund oder vor demselben Hintergrund getestet.

Die Ergebnisse stimmten mit den Hypothesen der Autoren überein. Bei jüngeren Menschen fand sich im Vergleich der beiden Kontextbedingungen ein Effekt auf das Diskriminationsmaß. Dies deutet also darauf hin, dass zusätzlich zu den Item- und Kontextinformationen auch die Integration als Ensemble erfolgte. Bei älteren Menschen fand sich kein Effekt auf den Diskriminationsindex, es ist also nach den Modellannahmen davon auszugehen, dass ältere Menschen nicht in der Lage waren, Ensembleinformationen zu bilden und/oder abzurufen.

Diese Studie zeigte, dass ältere Menschen kein allgemeines Defizit in der Verarbeitung von Kontextinformationen haben, sondern dass sie ein spezifisches Defizit in der Integration von Kontext und Iteminformationen aufweisen. Diese Ergebnisse stehen auch in Einklang mit den bereits geschilderten Studien zum Quellengedächtnis (Bayen et al., 1996). Kritisch könnte man anmerken, dass die Anforderungen an die Gedächtnisleistung für die Kontextinformationen nicht besonders hoch waren, da nur vier unterschiedliche Hintergrundbilder verwendet wurden. Zudem unterschieden sich die als Distraktoren verwendeten Hintergrundbilder sehr stark und in vielen Merkmalen von den in der Lernphase gezeigten Bildern. Das Rekognitionsgedächtnis für diese geringe Menge an Kontextinformation müsste also auch bei älteren Menschen gut ausgeprägt sein. Es sollte daher in weiteren Studien getestet werden, ob sich die Ergebnisse von Bayen, Phelps und Spaniol (2000) auch replizieren lassen, wenn die Anforderungen für zentrale und periphere Informationen vergleichbar sind. Andererseits muss betont werden, dass auch in vielen Studien, die die Kontextdefizit-Hypothese bestätigten, nur eine geringe Anzahl von Kontextinformationen erinnert werden mussten (für einen Überblick siehe Spencer & Raz, 1995).

Eine vergleichbare Komplexität von zentraler Information und Kontextinformation lag z.B. der Studien von Chalfonte und Johnson (1996) zugrunde. Das Ziel ihrer Experimente entsprach der Zielsetzung von Bayen und Kollegen. Auch sie wollten trennen, ob ältere Menschen (a) eine schlechtere Erinnerung an die Einzelelemente einer Episode aufwei-

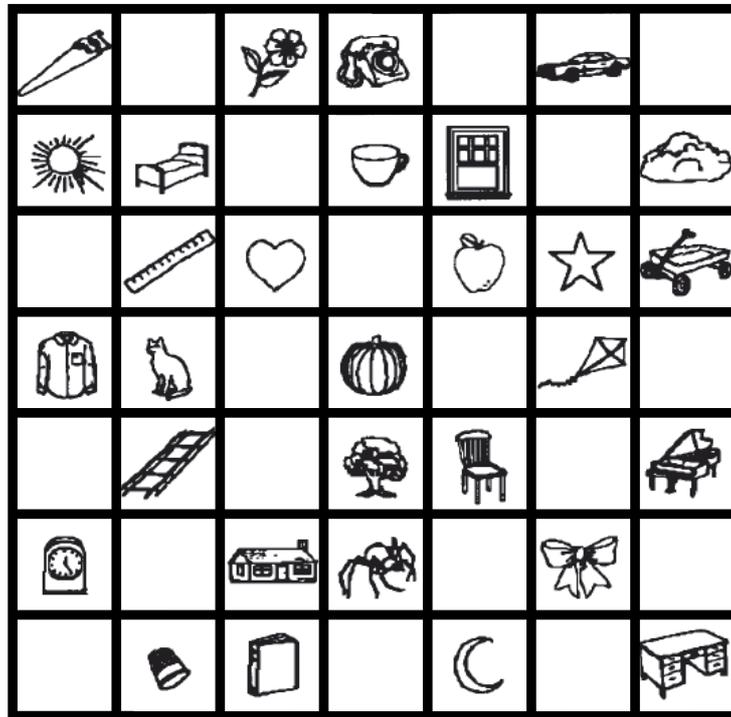


Abbildung 3.1: In dieser Abbildung ist exemplarisch das Lernmaterial aus der Studie von Chalfonte und Johnson (1996) abgebildet. Die Teilnehmer sahen ein 7*7 Raster, in dem 30 farbige Objekte (hier in schwarz-weiß) zu sehen waren. Abhängig von der Lerninstruktion sollten sich die Teilnehmer entweder nur die Objekte/Positionen/Farben merken oder Kombination von Objekt und Farbe bzw. Objekt und Position. [Abbildung aus Chalfonte und Johnson (1996), S. 405]

sen, (b) weniger gut in der Lage sind, zentrale und periphere Elemente miteinander zu integrieren, oder (c) ob beides der Fall ist.

Die zentrale Information in ihren Experimenten waren Strichzeichnungen von gewöhnlichen Objekten (Snodgrass & Vanderwart, 1980) wobei jedes Objekt in einer anderen Farbe dargeboten wurde. Dreißig dieser Zeichnungen wurden zufällig in einem Gitternetz mit 49 Feldern angeordnet. Als Kontextelemente dienten zum einen die Positionen der Zeichnungen sowie auch die Farben der Objekte (für ein Beispiel siehe Abbildung 3.1).

Die Teilnehmer erhielten während der Lernphase unterschiedliche Instruktionen: Entweder sollten sie sich nur eine Informationsart einprägen (Objekte, Orte oder Farben) oder sie sollten sich die Kombination zweier Informationen merken (Objekte und Farben oder Objekte und Positionen). In der Testphase wurden (unabhängig von der Instruktion) zum einen die Einzelinformationen alleine getestet, zum anderen die Kombinationen. Die Reihenfolge der Tests wurde über die Teilnehmer hinweg ausbalanciert. Für den Objekttest wurden dabei 10 alte und 10 neue Objekte in Schwarz-Weiß dargeboten. Die Aufgabe der Probanden bestand darin, die alten Objekte zu markieren. Für den Positionstest wur-

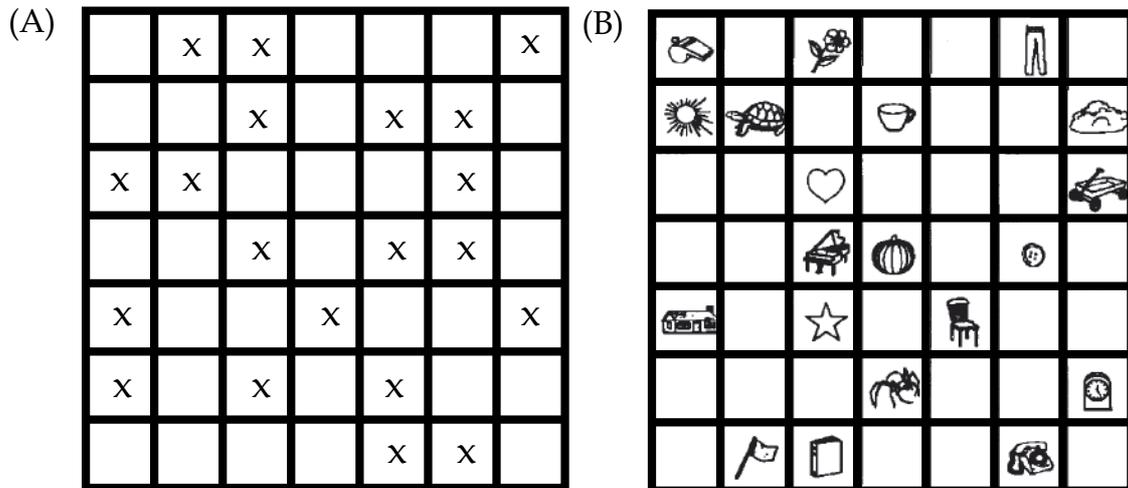


Abbildung 3.2: Abbildung (A) zeigt den Positionstest aus dem Experiment von Chalfonte und Johnson (1996). In dem 7*7 Raster wurden 20 Positionen über Kreuze markiert. 10 Kreuze markierten dabei Positionen, an denen während der Lernphase Objekte zu sehen waren, 10 Kreuze markierten Positionen, die während der Lernphase unbesetzt blieben. Abbildung (B): Für den Kombinationstest von Objekten und Positionen wurden 10 alte Objekte an alten Positionen und 5 alte Objekte an neuen Positionen dargeboten. Weitere 5 Distraktorobjekte wurden an alten Positionen gezeigt. [Abbildung aus Chalfonte und Johnson (1996), S. 406]

den in der Matrix der 49 Felder 10 alte und 10 neue Positionen mit schwarzen Kreuzen markiert. Die Positionen, an denen vorher Objekte zu sehen waren, sollten analog zum Objekttest gekennzeichnet werden (siehe Abbildung 3.2 A). Für den Farbttest schließlich wurden 10 alte und 10 neue farbige Felder dargeboten, wiederum mit der Aufgabe, die Farben hervorzuheben, in denen vorher Objekte gezeichnet waren. Der Kombinationstest für Objekt-Position-Verknüpfungen bestand aus 10 alten schwarz-weiß Objekten, die in ihrer ursprünglichen Position angezeigt wurden. Fünf Distraktoren wurden an Positionen dargeboten, an denen vorher alte Items gezeigt wurden, und fünf alte Objekte wurden an neuen Positionen präsentiert (siehe Abbildung 3.2 B). Der Objekt-Farb-Verbindungstest war analog dazu aufgebaut.

Positiv können an diesem experimentellen Paradigma zwei Punkte hervorgehoben werden: Zum einen liegen getrennte und unabhängige Gedächtnistests für die einzelnen Elemente und die Verknüpfung von Elementen vor. Zum anderen wurde die Komplexität für die zentralen Informationen und die Kontextinformationen vergleichbar gehalten.

Die Ergebnisse zeigten keinen Altersunterschied in der Objektrekognition und keinen Altersunterschied in der Rekognition der Farbe. Allerdings waren ältere Menschen signifikant schlechter darin, alte Positionen als solche wiederzuerkennen. Die Autoren zogen die Schlussfolgerung, dass ältere Menschen kein generelles Defizit für Kontextinformationen haben, wohl aber örtliche Informationen schlechter verarbeiten oder abrufen können.

Die Assoziationstests wiesen beide einen signifikanten Altersunterschied auf. Obwohl also kein Altersunterschied in der Rekognition von Objekten bzw. Farben vorhanden war, waren ältere Menschen schlechter in der Lage, sich an die Assoziation der Elemente zu erinnern. In zwei Folgeexperimenten wurde dieser Befund repliziert. Damit konnte sowohl von Chalfonte und Johnson (1996) als auch von Bayen, Phelps und Spaniol (2000) gezeigt werden, dass ältere Menschen kein globales Defizit für Kontextinformationen aufweisen, sondern ein spezifisches Defizit für bestimmte Kontextinformationen (wie z.B. die räumliche Position) sowie ein spezifisches Defizit für die Assoziation von Kontext und zentraler Information.

Die bereits angeführten Befunde sowie weitere Studien (Naveh-Benjamin, 2000; Old & Naveh-Benjamin, 2008a) sprechen also gegen die Hypothese eines reinen Kontextdefizits und führten zu der spezifischeren Hypothese eines Assoziationsdefizits. Auf diese Hypothese soll im nächsten Kapitel eingegangen werden.

3.2 Die Assoziationsdefizit-Hypothese

Die im vorangegangenen Kapitel geschilderten empirischen Befunde zur Kontextdefizit-Hypothese konnten zeigen, dass ältere Menschen kein globales Gedächtnisdefizit für Kontextinformationen aufweisen. Vielmehr wurde erste Evidenz für ein spezifisches Defizit in der Bildung von Assoziationen z.B. zwischen Kontext und zentraler Information berichtet. So schlussfolgern Chalfonte und Johnson (1996, S. 415): „older adults demonstrated difficulties in binding features together“. Auch Bayen, Phelps und Spaniol (2000) ziehen den Schluss, dass ältere Menschen Schwierigkeiten haben, zentrale Informationen und Kontextinformationen zu einem Ensemble zu integrieren.

In diesem Kapitel soll eine genaue Erläuterung sowie Definition der Assoziationsdefizit-Hypothese im Mittelpunkt stehen. Zeitlich wurde die Definition parallel zu den beiden oben aufgeführten Studien im Jahr 2000 formuliert und auch die empirischen Befunde stammen aus der jüngeren Forschung. Dies bedeutet jedoch nicht, dass frühere Studien sich nicht mit diesem Thema beschäftigt haben oder dass Forschungsergebnisse vor 2000 nicht rückblickend im Licht der Assoziationsdefizit-Hypothese betrachtet werden können. Deshalb wird im Folgenden kurz eine Studie angesprochen, die sich bereits 1941 das assoziativen Defizit untersuchte und es wird knapp erläutert, unter welchem Fokus assoziative Gedächtnisprozesse danach in der Forschung weiterverfolgt wurden.

Die m.W. erste experimentelle Untersuchung zu dem assoziativen Defizit älterer Menschen wurde von Gilbert (1941) durchgeführt: Gilbert untersuchte 174 ältere Menschen im Alter von 60 bis 69 Jahren und zog als Vergleichsgruppe 174 jüngere Menschen zwischen 20 und 29 Jahren heran. Besonders positiv ist an dieser frühen experimentellen Untersuchung hervorzuheben, dass bei der Stichprobenezusammenstellung darauf geachtet wurde,

dass sich die Individuen beider Altersgruppen jeweils paarweise in ihrer verbalen Fähigkeit glichen. Insgesamt wurden Altersvergleiche für elf unterschiedliche Tests durchgeführt. Die Tests umfassten dabei Kurzzeitgedächtnistests für unterschiedliche Modalitäten sowie verschiedene Arten episodischer Langzeitgedächtnistests. Drei der Langzeitgedächtnistests wurden als besonders sensibel für die Bildung bzw. den Abruf von Assoziationen eingestuft. Dies waren zum einen ein Paar-Assoziations-Test mit sofortiger bzw. verzögerter Wiedergabe und ein Test für Englisch-Türkische Vokabeln. Gilbert (1941) konnte in allen elf Tests reliable Altersunterschiede feststellen. Ein Altersvergleich zwischen den Tests zeigte, dass die prozentuale Abnahme der Leistung älterer Menschen in den Kurzzeitgedächtnistests am geringsten war (Spannweite von 8.5% bis 21.2%) und in den drei assoziationssensitiven Tests am stärksten ausgeprägt war (Spannweite von 54.6% bis 60.4%). Auch in der Teilstichprobe der Personen, die sehr gute verbale Fähigkeiten aufwiesen - 40 ältere und jüngere Menschen - blieb diese Rangfolge der Altersunterschiede über die unterschiedlichen Tests bestehen. Gilbert schlussfolgerte aus diesen Daten, dass ältere Menschen besondere Probleme haben, neue Assoziationen zu bilden und abzurufen. Er spricht in dieser Arbeit bereits die Hypothese eines assoziativen Defizits bei älteren Menschen an. Auch aus heutiger Sicht können die Daten von Gilbert (1941) als Evidenz für ein assoziatives Defizit interpretiert werden.

Obwohl sich empirische Evidenz und erste theoretische Hypothesen in Bezug auf ein Assoziationsdefizit also bereits in den Anfängen der Alternsforschung (Gilbert, 1941) finden, ist die größte Publikationsdichte zu diesem Thema erst in den letzten Jahren zu verzeichnen. Dies liegt daran, dass in der Zwischenzeit assoziative Gedächtnisprozesse unter einem anderen theoretischen Fokus betrachtet wurden.

In den 70er und 80er Jahren wurden assoziative Gedächtnisprozesse in Studien zum Paar-Assoziationslernen untersucht (z.B. Craik & McDowd, 1987). Hier wurde vor allem die Hypothese überprüft, dass ältere Menschen Probleme in der Initiierung von Abrufprozessen aufweisen und daher die Altersunterschiede in Rekognitionsaufgaben geringer ausfallen, als in geförderten oder freien Reproduktionstests (Craik & Byrd, 1982, siehe Kapitel 3.2.4). Diese Studien konzentrierten sich auf den Vergleich von unterschiedlichen Testverfahren. Es wurden durchaus assoziative Aufgaben verwendet, aber unter einer anderen Perspektive.

Gleiches gilt für die zahlreichen Studien, die im Rahmen der Kontextdefizit-Hypothese seit den frühen 80er Jahren durchgeführt wurden (s. Kap. 3.1). Auch hier spielen assoziative Gedächtnisprozesse eine große Rolle, aber die experimentellen Ansätze und Auswertungsmethoden lassen in den meisten Fällen keine validen Schlüsse im Hinblick auf ein Assoziationsdefizit zu. Erst nachdem die Kontextdefizit-Hypothese empirisch angezweifelt wurde (z.B. Bayen, Phelps & Spaniol, 2000; Chalfonte & Johnson, 1996), kam das Konzept des assoziativen Defizits (Gilbert, 1941) wieder auf und wurde mit *angemessenen* experimentellen Ansätzen und Auswertungsmethoden untersucht.

In die Literatur eingeführt wurde der Begriff ‚Assoziationsdefizit‘ von Naveh-Benjamin (2000) in seinem Artikel „Adult Age Differences in Memory: Tests of an Associative Deficit Hypothesis“. In den folgenden Kapiteln wird zunächst die Definition des Assoziationsdefizits von Naveh-Benjamin (2000) ausgeführt und diskutiert (Kap. 3.2.1). In Kapitel 3.2.2 wird in einem Überblick über empirische Forschungsarbeiten gezeigt, dass es sich bei diesem Defizit um ein stabiles und empirisch gefestigtes Altersdefizit handelt. In den Kapiteln 3.2.3 bis 3.2.5 wird das Assoziationsdefizit aus dem Blickwinkel der theoretischen Perspektiven von Salthouse betrachtet. Dabei wird zunächst begründet, warum dem Defizit die Stellung eines veränderten Basismechanismus zukommt und es nicht alleine als qualitative Altersveränderung anzusehen ist (Kap. 3.2.3). Daran anschließend soll die Interaktion und Abgrenzung zu anderen Basismechanismen thematisiert werden, um die unique Rolle des assoziativen Defizits zu rechtfertigen (Kap. 3.2.4). Zuletzt soll der Basismechanismus aus der analytischen Lokalisationsperspektive betrachtet werden. Dabei wird die Frage gestellt, ob das Defizit auf Enkodierungs- und/oder Abrufprozesse zurückzuführen ist.

3.2.1 Die Assoziationsdefizit-Hypothese: Eine Definition

Naveh-Benjamin (Naveh-Benjamin, 2000, S. 1170) definiert die Assoziationsdefizit-Hypothese wie folgt: „This hypothesis claimed that an important component of older adults’ poorer episodic memory is their deficiency in creating and retrieving links between single units of information. The basic units can be two items; an item and its context; two contextual elements; or, more generally, the representation of two mental codes. The extent to which a given memory task requires the creation or use of such associations is a significant determinant of older people’s memory performance“. Diese Definition wird nun genauer erläutert, weiter spezifiziert und diskutiert.

„This hypothesis claimed that an important component of older adults’ poorer episodic memory [...]“ Die Assoziationsdefizit-Hypothese soll also Altersunterschiede in episodischen Gedächtnisaufgaben erklären. Dies ist eine wichtige Spezifizierung, da Assoziationsprozesse keineswegs nur auf der Ebene des Langzeitgedächtnisses, sondern auch auf Ebene des Arbeitsgedächtnisses und bei Wahrnehmungsprozessen eine wichtige Rolle spielen. Zimmer, Mecklinger und Lindenberger (2006) geben einen Überblick über unterschiedliche Bindungsarten.⁶ Sie unterscheiden dabei zwischen frühen Bindungsprozessen auf Ebene der Wahrnehmung und zwischen späten Bindungsprozessen auf Ebene des Kurzzeit- oder Langzeitgedächtnisses.

Frühe Bindungsprozesse auf Wahrnehmungsebene (Kahneman, Treisman & Gibbs, 1992; Treisman, 1992, 2006; Treisman & Gelade, 1980) ermöglichen es uns, mehrere visuelle Teilelemente zu integrieren und das Ergebnis als ein zusammengehöriges Objekt wahrzunehmen. Treisman (Kahneman et al., 1992; Treisman, 1992) bezeichnet das Resultat dieser

⁶Zimmer et al. (2006) verwenden die Begriffe Bindung und Assoziation äquivalent.

frühen Bindungsprozesse als ‚Objekt-Token‘. In den bereits vorgestellten Aufgaben von Chalfonte und Johnson (1996) wäre die Wahrnehmung einer einfachen Strichzeichnung als einheitliches Objekt (z.B. ein Haus) das Ergebnis eines solchen frühen Bindungsprozesses und die temporäre Repräsentation dieser Zeichnung würde als Objekt-Token bezeichnet werden.

Objekt-Token können dann auf Ebene des Kurzzeitgedächtnisses und des Langzeitgedächtnisses weiterverarbeitet werden. Daher werden Objekt-Token als Basiselemente von nachfolgenden Verarbeitungsprozessen bezeichnet (Zimmer et al., 2006). Von Bindungen auf höherer Ebene (oder späten Bindungsprozessen) wird gesprochen, wenn mehrere Objekt-Token miteinander verknüpft werden. Prozesse der Bindung zwischen dem Objekt und dessen Farbe wären also höhere Bindungsprozesse, genau wie die Bindung von zwei Wörtern zu einem Wortpaar. Jegliche Bindung zwischen mindestens zwei (jeweils als Einheit wahrgenommenen) Elementen wird als später Bindungsprozess bezeichnet (Cabeza, 2006; Zimmer et al., 2006). In Anlehnung an den Begriff Objekt-Token bezeichnen Ecker, Zimmer und Groh-Bordin (2007) das Ergebnis später Bindungsprozesse auch als ‚Episodic-Token‘.

Neben dieser Unterteilung in frühe und späte Bindungsprozesse hat Cabeza (2006) die Begriffe ‚Item-Bindung‘ und ‚relationale Bindung‘ sowie die entsprechenden Termini des ‚Item-Gedächtnisses‘ und des ‚relationalen Gedächtnisses‘ eingeführt. Unter Item-Bindung werden dabei die frühen Bindungsprozesse zusammengefasst, die es uns ermöglichen, ein Objekt oder Wort als eine Einheit - ein Item - wahrzunehmen. Unter relationaler Bindung versteht er die Assoziationsprozesse, die Bindungen zwischen wahrgenommenen Elementen ermöglichen. Cabeza (2006) unterteilt die relationalen Bindungsprozesse noch weiter in z.B. temporäre oder semantische Bindungsprozesse. Diese Unterscheidung wird in Kapitel 3.2.2 aufgegriffen und differentielle Alterseffekte für die unterschiedlichen Bindungsarten werden diskutiert. Die Assoziationsdefizit-Hypothese wird also für höhere, Ressourcen erfordernde, relationale Bindungsprozesse auf Ebene des episodischen Gedächtnisses formuliert. Interessante Erweiterungen der Hypothese könnten sich durch eine Ausweitung auf andere Assoziationsprozesse ergeben (siehe z.B. Brockmole, Parra, Della Sala & Logie, 2008; Mitchell et al., 2000).

„The basic units can be two items; an item and its context; two contextual elements; or, more generally, the representation of two mental codes.“ In diesem Satz der Definition wird die kleinste Einheit, die in einen Bindungsprozess auf höherer Ebene eingehen kann, beschrieben. Naveh-Benjamin spricht dabei von der Repräsentation einer mentalen Kodierung, Zimmer und Kollegen (2006) spezifizieren die kleinste Einheit als ‚ein als Einheit wahrgenommenes Element‘. Wichtig ist hier das Wort ‚wahrgenommen‘, denn bei einer Person, die nicht lesen kann, oder einem Europäer, dem chinesische Schriftzeichen dargeboten werden, wird ein zu lernendes Wort nicht als eine Einheit re-

präsentiert. Die kleinsten wahrgenommenen Einheiten werden in diesem Fall maximal Buchstaben bzw. Schriftzeichen sein. Umgekehrt kann es natürlich auch gelingen, durch geschickte Verkettungen („chunking“) eine zunächst nicht zusammengehörige Buchstabenfolge von sechs Elementen (F, B, I, C, I, A) in größere Einheiten zusammenzufassen (FBI, CIA) und damit die Anzahl der Elemente zu reduzieren (G. A. Miller, 1956). Auch zwei Wörter, die zu einem zusammengesetzten Wort integriert werden können (z.B. Fuß und Ball), erfordern keine relationalen Bindungsprozesse, sondern können als ein Wort, d.h. eine Einheit verarbeitet werden (siehe dazu den Abschnitt „Integrationspotential“ in Kapitel 3.2.2).

Ob ein Objekt als Einheit wahrgenommen wird oder nicht, hängt aber nicht nur von dem Wissen und den Fähigkeiten der einzelnen Person ab, sondern auch von spezifischen Anforderungen und Charakteristika der Aufgabe. So werden normalerweise Gesichter, die aus Einzelementen wie Augenpartie, Mund, Nase, Kinn etc. bestehen, als Einheit wahrgenommen (als Überblick siehe Searcy & Bartlett, 1996). Dies gilt nicht mehr, wenn Gesichter um 180 Grad gedreht präsentiert werden (Searcy & Bartlett, 1996; Yonelinas, Kroll, Dobbins & Soltani, 1999). Experimentelle Studien konnten zeigen, dass unter diesen ungewohnten Wahrnehmungsbedingungen das Gesicht nicht mehr die kleinste wahrgenommene Einheit darstellt. Um die kleinste wahrgenommene Einheit zu bestimmen, müssen also immer eine Reihe von Rahmenbedingungen wie die Besonderheiten der Aufgabe oder Fähigkeiten und Kenntnisse der Person mit einbezogen werden.

„[...] creating and retrieving links between single units of information [...]“

Eine nächste wichtige Spezifizierung wird dadurch getroffen, dass das Assoziationsdefizit als Defizit in der Bildung von Assoziationen und im Abruf von Assoziationen angesehen wird. Das Assoziationsdefizit älterer Menschen stellt also ein übergeordnetes Defizit dar, dass in mindestens zwei Subprozesse unterteilt werden kann. In Anlehnung an Cohn, Emrich und Moscovitch (2008) soll der Prozess der Bildung bzw. Enkodierung von Assoziationen als Bindungsprozess, das Defizit während dieses Prozesses als Bindungsdefizit sowie *assoziatives Enkodierungsdefizit* bezeichnet werden. Der zweite Subprozess soll Abrufprozess bzw. *assoziatives Abrufdefizit* genannt werden. Andere Autoren nennen die Subprozesse auch Bindungsdefizit der Enkodierung und Bindungsdefizit des Abrufs (z.B. Zimmer et al., 2006). Mit dieser Spezifizierung wird in Anlehnung an die theoretischen Perspektiven von Salthouse (1991) das Assoziationsdefizit aus der analytischen Lokalisationsperspektive betrachtet. In Kapitel 3.2.5 wird genauer vorgestellt, wie versucht wurde, ein assoziatives Enkodierungs- oder Abrufdefizit aufzuzeigen und welche Schlussfolgerungen daraus gezogen werden können.

„The extent to which a given memory task requires the creation or use of such associations is a significant determinant of older people’s memory performance.“ Nachdem in den vorangegangenen Abschnitten der Definition das Asso-

ziationsdefizit spezifiziert wurde, wird in diesem Satz beschrieben, durch welche Faktoren das Ausmaß des Altersunterschieds in Gedächtnisaufgaben vorhergesagt werden kann. Verschiedene Gedächtnisaufgaben unterscheiden sich darin, inwieweit sie auf assoziativen Prozessen beruhen und abhängig davon sollten die Altersunterschiede eher schwach oder stark ausgeprägt sein. Einfache Itemrekognitionsaufgaben z.B. involvieren im Gegensatz zu geförderten Reproduktionsaufgaben in geringem Maße assoziative Prozesse.

Im nächsten Kapitel wird aufgezeigt, wie die Hypothese des assoziativen Defizits empirisch untersucht wurde und inwiefern sich die gerade beschriebenen Vorhersagen bestätigen.

3.2.2 Empirische Befunde zur Assoziationsdefizit-Hypothese

In diesem Kapitel soll ein Überblick über die empirischen Befunde zum Assoziationsdefizit älterer Menschen gegeben werden. Neben der grundlegenden Frage, inwieweit das Defizit empirisch nachgewiesen werden kann, sollen dabei auch moderierende Faktoren und Generalisierungsaspekte beleuchtet werden. Um nicht nur einen qualitativen Überblick über die Literatur zu geben, sondern auch quantitative Maße der Effektstärke zu berichten, wird der Aufbau dieses Kapitels an eine aktuelle Meta-Analyse von Old und Naveh-Benjamin (2008a) angelehnt. Die Meta-Analyse soll dabei als Gerüst dienen und wird an einigen Stellen durch die detaillierte Schilderung wichtiger Experimente ergänzt. Old und Naveh-Benjamin (2008a) verfolgen in ihrer Meta-Analyse fünf Ziele:

- (1) die Quantifizierung der Effektstärke des Assoziationsdefizits
- (2) die Erfassung des Ausmaßes des Assoziationsdefizits in Abhängigkeit von verschiedenen Assoziationsarten
- (3) der Vergleich von unterschiedlichen Materialmodalitäten
- (4) die Untersuchung der Unabhängigkeit des Defizits von unterschiedlichen Testformaten
- (5) den Einfluss inzidenteller und intentionaler Lerninstruktionen.

Um eine geeignete Datenbasis für die Meta-Analyse zu erhalten, wurden verschiedene Aufnahmekriterien festgelegt. Die Studien mussten eine ältere Stichprobe von mindestens 60 Jahren Durchschnittsalter mit einer jüngeren Stichprobe von maximal 35 Jahren Durchschnittsalter vergleichen. Jedes Experiment musste getrennte Maße für Itemgedächtnisleistung und assoziative Gedächtnisleistung enthalten sowie die notwendigen Statistiken, um für Item- und assoziative Gedächtnisleistungen Effektstärkemaße zu berechnen. Die Studien mussten sich dabei nicht direkt mit dem assoziativen Defizit befassen, sondern konnten auch aus der Quellengedächtnisforschung oder der Forschung zur Kontext-Defizithypothese

stammen.⁷ Aufgrund dieser Kriterien konnte eine Datenbasis von 90 Experimenten aus 65 veröffentlichten Studien zusammengestellt werden. Die gewichteten Effektstärken wurden jeweils für den Altersunterschied in der Item- bzw. assoziativen Gedächtnisleistung berechnet (Hedges & Olkin, 1985; Old & Naveh-Benjamin, 2008a). Positive Effektstärken kennzeichneten dabei eine bessere Gedächtnisleistung der jüngeren Menschen. Im Folgenden werden die Ergebnisse zu den fünf Zielen der Meta-Analyse beschrieben.

(1) Quantifizierung des Assoziationsdefizits. Im Hinblick auf das erste Ziel der Meta-Analyse, die Quantifizierung der Effektstärke, kann festgehalten werden, dass die Effektstärke des Altersunterschieds in Itemgedächtnistests 0.73 betrug, wohingegen für die assoziativen Tests eine durchschnittliche Effektstärke von 0.92 berechnet wurde. Da sich die 95%-Konfidenzintervalle dieser beiden Mittelwerte nicht überlappten, wurde dieser Unterschied als signifikant eingestuft (siehe Abbildung 3.3). Gemittelt über 90 Studien mit unterschiedlichem Itemmaterial sowie unterschiedlichen Enkodierungs- und Testbedingungen wurde also ein signifikant größerer Altersunterschied für assoziative Gedächtnismaße im Vergleich zu Itemgedächtnismaßen ausgemacht. Die Assoziationsdefizit-Hypothese wird durch dieses empirische Ergebnis gestützt.

(2) Effekte unterschiedlicher Modalitäten des Materials. Um die Generalisierbarkeit des Assoziationsdefizits zu prüfen, wurden Materialeffekte untersucht. Das verwendete Material wurde dazu in die zwei Klassen, verbales vs. nonverbales Material, eingeteilt. In die verbale Kategorie fallen dabei 53 Datensätze, in denen entweder visuell oder akustisch präsentiertes Wortmaterial verwendet wurde (z.B. Dumas & Hartman, 2003; Naveh-Benjamin, 2000; Naveh-Benjamin, Guez & Shulman, 2004). Die nonverbalen Studien bilden eine relativ heterogene Klasse von 22 Studien. Hierunter werden sowohl Materialien zusammengefasst wie visuell präsentierte Gesichter und Objekte (z.B. Bastin & Van der Linden, 2005, 2006; Hess & Slaughter, 1990; Lyle, Bloise & Johnson, 2006; Newman, Allen & Kaszniak, 2001), als auch Studien, in denen die Erinnerung für die Reihenfolge von Handlungsausführungen erfragt wurde (z.B. Kausler & Phillips, 1988). Die Ergebnisse zeigten, dass sich die Konfidenzintervalle zwischen den beiden Materialarten weder für die Itemeffekte noch für die assoziativen Effekte unterschieden. Daher schlussfolgerten die Autoren, dass die Verwendung von verbalem oder nonverbalem Material keine Auswirkungen auf das Altersdefizit hat.

(3) Modulation des Assoziationsdefizits durch verschiedene Assoziationsarten. Eine zweite Einteilung des Materials erfolgte aufgrund der Assoziationsart, die zwischen

⁷Einige der Studien, die in die Meta-Analyse von Old und Naveh-Benjamin (2008a) eingingen, wurden auch in der Meta-Analyse von Spencer und Raz (1995) ausgewertet (s. Kap. 3.1). Die Altersunterschiede für Quelleninformationen oder Kontextinformationen gehen nun als assoziative Maße in die Analyse ein, nicht mehr als Kontextinformationen.

den Elementen gebildet werden musste. Die Autoren unterschieden dabei sechs unterschiedliche Assoziationsarten, wobei die Auswahl theoretisch nicht weiter begründet wurde. Die Einteilung war vielmehr abhängig von der Art der Einzelelemente, die verknüpft werden sollten. Je nachdem, ob ein dargebotenes Wort mit der Hintergrundfarbe oder aber der Darbietungsposition assoziiert werden sollte, wurden die Experimente in verschiedene Kategorien aufgeteilt. Inwieweit diesen Kategorien wirklich unterschiedliche Assoziationsarten d.h. Bindungsmechanismen zugrunde liegen, wurde als explorative Frage formuliert, die abhängig von den Ergebnissen des metaanalytischen Vergleichs beantwortet werden sollte. Falls sich unterschiedliche Alterseffekte zwischen den Assoziationskategorien zeigen sollten, würden die Autoren von unterschiedlichen Bindungsmechanismen ausgehen.

Zu einer ersten Kategorie wurden unter dem Label ‚Quellenassoziation‘ 28 Studien zusammengefasst, in denen erinnert werden musste, ob eine Information gelesen oder selbst generiert wurde bzw. von welcher Person eine Information präsentiert wurde (z.B. Rabinowitz, 1989; Schacter et al., 1991; Simons, Dodson, Bell & Schacter, 2004; Wegesin, Friedman, Varughese & Stern, 2002; Wegesin, Jacobs, Zubin, Ventura & Stern, 2000). In diese Klasse fallen u.a. Altersstudien, die typische Quellengedächtnisaufgaben verwendeten. In die Kategorie der ‚Kontextassoziation‘ (11 Studien) wurden Experimente eingeordnet, bei denen Versuchspersonen entscheiden mussten, ob sie ein Wort in Groß- oder Kleinbuchstaben gelesen haben bzw. in welcher Schriftart ein Wort präsentiert wurde (z.B. Kausler & Puckett, 1980; Naveh-Benjamin, 2000, Exp. 3). Da die Autoren relativ spezifische Assoziationsarten unterscheiden, bilden Studien, die eine Erinnerung an die Modalität (akustisch oder visuell) bzw. die Darbietungslokalität erforderten, eigene Kategorien. In die spezifische Gruppe der ‚Modalitätsassoziationen‘ fallen dabei nur 6 Experimente (z.B. Lehman & Mellinger, 1984). In der Klasse der ‚Lokalitätsassoziationen‘ wurden 12 Experimente zusammengefasst (z.B. Bastin & Van der Linden, 2006; Light & Zelinski, 1983). Schließlich wurden 20 Studien, in denen die zeitliche Reihenfolge von präsentierten Wörtern oder ausgeführten Handlungen erinnert werden musste (z.B. Dumas & Hartman, 2003; Newman et al., 2001), unter dem Aspekt der ‚zeitlichen Assoziation‘ geclustert. Die letzte Gruppierung bezieht sich auf 10 Studien zu ‚Item-Item-Assoziationen‘. Dies konnten entweder Wortpaare, Kombinationen von Gesicht und Name oder auch zwei Gesichter sein (z.B. Bastin & Van der Linden, 2006; Naveh-Benjamin, 2000; Naveh-Benjamin, Guez, Kilb & Reedy, 2004). Der Unterschied zu den anderen Assoziationsarten besteht darin, dass zwei zentrale Informationseinheiten (im Sinne der Kontextdefizit-Hypothese) miteinander verknüpft werden müssen.

Generell gilt hinsichtlich der Einteilung in Assoziationsarten, dass jedes der 90 Experimente nur einer Assoziationsart zugeteilt wurde. Die Ergebnisse sind in Abbildung 3.3 veranschaulicht. Es fällt auf, dass bei fünf der sechs Assoziationsarten die Altersunterschiede in der assoziativen Komponente größer ausfallen als in der Itemkomponente. Die Ausnahme bildet die Kategorie der Modalitätsassoziationen. Hier findet sich deskriptiv

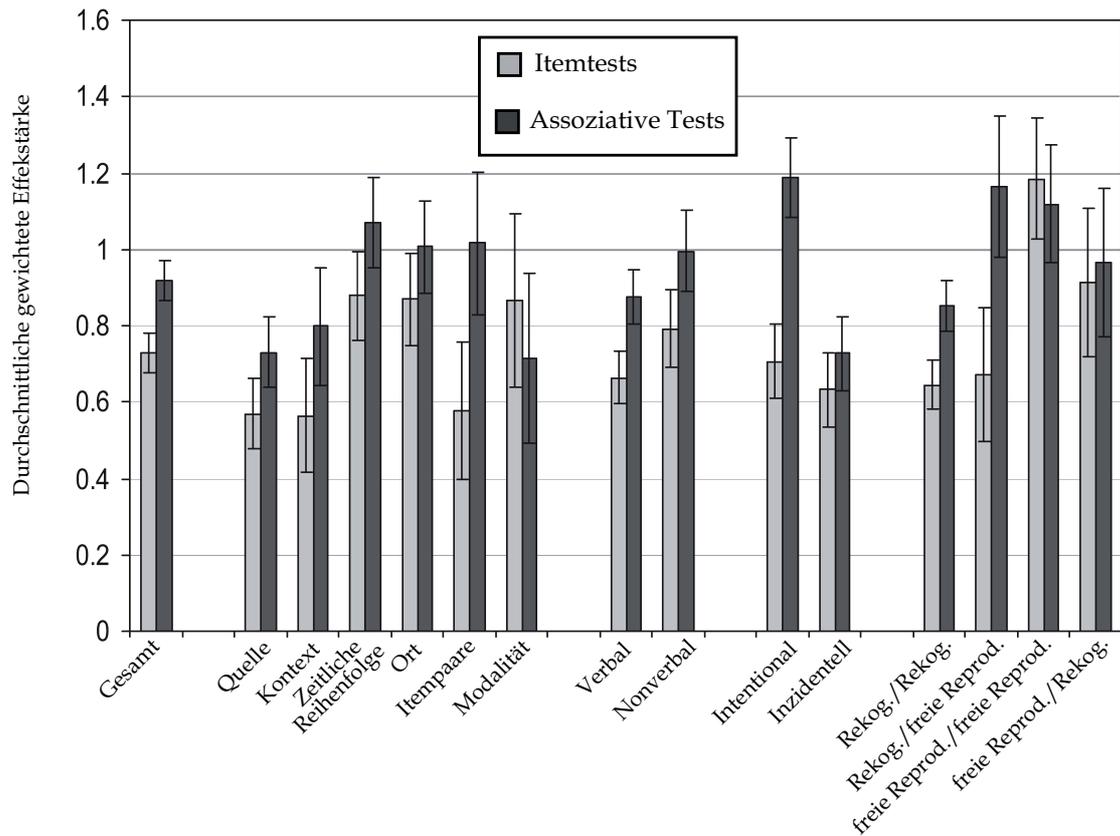


Abbildung 3.3: In dieser Abbildung sind die Ergebnisse der Meta-Analyse von Old und Naveh-Benjamin (2008a) abgebildet. Die Balken zeigen jeweils die Effektstärken d.h. den Altersunterschied zwischen jungen und älteren Menschen an. Hellgrau sind dabei die Altersunterschiede für Itemtests, dunkelgrau die Altersunterschiede für assoziative Tests dargestellt. Die Fehlerbalken repräsentieren die 95% Konfidenzintervalle. Ganz links sind die Effektstärken über alle 90 Studien hinweg angezeigt, darauf folgen die Effektstärken in Abhängigkeit von unterschiedlichen Kontextfaktoren. [Abbildung aus Old und Naveh-Benjamin (2008a), S. 111]

der entgegengesetzte Trend. Da dieser Unterschied aber nicht signifikant ist und diesem Vergleich nur sechs Experimente zugrunde liegen, wurde dieser Ausnahme von den Autoren keine größere Bedeutung beigemessen. Auch in den anderen fünf Assoziationsarten sind einige Unterschiede zwischen Item- und assoziativer Gedächtnisleistung nicht signifikant, wobei einheitlich ein stärkeres Altersdefizit für die Assoziationen im Vergleich zu Items festzustellen ist. Die größte Differenz in den Effektstärken weisen dabei klar die Item-Item-Assoziationen auf. Da sich keine gravierenden Unterschiede zwischen den verschiedenen Kategorien finden, wurde die Annahme grundsätzlich verschiedener Assoziationsmechanismen von den Autoren nicht weiter diskutiert. Ebenso wie die Aufteilung der Assoziationsarten relativ theoriefrei vorgenommen wurde, so wurden auch die Ergebnisse ohne Bezug zu theoretischen Implikationen erläutert. Der einzige Schluss, der aus diesem Vergleich unterschiedlicher Assoziationsarten gezogen wurde, ist die Allgemeingültigkeit des assoziativen Defizits, das sich unabhängig von den zu verknüpfenden Informationen zeigte. Alternative und theoretisch fundierte Unterteilungen werden später in diesem Kapitel vorgestellt.

(4) Einfluss des Testformats auf das Assoziationsdefizit. Weitere Faktoren, die untersucht wurden, sind Charakteristika der Testsituation. Da das assoziative Defizit über den Vergleich von Item- und assoziativer Gedächtnisleistung definiert ist, sind immer zwei Gedächtnistests zur Messung notwendig. Normalerweise werden in der Lernphase Wortpaare präsentiert, in der Testphase wird dann in zufälliger Reihenfolge ein Itemtest und ein assoziativer Test durchgeführt. Bei dem Itemtest muss abhängig vom Testformat (Rekognition oder Reproduktion) ein gelerntes Wort reproduziert bzw. wiedererkannt werden. Dabei müssen die Items unabhängig von ihrer Darbietungsposition erinnert werden. Im assoziativen Rekognitionstest werden nur alte Wörter dargeboten. Hier muss die Entscheidung getroffen werden, ob die beiden Wörter in dieser Kombination gelernt wurden, oder ob eine neue Kombination vorliegt. Wird der assoziative Test als freier Reproduktionstest durchgeführt, so muss die Versuchsperson die Wortpaare frei erinnern.

Für den Vergleich von Item- und assoziativem Test gilt es, die beiden Verfahren so auszuwählen, dass sie sich möglichst nur in der Erfassung von assoziativen Prozessen unterscheiden und damit maximal sensitiv für diesen Gedächtnisprozess sind. Die Tests sollten also z.B. hinsichtlich der Aufgabenkomplexität und des Antwortmodus parallelisiert sein, aber in unterschiedlichem Ausmaß auf assoziativen Gedächtnisprozessen beruhen. Gänzlich ungeeignet zur Erfassung der Itemerinnerung sind aus diesem Grund geförderte Reproduktionstests. Diese Art von Gedächtnistest involviert sowohl Item- als auch assoziative Prozesse (für eine Diskussion dieses Problems siehe Provyn, Sliwinski & Howard, 2007). Es muss zwar nur ein Item reproduziert werden, aber dessen korrekte Wiedergabe erfordert nicht nur Itemgedächtnisprozesse, sondern auch die Erinnerung an die Verbindung zum Hinweisreiz. Man könnte einen geförderten Reproduktionstest also sowohl als

Itemtest als auch als assoziativen Test einordnen. Daher ist es schwierig, einen entsprechenden Vergleichstest zu finden. Würde man den geförderten Reproduktionstest als Maß für die assoziative Gedächtnisleistung anlegen, so stellt sich die Frage, ob ein Rekognitionstest als vergleichbarer Itemtest geeignet ist. Da sich Rekognitionstests und geförderte Reproduktionstests nicht nur in der Beteiligung von assoziativen Prozessen unterscheiden, sondern z.B. auch in Bezug auf Abrufprozesse, scheint diese Wahl dem Ziel maximaler Vergleichbarkeit der Testsituationen nicht näher zu kommen. Eine ähnliche Argumentation kann geführt werden, wenn der geförderte Reproduktionstest als Item-Maß und ein freier Reproduktionstest als assoziativer Tests genutzt werden. Aus diesem Grund finden geförderte Reproduktionstests zur gezielten Erfassung des assoziativen Defizits keine Verwendung und Old und Naveh-Benjamin (2008a) haben Studien, die geförderte Reproduktionstests nutzten, aus ihrer Meta-Analyse ausgeschlossen.

Aufgrund der genannten Probleme sind auch Experimente, in denen freie Reproduktionstest zur Erfassung der Itemgedächtnisleistung mit Rekognitionstests zur Erfassung der assoziativen Komponente (oder umgekehrt) verglichen werden, nicht optimal. Rekognitions- und Reproduktionstests weisen Unterschiede in der Aufgabenkomplexität auf, beruhen in unterschiedlichem Maße auf Abrufprozessen und auch das Ausmaß der benötigten selbstinitiierten Prozesse ist verschieden. Allein aufgrund dieser Differenzen sind, unabhängig von einem assoziativen Defizit, bereits Altersunterschiede zu erwarten. Aus diesen Gründen sollten Item- und assoziative Gedächtnistests keine unterschiedlichen Testarten vermischen, sondern entweder über zwei Rekognitionstests oder zwei freie Reproduktionstests erfasst werden. Ein wichtiger Vorteil des Rekognitionsparadigmas ist, dass dieses weniger auf selbstinitiierten Prozessen beruht und Abrufprozesse eine geringere Rolle spielen. Diese Faktoren führen dazu, dass sich in der Itemrekognition nur relativ geringe bis gar keine Altersunterschiede zeigen (Provyne et al., 2007; Zacks et al., 2000) und das Rekognitionsparadigma daher als sensitives Messinstrument für die Erfassung eines assoziativen Defizits eingestuft werden kann. Im Vergleich von assoziativem Gedächtnistest und besteht die Möglichkeit, den Antwortmodus der beiden Testarten zu parallelisieren. Zwischen einem assoziativen Rekognitionstest und einem ‚two-alternative forced-choice‘ Itemrekognitionstest (2AFC-Itemrekognitionstest) kann der Darbietungsmodus (Präsentation von zwei Wörtern in der Testphase) konstant gehalten werden.

Aus diesen genannten Gründen hat sich der Vergleich von Itemrekognitionstest (Ja-nein- oder 2AFC-Test) und assoziativem Rekognitionstest für die Untersuchung des assoziativen Defizits durchgesetzt. Bereits in den ersten Studien von Chalfonte und Johnson (1996), Bayen, Phelps und Spaniol (2000) und Naveh-Benjamin (2000) wurden Rekognitionstests verwendet, und dies wurde in den darauf folgenden Untersuchungen aus gutem Grund beibehalten (z.B. Castel & Craik, 2003; Cohn & Moscovitch, 2007; Naveh-Benjamin, Brav & Levy, 2007; Naveh-Benjamin, Guez, Kilb & Reedy, 2004; Shing et al., 2008).

Das Standardverfahren geht dabei auf Humphreys (1976, 1978) zurück und sieht eine Lernphase vor, in der jeweils zwei Wörter präsentiert werden. Danach findet in randomi-

sierter Reihenfolge ein Item-Rekognitionstest sowie ein assoziativer Rekognitionstest statt. Im assoziativen Rekognitionstest behalten die Wörter ihre Position bei, d.h. Wörter, die als linkes Wort eines Wortpaares präsentiert wurden, werden auch in dieser Position getestet. Gleiches gilt für einen 2AFC-Item-Rekognitionstest; bei einem Ja-nein-Rekognitionstest werden sowohl links als auch rechts dargebotene Wörter getestet. Jedes Wort wird dabei nur einmal verwendet. Dieses Standardverfahren aus der Kombination von Item- und assoziativem Rekognitionstest lag 70% der Studien, die in die Analyse des Testformats einfließen, zugrunde. Die verbleibenden Studien, die meist aus den 80er Jahren stammten und eher auf die Erfassung des Kontextdefizits als des Assoziationsdefizits abzielten, verwendeten zu 8% den Vergleich von zwei Reproduktionstests, zu 10% sowie 11% die Mischformen aus Rekognition und Reproduktion.

Da sich die Studien so ungleich auf die Testarten verteilen, sollte die Anzahl der zugrundeliegenden Experimente bei der Interpretation der Ergebnisse berücksichtigt werden. Die Analysen bestätigten ein reliables assoziatives Defizit bei der Verwendung zweier Rekognitionstests. Noch stärker ausgeprägt ist der Altersunterschied zwischen Item- und assoziativer Gedächtnisleistung, wenn für das Itemgedächtnis ein Rekognitionstest, für die assoziative Gedächtnisleistung ein freier Reproduktionstest verwendet wurde (basierend auf 8 Experimenten). Es wurde bereits erläutert, dass dieser Vergleich aber nicht nur auf assoziative Prozesse zurückgeführt werden kann. Im Vergleich von zwei Reproduktionstests bzw. der Kombination Reproduktion und Rekognition (7 bzw. 9 Experimente) findet sich keine Evidenz für ein assoziatives Defizit. Gründe, warum dieses Ergebnismuster nicht als kritisch für ein assoziatives Defizit angesehen werden muss, wurden ebenfalls schon angesprochen.

An dieser Stelle sollen abschließend noch zwei Experimente angeführt werden, die die Bedeutung des Testformats experimentell und damit unter Kontrolle konfundierender Faktoren untersucht haben.

In der ersten Studie von Naveh-Benjamin (2000, Exp. 4) wurde die Alterssensitivität unterschiedlicher Itemtests untersucht. Die Probanden, zwei Gruppen älterer und jüngerer Erwachsener, mussten eine Liste von Wortpaaren lernen. Anschließend wurde ihre Gedächtnisleistung für das jeweils zweite (rechts stehende) Wort eines Wortpaares getestet. Dies erfolgte entweder durch einen Ja-nein-Rekognitionstest, durch einen geförderten Reproduktionstest, bei dem das erste Wort des Wortpaares als Hinweisreiz vorhanden war, oder durch einen freien Reproduktionstest. Wie in diesem Kapitel erläutert wurde, sind nicht nur Itemgedächtnisprozesse, sondern auch assoziative Gedächtnisprozesse für die erfolgreiche Leistung in einem geförderten Reproduktionstest notwendig. In der freien Item-Reproduktion oder Item-Rekognition dahingegen sind keine assoziativen Erinnerungsmechanismen involviert. Nach der Hypothese des assoziativen Defizits sollten daher Altersunterschiede in der geförderten Reproduktion größer ausfallen als in der Itemrekognition und freien Reproduktion. Genau dieses Ergebnismuster resultierte aus dem

Experiment von Naveh-Benjamin (2000): Die Altersunterschiede waren in der geförderten Reproduktion am stärksten ausgeprägt. Damit stützt dieses Experiment die Sonderrolle von geförderten Reproduktionstests und schließt sie als geeignete Gedächtnistests zur Erfassung der Itemrekognitionsleistung aus. Interessant an dieser Studie ist die Tatsache, dass sich in freien Reproduktionstests nicht die größten Altersunterschiede finden, obwohl weniger Hinweisreize gegeben werden als in den beiden anderen Tests und damit selbstinitiierte Prozesse sowie Abrufprozesse stärker involviert sind.⁸ Das Altersdefizit für assoziative Prozesse scheint hier einen besonders großen Unterschied auszulösen.

In einem weiteren Experiment haben Naveh-Benjamin, Guez und Shulman (2004) genauer untersucht, ob das assoziative Defizit im Rekognitionsparadigma von der Art des Itemrekognitionstests abhängt. Grundsätzlich kann entweder ein Ja-nein-Itemrekognitionstest oder ein 2AFC-Itemrekognitionstest verwendet werden. Der Vorteil des Ja-nein-Itemrekognitionstests besteht darin, dass das Antwortformat (Ja-nein-Entscheidung, welches Item alt ist) mit dem Antwortformat des assoziativen Rekognitionstests vergleichbar ist (Ja-nein-Entscheidung welches Wortpaar eine alte Kombination darstellt). Ein 2AFC-Itemrekognitionstest dahingegen ist hinsichtlich der Menge der dargebotenen Items mit dem assoziativen Test äquivalent, bei beiden Testarten werden jeweils zwei Items präsentiert. Es können also Vorteile in der Verwendung beider Itemtests gesehen werden. Um zu untersuchen, ob Altersunterschiede eher durch das Testformat oder durch die dargebotene Itemmenge gefördert werden, haben Naveh-Benjamin und Kollegen die Testart experimentell manipuliert. Jede Versuchsperson musste einen Ja-nein-Rekognitionstest, einen 2AFC-Rekognitionstest und einen assoziativen Rekognitionstest bearbeiten, wobei die Reihenfolge der Tests über die Teilnehmer hinweg ausbalanciert wurde. Die Analysen zeigten, dass das assoziative Defizit nicht von der Art des verwendeten Itemtests moderiert wird.

Zusammenfassend kann hinsichtlich des experimentellen Paradigmas Folgendes festgehalten werden: Sowohl die assoziative Gedächtnisleistung als auch die Itemgedächtnisleistung sollten über Rekognitionstests gemessen werden. Dies kann theoretisch gut begründet werden, da beim Vergleich zweier Rekognitionsaufgaben u.a. der Einfluss von Abrufprozessen und die Komplexität der Aufgabe konstant gehalten werden. Darüber hinaus bewährte sich das Paradigma sowohl im metaanalytischen Vergleich (Old & Naveh-Benjamin, 2008a) als auch über experimentelle Studien (Naveh-Benjamin, 2000; Naveh-Benjamin, Guez & Shulman, 2004) als sensitiv für die Erfassung des assoziativen Defizits.

⁸Dieses Ergebnis steht zunächst auch im Widerspruch zu Studien von Craik und McDowd (1987), da hier die größten Altersunterschiede in freien Reproduktionstests erreicht wurden. Allerdings konnten in den Studien von Craik und McDowd alle Wörter frei erinnert werden; im vorliegenden Test von Naveh-Benjamin wurden nur die rechts präsentierten Wörter als richtige Erinnerungen gewertet. Dadurch weist dieser freie Reproduktionstest weniger Kompensationsmöglichkeiten auf und ist direkt vergleichbar zu den anderen beiden Testarten, bei denen auch jeweils nur die rechten Wörter der Kombination getestet wurden.

(5) Einfluss inzidenteller und intentionaler Lerninstruktionen. Ein letzter Aspekt, der in der Meta-Analyse untersucht wurde, ist die Bedeutung der Lernsituation. Die Experimente wurden dazu einer der beiden Klassen ‚inzidentelle‘ oder ‚intentionale Lernbedingung‘ zugeordnet. Intentional meint hier nicht nur, dass die Probanden über einen späteren Gedächtnistest informiert waren, sondern dass sie genau wussten, welche Art des Gedächtnistests folgte, d.h. ob es sich dabei um einen Test für Iteminformation oder um einen Test für assoziative Informationen handelte. Dabei wurden Experimente nur dann in eine der Kategorien aufgenommen, wenn die Lernbedingungen (intentional vs. inzidentell) für beide Tests übereinstimmten. Wurde den Probanden also vor der Lernphase sowohl der Item- als auch der assoziative Test erklärt, galt die Studie als intentionale Lernstudie. Wurde nur auf einen der beiden Tests aufmerksam gemacht, so wurde die Studie nicht in diese Klassifikation aufgenommen, da die Lernsituation nicht für beide Tests einheitlich war. Für den verbleibenden Fall, dass die Probanden auf keinen der beiden Tests hingewiesen wurden, sondern sie das Material unter einem anderen Vorwand bearbeiteten, war die Lernsituation inzidentell. Aufgrund dieser Kriterien konnte nur etwas mehr als die Hälfte der Experimente der inzidentellen (28 Experimente) oder intentionalen (23 Experimente) Lernbedingung zugeteilt werden. In diesem Vergleich zeigten die Ergebnisse ein sehr klares Bild: Die Effektgröße für Assoziationen betrug unter inzidentellen Bedingungen 0.73, unter intentionalen dahingegen 1.19; demgegenüber zeigte sich nur ein schwacher Effekt der Lernsituation auf die Itemleistung mit einer Effektgröße von 0.63 vs. 0.71.

Um die Ursache dieses Unterschieds zu erklären, ist es aufschlussreich, Studien zu Rate zu ziehen, die gezielt den Aspekt der Lernsituation (inzidentell vs. intentional) manipulierten. Da es zahlreiche Experimente gibt, die sich innerhalb dieses Forschungsfelds mit den Auswirkungen von Instruktionen befassen (u.a. Bastin & Van der Linden, 2005; Kausler & Phillips, 1988; Lipman & Caplan, 1992; Mitchell et al., 2000; Puglisi, Park, Smith & Hill, 1985; Schmitter-Edgecombe & Simpson, 2001), diese an dieser Stelle aber nicht alle diskutiert werden sollen, wurden anhand der folgenden Kriterien zwei Experimente ausgewählt: Die Studien sollten die Gedächtnisleistungen (Item und assoziative) anhand von Rekognitionstests erfassen und die Rekognitionsleistung von beiden Einzelkomponenten (der zentralen Information und der Kontextinformation) sollte erhoben werden, um auszuschließen, dass die assoziative Gedächtnisleistung durch Leistungsunterschiede in der Rekognition der Komponenten verzerrt ist (siehe Kap. 3.1; Kritik an Methode der Kontextdefizit-Hypothese). Sowohl die bereits in anderem Kontext (siehe Kap. 3.1; Empirische Befunde) erwähnten Studien von Naveh-Benjamin (2000, Experiment 2) und Chalfonte und Johnson (1996, Experiment 3b) erfüllen die genannten Kriterien. In der Studie von Chalfonte und Johnson mussten sich die älteren und jüngeren Teilnehmer 30 Strichzeichnungen in 30 verschiedenen Farben merken. In der Lernphase wurde variiert, ob die Teilnehmer die Instruktion erhielten, nur die Zeichnungen zu lernen, oder ob sie Anweisung hatten, sich sowohl die Zeichnungen als auch die jeweiligen Farben einzuprägen. Die Ergebnisse zeigten,

dass jüngere Menschen ihre Leistung im assoziativen Gedächtnistest signifikant verbesserten, wenn hinsichtlich beider Komponenten intentionale Lernbedingungen vorlagen. Ältere Menschen waren zu einer solchen Leistungssteigerung nicht in der Lage. Bei älteren Menschen wird die Leistung nicht von der Art der Lerninstruktion moderiert. Unklar blieb bei dieser Studie, inwieweit die Probanden in der für beide Komponenten intentionalen Bedingung auch explizit darauf hingewiesen wurden, die Kombination der beiden Elemente zu lernen. Die Autoren schreiben, dass die Probanden aufgefordert wurden, entweder nur die Zeichnungen oder die Zeichnungen und die Farben zu lernen. Es bleibt intransparent, ob für die Probanden eine intentionale Lernbedingung in Hinsicht auf die beiden Komponenten vorlag oder in Hinsicht auf die Assoziation. Diesen Aspekt manipulierte Naveh-Benjamin in einem seiner Experimente und, um die Ergebnisse vorwegzunehmen, er replizierte damit das Ergebnismuster von Chalfonte und Johnson.

Naveh-Benjamin (2000) bot den Probanden Wortpaare dar, und die Instruktion lautete entweder, beide Wörter einzeln für einen Itemerkennungstest zu lernen, oder aber die Wortpaare als Paar zu lernen, um sich auf einen Wortpaartest vorzubereiten. Zusätzlich wurde nach einigen Übungsaufgaben der angekündigte Item- oder assoziative Test durchgeführt, so dass die Teilnehmer (vermeintlich) wussten, welche Testart zu erwarten war. Nach der Präsentation von 40 Wortpaaren erhielten dann beide Teilnehmergruppen sowohl den angekündigten, mit den Instruktionen kohärenten Test, als auch zusätzlich den inkohärenten Item- bzw. assoziativen Test. Die Ergebnisse (Abbildung 3.3) bestätigten, dass jüngere Menschen von den instruktionalen Hinweisen und der Vertrautheit mit dem Test profitieren, da sie ihre assoziative Gedächtnisleistung verbessern konnten; bei älteren Menschen bestand keine Leistungsveränderung. Im Vergleich zwischen den Altersgruppen zeigte sich das assoziative Defizit in beiden Instruktionsbedingungen, war aber in der *assoziativ-intentionalen* Bedingung stärker ausgeprägt.

In beiden Experimenten hatte der explizite Hinweis auf die Art des Tests eine Leistungssteigerung bei jüngeren Menschen zur Folge. Wahrscheinlich aufgrund der Fähigkeit zu einer flexiblen Strategieverwendung sind jüngere Menschen in der Lage, ihre Gedächtnisleistung, insbesondere für Assoziationen, zu steigern. Ältere Menschen sind zu einer solchen Leistungsverbesserung - zumindest unter diesen Bedingungen - nicht fähig (siehe aber Kilb & Naveh-Benjamin, 2007). Weiterführende Studien haben nicht nur manipuliert, ob die Teilnehmer über die Art des folgenden Tests aufgeklärt wurden, sie wurden auch gezielt über geeignete assoziative Strategien informiert bzw. sogar in der Strategieverwendung trainiert. Inwieweit ältere Menschen von diesen Bedingungen profitieren und welche Konsequenz sich aus diesen Studien für den strategischen Aspekt des Assoziationsdefizits ergibt, wird in Kapitel 3.2.3 detailliert erläutert.

Zusammenfassend wurde in der Meta-Analyse von Old und Naveh-Benjamin also zum einen die Effektstärke des assoziativen Defizits über 90 Studien hinweg bestimmt. Es zeigte

sich dabei, dass die Effektstärken der Altersunterschiede für assoziative Informationen größer sind, als die der Altersunterschiede für einzelne Komponenten. Weder die Modalität der Komponenten noch die Assoziationsart scheinen dabei eine bedeutende moderierende Rolle zu spielen. Allerdings konnte ein starker Einfluss von intentionalen vs. inzidentellen Lernbedingungen ausgemacht werden, und die Bedeutung des Rekognitionsparadigmas zur Untersuchung des assoziativen Defizits wurde aufgezeigt.

Da die Studien für die Meta-Analyse Ende des Jahres 2006 recherchiert wurden, konnten folglich keine Veröffentlichungen aus 2007 oder 2008 in die Zusammenfassung einfließen. An dieser Stelle sei daher kurz auf drei Studien hingewiesen, die im Jahre 2008 erschienen und das assoziative Defizit ebenfalls empirisch unterstützen. James, Fogler und Tauber (2008) konnten das Defizit für die Assoziation von Namen und Gesichtern nachweisen und damit die Befunde von Naveh-Benjamin, Guez und Shulman (2004) replizieren. Kersten, Earles, Curtayne und Lane (2008) zeigten, dass sich ältere Menschen auch schlechter an die Kombination von Handlungen und ausführender Person erinnern können. In ihrer Studie wurden Tätigkeiten wie z.B. eine Banane schälen oder eine Fahne schwenken von unterschiedlichen Personen vorgenommen. Auch unter diesen eher naturalistischen und alltagsnahen Bedingungen konnte das Altersdefizit repliziert werden (siehe auch Old & Naveh-Benjamin, 2008b). Weitere Veröffentlichungen aus den Jahren 2007 und 2008 (Buchler, Light & Reder, 2008; Cohn et al., 2008; Cohn & Moscovitch, 2007; Diana, Yonelinas & Ranganath, 2008; Ecker et al., 2007; Jäger & Mecklinger, 2009; Kilb & Naveh-Benjamin, 2007; Pantelis, Van Vugt, Sekuler, Wilson & Kahana, 2008; Provyn et al., 2007; M. G. Rhodes, Castel & Jacoby, 2008; S. M. Rhodes & Donaldson, 2007, 2008; Shing et al., 2008) werden noch aufgegriffen. Wichtige ältere Studien, die nicht in die Meta-Analyse einfließen, wurden bereits erläutert (Bayen, Phelps & Spaniol, 2000; Chalfonte & Johnson, 1996) oder werden es noch (Castel & Craik, 2003; Giovanello, Keane & Verfaellie, 2006; M. R. Healy, Light & Chung, 2005; Jäger, Mecklinger & Kipp, 2006; Light, Patterson, Chung & Healy, 2004).

Abschließend sollen zwei Aspekte der Meta-Analyse kritisch diskutiert werden. Dies betrifft zum einen Einschränkungen in Hinsicht auf die Interpretation der Ergebnisse, zum anderen Kritik an der Klassifikation der Studien.

Die Autoren berichten große Effekte der Lernbedingungen auf das Assoziationsdefizit. Daher ist es fraglich, inwieweit die anderen Analyseergebnisse z.B. hinsichtlich verbaler vs. nonverbaler Modalität sinnvoll interpretiert werden können. Eine unverzerrte Interpretation wäre nur dann sichergestellt, wenn der Modalitätsfaktor nicht mit dem Faktor Lernbedingung konfundiert wäre. Da die Autoren aber weder die Kontrolle einer solchen Konfundierung vornehmen, noch aufzeigen, welche Studien jeweils in welche Kategorien eingeordnet wurden, kann der Leser nicht abschätzen, inwieweit die Ergebnisse verzerrt sind. Dies gilt auch für die Analysen in Abhängigkeit von unterschiedlichen Assoziationsar-

ten. Auch hier kann eine Konfundierung mit den Faktoren Lernbedingung oder Testbedingung artifizielle Ergebnisse auslösen. Die Autoren erwähnen dieses Problem zwar kurz in der Diskussion, ziehen aber keine Konsequenzen daraus. Auch wenn eine Meta-Analyse diese - den Studien immanente - Probleme nicht aufheben kann, so sollte dieser Aspekt nicht nur kurz angesprochen werden, sondern Konsequenzen für die Darstellung der Ergebnisse haben. Entweder sollte eine transparente Zuteilung der Studien zu den unterschiedlichen Analysen erfolgen oder entsprechende Kontrollen sollten vorgenommen werden. Da dies in der Meta-Analyse nicht der Fall war, wurden dem Leser dieser Arbeit zu den beiden zentralen Aspekten Lernsituation und Testart jeweils experimentelle Studien vorgestellt, die eine klare Interpretation erlauben. Erfreulicherweise stimmten dabei die Schlüsse, die sich aus dem metaanalytischen Vergleich und den intern validen Experimenten ergaben, überein.

Ein zweiter, diskussionswürdiger Aspekt ist die Einteilung der Studien in sechs verschiedene Assoziationsarten. Die Einteilung kann zwar Post-Hoc nachvollzogen werden, warum sich aber unterschiedliche Alterseffekte für die Assoziation von Wort und Schriftart vs. Wort und Lokalisation ergeben sollten, wird theoretisch nicht bzw. sehr schwach begründet. Dabei gibt es mindestens zwei theoretische Aspekte, anhand derer Vorhersagen getroffen werden könnten. Dies ist (1) das Zuordnungsverhältnis der Komponenten und (2) das Integrationspotential der Komponenten. Um diese dem Leser nicht vorzuenthalten, werden sie im Folgenden erläutert.

(1) Zuordnungsverhältnis

Eine aus theoretischer Hinsicht interessante Klassifikationsdimension stellt das Zuordnungsverhältnis („Mapping“) zwischen den einzelnen Komponenten dar (z.B. zwischen den einzelnen Wörtern eines Wortpaares). Eine einfache Zuordnung („one to one mapping“) liegt bei den meisten assoziativen Rekognitionsstudien vor. Hier werden z.B. 20 Namen mit 20 verschiedenen Gesichtern gepaart. Jeder Name muss also nur mit einem Gesicht assoziiert werden. Eine mehrfache Zuordnung der Komponenten („one to many mapping“) liegt den klassischen Quellengedächtnisexperimenten zugrunde. Hier werden von wenigen Quellen (z.B. zwei verschiedenen Personen) jeweils viele Items (z.B. 40 unterschiedliche Aussagen) präsentiert. Eine Quelle wird also mit mehreren Aussagen assoziiert.⁹ Auch beim Paar-Assoziations-Lernen, in assoziativen Rekognitionsaufgaben und in Studien zum Fan-Effekt werden bisweilen mehrfache Zuordnungsverhältnisse eingesetzt. Beim Paar-Assoziations-Lernen werden meist Verhältnisse von 2:1 gewählt, ein Wort kommt als

⁹Da in Quellengedächtnisexperimenten und in Experimenten zum Kontextgedächtnis meist viele zentrale Inhalte (Aussagen) mit wenigen Quellen oder Kontexten verknüpft werden müssen, wurde von einigen Autoren vorgeschlagen, die seltenere Information als Iteminformation und die häufigere Information als Kontextinformation anzusehen (z.B. Van Petten, Senkfor & Newberg, 2000). Diese Definition hat aber den Nachteil, dass sie an das mehrfache Zuordnungsverhältnis gekoppelt ist, welches zwar typischerweise verwendet wird, keinesfalls aber notwendigerweise genutzt werden muss.

‚Stimulus‘ im Wortpaar A-B und als ‚Response‘ im Wortpaar C-A vor (z.B. Primoff, 1938; Slamecka, 1976; Young, 1961). In assoziativen Rekognitionsexperimenten werden zum Teil Verhältnisse von 2:1 bis 5:1 verwendet (z.B. Buchler et al., 2008; Verde, 2004). Eine genaue Spezifikation des Zuordnungsverhältnisses wird typischerweise bei Untersuchungen zum Fan-Effekt angegeben (Anderson, 1974; Anderson & Bower, 1973). Hier drückt die Bezeichnung Fan 1-1 eine einfache Zuordnung aus. Jedes Wort wird nur mit einem anderen Wort gepaart.¹⁰ Fan 1-5 meint, dass ein Wort über die gesamte Lernphase hinweg mit fünf anderen Wörtern präsentiert wird; möglich ist auch ein Fan 5-5 bei dem beide Wörter des Wortpaares mit jeweils 5 anderen Wörtern kombiniert werden. Wie durch diesen kurzen Abriss deutlich wird, variieren in unterschiedlichen Forschungsparadigmen die typischen Zuordnungsverhältnisse. Auch den in der Meta-Analyse (Old & Naveh-Benjamin, 2008a) zusammengefassten Studien liegen sehr unterschiedliche Verhältnisse zugrunde.

Der Aspekt des Zuordnungsverhältnisses ist aus theoretischer Sicht interessant, da die Inhibitionsdefizittheorie (Hasher & Zacks, 1988, siehe Kapitel 3.2.4) eine stärkere Interferenzanfälligkeit älterer Menschen vorhersagt. Das assoziative Gedächtnisdefizit älterer Menschen sollte bei mehrfachen Zuordnungsverhältnissen, die eine hohe Interferenz erzeugen, stärker ausgeprägt sein als bei einfachen Zuordnungsverhältnissen mit schwacher Interferenz. Im Hinblick auf verschiedene Interferenzeffekte (z.B. proaktive und retroaktive Interferenz) ist für die Assoziationsdefizit-Theorie das Phänomen der assoziativen Interferenz besonders interessant (Criss & Shiffrin, 2004; Dyne, Humphreys, Bain & Pike, 1990; Verde, 2004). Als assoziative Interferenz wird die spezifische Interferenz von sich überlappenden Wortpaaren bezeichnet. Wenn also die Wortkombinationen A-B, C-D, A-E, F-D gelernt werden müssen, so besteht u.a. Interferenz zwischen den Wortpaaren A-B und A-E sowie zwischen C-D und F-D. Bei jungen Menschen hat sich gezeigt, dass der Effekt der assoziativen Interferenz in Reproduktionsaufgaben zu einer schlechteren Gedächtnisleistung führt. In Rekognitionsaufgaben erhöhen sich vor allem die falschen Alarme. Hinsichtlich der Auswirkungen auf die Trefferraten gibt es unterschiedliche Befunde (für einen Überblick siehe Verde, 2004). Eine vergleichbare Studie mit älteren Menschen liegt m.W. bisher nicht vor.

Versucht man die Vorhersagen der Inhibitionsdefizittheorie auf das Spektrum der möglichen Zuordnungsverhältnisse anzuwenden, so wäre aus meiner Sicht ein umgekehrt U-förmiger Zusammenhang zwischen Zuordnungsverhältnis und Interferenz zu erwarten. Bei einfachen Zuordnungen sollten keine Interferenzeffekte auftreten. Die Interferenzeffekte sollten bei mehrfachen Zuordnungen (ca. 2:1 bis 6:1) ansteigen. Sobald eine Komponente nur noch wenige Ausprägungen annimmt (z.B. 2 unterschiedliche Quellen bei einem Verhältnis von 40:2), sollten die Interferenzeffekte wieder abnehmen. Die Erinnerung an die Quelleninformation ist in letzterem Fall trivial, und es kann mit gutem Grund angenom-

¹⁰In klassischen Experimenten zum Fan-Effekt werden nur selten Wortpaare verwendet. Meistens werden zu einem Wort bzw. einer Person oder einem Objekt unterschiedlich viele Zusatzinformationen präsentiert.

men werden, dass hier trotz mehrfacher Zuordnung keine starken Interferenzbedingungen vorliegen.

Die empirischen Befunde zu diesen Überlegungen sind leider relativ spärlich. Die Hypothese, dass ältere Menschen interferenzanfälliger sind, konnte empirisch sowohl für retroaktive als auch für proaktive Interferenz bestätigt werden (für einen Überblick siehe Kane & Hasher, 1995). Auch in Altersstudien zum Fan-Effekt drückt sich die Interferenzanfälligkeit in längeren Reaktionszeiten bzw. höheren Fehlerraten älterer Menschen aus (Radvansky, Zacks & Hasher, 1996). Leider gibt es bisher nur wenige experimentelle Studien, die die Auswirkungen des Zuordnungsverhältnisses in verschiedenen Altersgruppen untersuchen.

Eine der Studien, die sich diesem Thema empirisch genähert hat, ist die Studie von Schacter, Osowiecki, Kaszniak, Kihlstrom und Valdiserri (1994). Sie haben eine einfache Zuordnung im Quellengedächtnisparadigma getestet und konnten zeigen, dass Altersunterschiede in Quellengedächtnistests auch unter diesen Bedingungen vorliegen. Leider haben sie den Zuordnungsfaktor nicht experimentell manipuliert, sondern ausschließlich die Bedingung einfacher Zuordnung getestet. Daher kann aus dieser Studie lediglich geschlossen werden, dass bei einer einfachen Zuordnung ein Altersunterschied vorhanden ist, nicht aber inwieweit sich der Altersunterschied in Abhängigkeit der Zuordnung verändert.

In einem assoziativen Rekognitionsparadigma, das dem von Naveh-Benjamin verwendeten sehr ähnlich ist, haben Buchler et al. (2008) sowie Cohn und Moscovitch (2007) das Phänomen der assoziativen Interferenz bei *jüngeren* Menschen untersucht. Sie manipulierten dabei, ob eine einfache Zuordnung vorlag oder ob ein Wort mit 2 bis 5 anderen Wörtern präsentiert wurde. Die erhöhte Interferenz zeigte signifikante Effekte. Die Erinnerungsleistung war bei einfachen Zuordnungen besser als bei den Interferenz erzeugenden mehrfachen Zuordnungen. Zu erwarten wäre, dass dieser Effekt bei älteren Menschen noch stärker ausfällt.

Die m.W. einzige Studie, die die Auswirkungen des Zuordnungsverhältnisses bei älteren Menschen manipuliert hat, stammt von Probyn et al. (2007). Sie testeten unter Verwendung eines Reproduktionsparadigmas die Vorhersagen des ‚temporal-context-models‘ (Howard & Kahana, 2002). Die Ergebnisse zeigten, wie erwartet, eine signifikante Interaktion von Alter und Zuordnungsverhältnis. Diese Studie bestätigt damit die zu Beginn des Kapitels postulierte Hypothese.

Auf das Konzept der assoziativen Interferenz greifen auch Studien zurück, die die Listenlänge oder die Ähnlichkeit der Elemente zueinander verändern. Criss und Shiffrin (2004) untersuchten die assoziative Rekognitionsleistung bei Lernlisten, die drei unterschiedliche Arten von Materialien enthielten. Es wurden Paare aus zwei Wörtern, Wort-Gesicht-Kombinationen sowie Kombinationen von zwei Gesichtern dargeboten. Variiert wurde in ihrem Experiment die Listenlänge innerhalb einer der drei Kategorien. In ihrem ersten

Experiment wurden einer Gruppe jeweils 40 Kombinationen aus jeder Kategorie gezeigt; eine andere Gruppe lernte insgesamt ebenfalls 120 Kombinationen, allerdings 80 Wort-Gesicht Paare und jeweils 20 Kombinationen aus zwei Wörtern und zwei Gesichtern. Criss und Shiffrin (2004) konnten in insgesamt drei Experimenten zeigen, dass die Manipulation der Listenlänge einer Kategorie zu einer verschlechterten Diskriminationsleistung für die Items dieser Kategorie führt. Auf die Items einer anderen Kategorie wirkte sich diese Manipulation nicht aus, selbst wenn die Liste insgesamt länger wurde (Criss & Shiffrin, 2004, siehe Experiment 3). Sie erklären diesen Befund durch eine erhöhte assoziative Interferenz mit ansteigender Zahl von Items innerhalb einer Kategorie.

Eine weitere Manipulation, die zu einer erhöhten Interferenz führt, ist die Ähnlichkeit der Items zueinander. In verbalen Aufgaben wird die Ähnlichkeit der Items meist durch die Zugehörigkeit zu einer Kategorie hervorgerufen (für einen Überblick siehe Pantelis et al., 2008). Leider kann dadurch die Ähnlichkeit nur kategorial beeinflusst aber nicht über genaue Parameter spezifiziert werden. Pantelis et al. (2008) wählten daher sowohl artifiziell aussehende als auch natürlich aussehende Gesichter, deren Ähnlichkeit anhand von vier Dimensionen genau variiert werden konnte. Die Ergebnisse zeigten, dass sowohl bei freien Reproduktionsaufgaben (Exp. 1) als auch in assoziativen Rekognitionsaufgaben (Exp. 2 und 3) eine erhöhte Ähnlichkeit von Items innerhalb einer Liste zu einer schlechteren Gedächtnisleistung führte.

Sowohl die Ähnlichkeit der Materialien als auch die Listenlänge innerhalb einer Kategorie beeinflusste die Gedächtnisleistung in assoziativen Rekognitionsaufgaben. Beide Effekte werden genau wie die Effekte des Zuordnungsverhältnisses auf eine erhöhte Interferenz zurückgeführt. Daher wäre zu erwarten, dass ältere Menschen unter den genannten Bedingungen weitaus stärkere Gedächtniseinbußen aufweisen, als jüngere Menschen. Die zuletzt genannten Manipulationen wurden bisher nicht bei älteren Menschen untersucht. Allerdings liegen viele Altersstudien vor, denen unterschiedliche Zuordnungsverhältnisse zugrunde liegen (siehe oben). Diese Studien hätte man in die Meta-Analyse integrieren sollen. Zum einen hätte damit erste Evidenz aufgeführt werden können, inwieweit die erhöhte Interferenz das Assoziationsdefizit älterer Menschen tatsächlich moderiert; zum anderen hätten sich daraus dann weitere Hypothesen für die Manipulation der Listenlänge und der Ähnlichkeit der Items ableiten lassen.

In diesem Abschnitt sollte deutlich geworden sein, dass die Interferenz einer assoziativen Rekognitionsaufgabe über den Zuordnungsquotienten variiert werden kann und sich daraus unterschiedliche Vorhersagen für die Alterseffekte ergeben. Zukünftige Experimente zur Auswirkung des Zuordnungsquotienten könnten zu einer Generalisierung des assoziativen Defizits beitragen und das Zusammenspiel von Interferenzanfälligkeit und assoziativem Defizit im Alter genauer beleuchten.

(2) Integrationspotential

Eine zweite Klassifikationsdimension, die in der Meta-Analyse nicht aufgegriffen wurde, die in Bezug auf das Ausmaß der Alterseffekte aber wichtig erscheint, ist das Integrationspotential der beiden Komponenten. Der Begriff Integrationspotential bezieht sich darauf, wie leicht sich die beiden Komponenten zu einer Einheit integrieren lassen (,unitization‘). Ein hohes Integrationspotential liegt z.B. bei zwei einzelnen Wörtern vor, die als Komposita zusammengesetzt werden können. Die Elemente ,Dach‘ und ,Fenster‘ lassen sich zu einer Einheit ,Dachfenster‘ verbinden. Auch einzelne Gesichtskomponenten wie z.B. Augenpartie oder Mundpartie werden bei der Darbietung von Gesichtern als Einheit wahrgenommen. Auf der anderen Seite weisen schwach assoziierte Wörter eine geringe Integrationsmöglichkeit auf (Politik, Schublade). Auch zwei Gesichter, die miteinander assoziiert werden sollen, können kaum integriert werden.

In Anlehnung an das Konzept des Integrationspotentials wird in einigen Studien (Diana et al., 2008; Jäger et al., 2006; Naveh-Benjamin, 2006) die Unterscheidung zwischen ,Intra-Item‘-Assoziationen und ,Inter-Item‘-Assoziationen getroffen¹¹. Ein Beispiel für eine Inter-Item-Assoziation wären schwach assoziierte Wortpaare wie ,Politik - Schublade‘ sowie zwei zu assoziierende Gesichter. Hier muss eine Assoziation zwischen (=inter) zwei Items hergestellt werden. Die Assoziation von einem Wort und dessen Schriftart, Schriftfarbe oder Darbietungsmodalität wäre hingegen eine Intra-Item-Assoziation, da die Darbietungsmodalität direkt mit dem Item verbunden ist und die Komponenten nur schwer isoliert voneinander präsentiert werden können. In Bezug auf das Integrationspotential wird angenommen, dass ,Intra-Item‘-Assoziationen ein höheres Integrationspotential aufweisen als ,Inter-Item‘-Assoziationen.

Die genannten Beispiele könnten als Abstufungen eines Kontinuums gesehen werden, auf dem sich verschiedene Stufen des Integrationspotentials unterscheiden lassen: (1) Assoziationen mit hoher Integrationsmöglichkeit; (2) Intra-Item-Assoziationen; (3) Inter-Item-Assoziationen mit geringer Integrationsmöglichkeit.

Dem Konzept des Integrationspotentials wird an dieser Stelle deshalb so viel Bedeutung beigemessen, da zu vermuten ist, dass das assoziative Altersdefizit wesentlich vom Integrationspotential des Materials abhängt. Um die Effekte des Integrationspotentials zu erklären bzw. Hypothesen für die Altersunterschiede zu generieren, wird die unterschiedliche Beteiligung von bewusster Erinnerung und Vertrautheit am Erinnerungsprozess herangezogen. Bekannt ist, dass Prozesse der bewussten Erinnerung starken Alterseffekten unterliegen, wohingegen Vertrautheitsprozesse kaum vom kognitiven Altern betroffen sind (Light, Prull, La Voie & Healy, 2000; Zacks et al., 2000). Zudem wird postuliert, dass der assoziativen Rekognition andere Gedächtnisprozesse zugrunde liegen als der Itemre-

¹¹Zum Teil wird auch die alternative Bezeichnung ,Inter-unit‘- und ,Intra-unit-binding‘ verwendet. Ecker et al. (2007) sprechen auch von intrinsischen und extrinsischen Bindungsprozessen.

kognition (Hockley & Consoli, 1999; Kelley & Wixted, 2001; Rotello & Heit, 1999, 2000; Yonelinas, 1997).

Bei der Rekognition von einzelnen Items können sowohl Vertrautheit als auch bewusste Erinnerung beteiligt sein. Für assoziative Rekognitionsaufgaben, bei denen in der Testphase nur alte Items in unterschiedlichen Kombinationen getestet werden, ist dahingegen bewusste Erinnerung notwendig und Vertrautheit kann den assoziativen Rekognitionsprozess nicht unterstützen (Hockley & Consoli, 1999; Kelley & Wixted, 2001; Rotello & Heit, 1999, 2000; Yonelinas, 1997).

Aus diesen Ausführungen ergibt sich die Schlussfolgerung, dass die Leistung älterer Menschen in Itemrekognitionsaufgaben besser ausfallen sollte als in assoziativen Rekognitionsaufgaben, denn bei der Erinnerung an Items können Senioren auf Vertrautheit zurückgreifen. Dies ist für die assoziative Aufgabe nicht möglich.

Diese Annahme wird unter anderem von M. R. Healy et al. (2005) untermauert. Sie analysierten assoziative Rekognitionsdaten von älteren und jüngeren Menschen unter der Verwendung unterschiedlicher Rekognitionsmodelle (siehe Macho, 2004). Die Daten konnten am besten von den Modellen erklärt werden, die Prozesse der bewussten Erinnerung *und* Vertrautheit enthalten. Dabei zeigte sich bei älteren Menschen unabhängig vom verwendeten Modell eine herabgesetzte bewusste Erinnerung für Wortkombinationen (für vergleichbare Ergebnisse unter Verwendung des ‚Remember-Know-Guess-Paradigmas‘ siehe D’Argembeau & Van der Linden, 2004). Hinsichtlich der Vertrautheitsprozesse scheinen sich ältere Menschen von der Itemvertrautheit fehlleiten zu lassen. Allgemein finden sich bei beiden Altersgruppen vergleichbare Trefferraten; die falschen Alarme sind allerdings bei älteren Menschen signifikant erhöht (Light et al., 2004; M. G. Rhodes et al., 2008). Werden nun einige Wortpaare wiederholt präsentiert, so führt dies zu einem weiteren Anstieg der falschen Alarme bei älteren Menschen. Ältere Menschen beurteilen ein neu kombiniertes Wortpaar aus zwei mehrfach dargebotenen Elementen häufig als altes Wortpaar. Dies gilt nicht nur für Wortpaare, sondern konnte auch für Gesichterkombinationen repliziert werden (M. G. Rhodes et al., 2008) und zeigt klar, dass ältere Menschen sich durch die Itemvertrautheit beeinflussen lassen und die bewussten Erinnerungsprozesse zu schwach sind, um dem Vertrautheitsgefühl entgegenzuwirken.

In den letzten Jahren wurde die Annahme, dass für die Erinnerung an Wortkombinationen keine Vertrautheit genutzt werden kann, in Frage gestellt (z.B. Yonelinas et al., 1999). Einige Forscher postulierten, dass die Beteiligung von Vertrautheit bei assoziativen Rekognitionsprozessen vom Itemmaterial abhängig ist. Mehrere Studien konnten anhand unterschiedlicher Methoden aufzeigen, dass bei Materialien mit hohem Integrationspotential sowohl Vertrautheit als auch bewusste Erinnerung zusammenwirken (Ecker et al., 2007; Giovanello et al., 2006; Graf & Schacter, 1989; Jäger et al., 2006; Yonelinas et al., 1999). Bei Materialien mit geringem Integrationspotential hingegen stützt sich der Abrufprozess allein auf bewusste Erinnerung. Kombiniert man diese Befunde mit den

Befunden aus der Altersforschung, ergibt sich folgende Vorhersage: Stabile Alterseffekte sollten sich bei Inter-Item-Assoziationen finden, die schwer zu integrieren sind. Unter Verwendung von Material mit hohem Integrationspotential sollten sich keine oder nur geringe Altersunterschiede zeigen.¹² Unter welchen Bedingungen aber liegt ein hohes bzw. niedriges Integrationspotential vor? Unter welchen Bedingungen sollten sich eher geringe bzw. stark ausgeprägte Altersunterschiede zeigen? Bevor diese Fragen anhand von einigen exemplarischen Studien beantwortet werden, wird zunächst ein kurzer Überblick über die verschiedenen Herangehensweisen an diese Fragestellung gegeben.

Um aufzuzeigen, ob allein bewusste Erinnerung oder bewusste Erinnerung und Vertrautheit gemeinsam dem Erinnerungsprozess zugrunde liegen, können unterschiedliche methodische Zugänge gewählt werden. Einige Studien (Giovanello et al., 2006; Quamme, Yonelinas & Norman, 2007) untersuchten Patienten mit hippocampalen Läsionen (1). Von diesen Patienten ist bekannt, dass sie zwar über intakte Vertrautheitsprozesse verfügen, sich aber nicht auf bewusste Erinnerung stützen können. Studien aus der Arbeitsgruppe von Yonelinas (Diana et al., 2008; Yonelinas et al., 1999) haben den Verlauf von ROC-Kurven herangezogen (2), um auf die Beteiligung von bewusster Erinnerung und Vertrautheit zu schließen. Zu beachten ist bei diesem Vorgehen, dass es an bestimmte Modellannahmen gebunden ist. Dasselbe gilt auch für die Verwendung des Remember-Know-Paradigmas (3), das von Giovanello et al. (2006) eingesetzt wurde. Graf und Schacter (1989) verglichen die Leistung in impliziten und expliziten Gedächtnistests (4), um auf die Beteiligung von Vertrautheit zu schließen (für einen Überblick siehe auch Dew et al., 2007). Eine größere Gruppe von Forschern (Jäger et al., 2006; S. M. Rhodes & Donaldson, 2007) verwendete ereigniskorrelierte Potentiale (EKP) (5), um Vertrautheit von bewusster Erinnerung zu trennen. Auf die Details dieser Methoden und die kritischen methodischen Aspekte soll hier nicht genauer eingegangen werden. Hinsichtlich des Einsatzes dieser Methoden in der Frage des Integrationspotentials ist es wichtig festzuhalten, dass vergleichbare Ergebnisse unter der Verwendung von unterschiedlichen Methoden berichtet wurden. Dies stützt die Schlussfolgerung, dass die Ergebnisse nicht auf methodischen Artefakten beruhen und daher auf eine detaillierte Methodendiskussion an dieser Stelle verzichtet werden kann.

Die Studien unterscheiden sich nicht nur in ihrer Methodik, sondern auch darin, wie das Integrationspotential manipuliert wird. Graf und Schacter (1989) postulieren, dass die Integration zweier Komponenten unter zwei Bedingungen erfolgen kann: Entweder werden die beiden Komponenten direkt und spontan integriert, d.h. das Integrationspotential beruht auf prä-experimentellen Eigenschaften des Itemmaterials. Oder aber die Integration

¹²Dieser Tatsache wurde auch in der Definition des assoziativen Defizits Rechnung getragen. Denn dort wird das Defizit spezifiziert als „deficiency in creating and retrieving links between *single units* of information“. Der Extremfall des perfekten Integrationspotentials, dass also zwei Komponenten als Einheit wahrgenommen werden (Gesichtserkennung), ist in der Definition des Assoziationsdefizits also streng genommen gar nicht vorgesehen.

muss erzeugt werden, z.B. durch die Verwendung integrativer Enkodierungsstrategien. Zunächst werden einige Beispiele für prä-experimentelle Assoziationen berichtet, dann wird darauf eingegangen, welche integrativen Enkodierungsstrategien bisher untersucht wurden.

Giovanello et al. (2006) haben Patienten mit hippocampalen Läsionen zum einen Komposita (land-scape; black-mail; jail-bird) dargeboten, zum anderen unrelatierte Wortpaare (telefone-trumpet). Im assoziativen Rekognitionstest mussten intakte Wortpaare (land-scape) von neu kombinierten Wortpaaren (black-bird) unterschieden werden. Die Patientengruppe zeigte eine schlechte Gedächtnisleistung für die unrelatierten Wortpaare, wie bei hippocampalen Läsionen zu erwarten. Ihre Gedächtnisleistung für Komposita dahingegen war signifikant besser. Die gleichaltrige Kontrollgruppe war in beiden Bedingungen gleich erfolgreich. Diese Studie zeigt, dass Komposita ein hohes prä-experimentelles Integrationspotential aufweisen und daher Vertrautheit am Erinnerungsprozess dieser Wortpaare beteiligt ist.

S. M. Rhodes und Donaldson (2007) untersuchten Wortpaare, die (a) entweder hoch assoziiert waren, (b) einen semantischen Bezug aufwiesen oder (c) sowohl assoziiert als auch semantisch relatiert waren. Sowohl für die semantische Zusammengehörigkeit als auch für die Assoziationen wurden Normen als objektive Maße verwendet. Die assoziierten Wortpaare (a) waren dabei nicht nur miteinander assoziiert, sondern konnten auch zu Komposita zusammengesetzt werden (z.B. traffic-jam; spare-tire). Die semantisch relatierten Wortpaare (b) gehörten einer semantischen Kategorie an, ohne aber hoch miteinander assoziiert zu sein (z.B. lemon-orange; needle-thread). Wortpaare, die sowohl semantisch als auch assoziativ verbunden waren (c), gehörten einer Kategorie an und wiesen dieselbe Assoziationsstärke auf wie die assoziierten Wortpaare (z.B. cereal-bread; fork-plate). Aus den abgeleiteten EKP-Maßen kann die Schlussfolgerung gezogen werden, dass nur bei der Kategorie der assoziierten Wortpaare Vertrautheit am Erinnerungsprozess beteiligt ist. Bei keiner der beiden anderen Bedingungen konnte ein frühes (300-500msec) frontales Signal abgeleitet werden, das als Korrelat von Vertrautheit angesehen wird (Mecklinger, 2000). Die EKP-Signale, die bewusste Erinnerung anzeigen (spätes 400-800msec parietales Signal), wurden in vergleichbarem Ausmaß bei allen drei Wortkombinationen gefunden. Semantischer Bezug zwischen Wortpaaren scheint also keine ausreichende Bedingung zu sein, um Vertrautheit hervorzurufen; Komposita dahingegen rufen Vertrautheit hervor.

Abgesehen vom Wortmaterial können auch Gesichter als Stimulusmaterial genutzt werden. Yonelinas et al. (1999) konnten am Verlauf von ROC-Kurven zeigen, dass die Gesichtskomponenten wie Nase, Augenpartie und Mund in einem aufrecht präsentierten Gesicht als eine Einheit verarbeitet werden und Vertrautheit bei der Erinnerung eine wesentliche Rolle spielt. Werden die Gesichter dahingegen um 180 Grad gedreht, so können die Komponenten nur anhand von bewusster Erinnerung rekonstruiert werden. Jäger und Mecklinger präsentierten ihren Versuchspersonen entweder zwei Gesichter von unterschiedlichen Per-

sonen oder aber zwei unterschiedliche Darstellungen des Gesichts einer Person. Sie konnten zum einen anhand des Verlaufs von ROC-Kurven (Jäger & Mecklinger, 2009) zum anderen anhand von EKP-Signalen (Jäger et al., 2006), zeigen, dass bei der Erinnerung an die Gesichter zweier Personen keine Vertrautheit, sondern nur bewusste Erinnerung beteiligt ist. Die Erinnerung an die Gesichtsdarstellungen einer Person wird dahingegen sowohl von Vertrautheit als auch von bewusster Erinnerung unterstützt.

Zusammenfassend liegt also bei Komposita, bei unterschiedlichen Gesichtsabbildungen einer Person sowie bei Gesichtspartien innerhalb eines Gesichts ein prä-experimentelles Integrationspotential vor, das hoch genug ist, um Vertrautheit hervorzurufen. Eine wichtige Frage ist m.E., ob diese Stimuli überhaupt noch als Kombination von einzelnen Elementen angesehen werden können, oder ob diese nicht eher einheitlich wahrgenommene Stimuli sind. Dieser Aspekt wird in der Literatur nur an wenigen Stellen diskutiert (z.B. S. M. Rhodes & Donaldson, 2007, S. 421). Zumeist wird dann auf die zentrale Rolle der Testsituation hingewiesen, da hier intakte Kombinationen und aus Teilelementen zusammengesetzte neue Kombinationen differenziert werden müssen. Selbst wenn man diese Argumente nicht für überzeugend hält und eher den Standpunkt vertritt, dass es sich bei Komposita *nicht* um integrierte Einzelelemente handelt, so können die aufgeführten Studien als extreme Endpunkte der Integrationsskala gesehen werden, und es sollte leicht vorstellbar sein, wie schwächere Skalenabstufungen gewählt werden könnten.

Eine schwächere Form der Integration besteht zum Beispiel darin, integrative Enkodierungsstrategien auf prä-experimentell unrelatiertes Wortmaterial anzuwenden. Diesen Ansatz testeten Quamme et al. (2007) an einer Patientengruppe mit hippocampalen Läsionen (vgl. Giovanello et al., 2006). Die Probanden erhielten unterschiedliche Enkodierungsstrategien für Wortpaare. Manche Wörter sollten sie in einen Satz integrieren, z.B. ‚Der HIMMEL kann von der BANK aus gesehen werden‘. Bei anderen Wortpaaren sollten die Wörter genutzt werden, um ein neues Kompositum zu definieren. Bei den Wörtern ‚Himmel‘ und ‚Bank‘ wäre die Definition für das Kompositum ‚Himmelsbank‘ eine Bank, von der aus man in den Himmel schauen kann. Es zeigte sich, dass die Gedächtnisleistung der Patienten für die Wörter der Satzbedingung signifikant schlechter war als die Gedächtnisleistung für die Komposita-Definitionen. Diese Manipulation scheint also besser geeignet zu sein, um zwei Wörter miteinander zu integrieren.

S. M. Rhodes und Donaldson (2008) legten ihren Probanden nahe, sich entweder jedes Wort einzeln bildlich vorzustellen, oder aber beide Wörter in ein interaktives Bild zu integrieren. Nur in der interaktiven Vorstellungsbedingung zeigte die Ableitung der EKP-Signale die typischen frühen frontalen Signale, die als Korrelat von Vertrautheitsprozessen interpretiert werden. Ebenfalls bildliche Vorstellungen nutzten Diana et al. (2008, Exp. 2). Ihren Probanden wurden Wörter in den Farben Rot und Grün dargeboten. In der interaktiven Vorstellungsbedingung sollten sich die Probanden das jeweilige Wort in der entsprechenden Farbe vorstellen und einen Grund dafür überlegen, warum das Wort

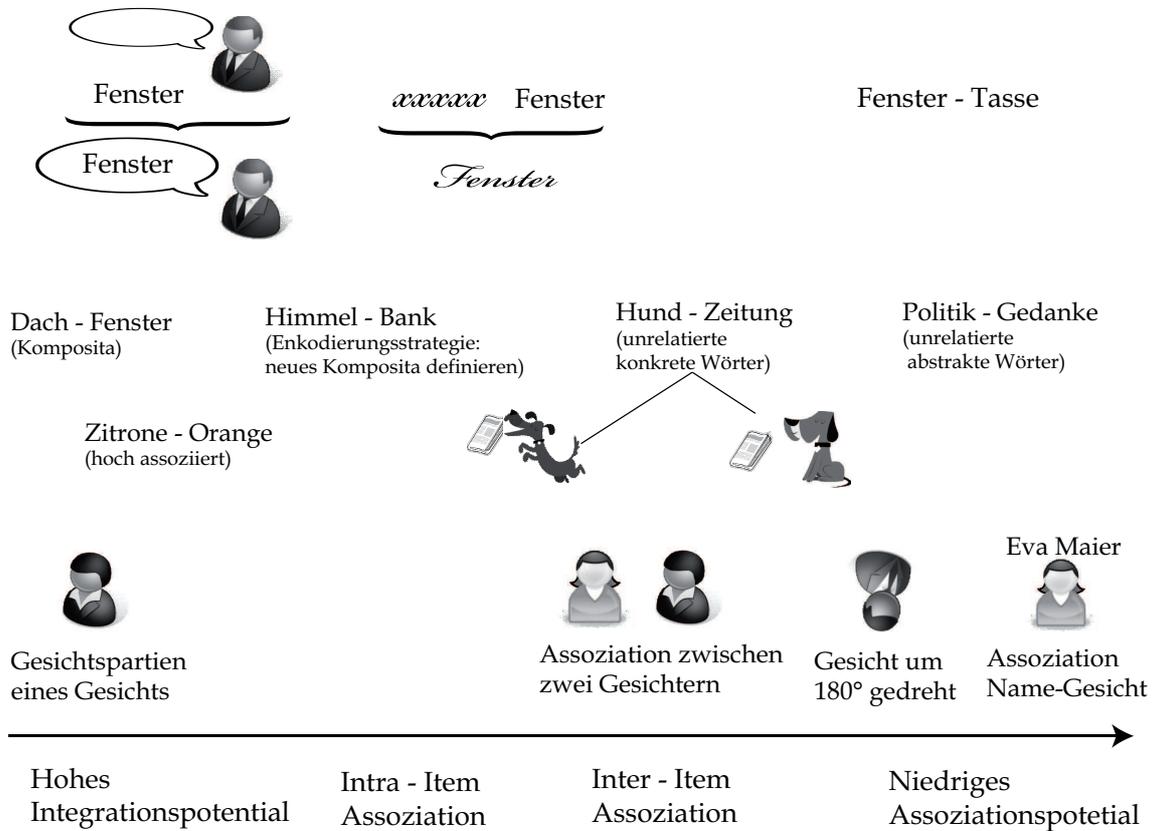


Abbildung 3.4: In dieser Abbildung sind unterschiedliche Materialien bzw. unterschiedliche Enkodierungsstrategien dargestellt, die sich hinsichtlich ihres Integrationspotentials unterscheiden lassen. Ganz links sind Materialien wie z.B. Komposita abgebildet, die ein sehr hohes Integrationspotential aufweisen. Ganz rechts werden Materialien dargestellt, die ein niedriges Integrationspotential aufweisen, z.B. abstrakte, nicht assoziierte Wörter.

diese Farbe annimmt. Das heißt, sie überlegten sich z.B., dass der Tisch grün ist, da er mit Moos bewachsen ist. In einer nicht interaktiven Vorstellungsbedingung sollten sich die Probanden das jeweilige Wort entweder mit einem grünen Dollarschein vorstellen (auf dem Tisch liegen grüne Dollarscheine) oder mit einem roten Stoppschild. Die Analyse von ROC-Kurven zeigte, dass unter der integrativen Vorstellungsbedingung Vertrautheit und bewusste Erinnerung zur Gedächtnisleistung beitragen. In einer weiteren Studie von Opitz und Cornell (2006) wurden den Teilnehmern jeweils vier Wörter vorgelegt, und sie sollten in einer integrativen Enkodierungsbedingung entscheiden, welches Wort nicht zu den anderen dreien passt. In einer vergleichenden Bedingung sollten sie angeben, welches Wort die kleinste physikalische Erscheinung aufweist. Die Analyse von EKP-Signalen zeigte wiederum eine Beteiligung von Vertrautheit in der integrativen Bedingung an, nicht aber in der vergleichenden Enkodierungsbedingung.

In den berichteten Studien wurden meist integrative Enkodierungsbedingungen mit nicht-integrativen Bedingungen verglichen. Interessant wäre es, in zukünftigen Studien

unterschiedliche Abstufungen der Enkodierungsbedingungen miteinander zu vergleichen, d.h. zu untersuchen, ob integrative Vorstellungen und integrative Sätze zu einer vergleichbaren Beteiligung von Vertrautheit führen, oder ob die Erzeugung von Definitionen für neue Komposita zu einer höheren Integration führt als die integrativen Sätze. Da sich im Moment einige Forschergruppen intensiv mit diesem Thema zu befassen scheinen, sind solche Erkenntnisgewinne wahrscheinlich bald zu erwarten.

In diesem Vergleich von hauptsächlich zehn Studien, die in den letzten zwei Jahren erschienen sind, konnte gezeigt werden, unter welchen Bedingungen eine hohe prä-experimentelle Integration vorliegt, und welche Enkodierungsstrategien geeignet sind, um eine integrative Verarbeitung hervorzurufen. In der Anwendung auf Altersstudien wäre zu erwarten, dass unter den Bedingungen einer hohen Integration geringere Altersunterschiede vorliegen, da ältere Menschen in der Lage sind, Vertrautheit bei der Erinnerung zu nutzen.

Leider wurde bisher keine dieser ausgeführten Manipulationen in assoziativen Rekognitionstests an älteren Menschen getestet. Inwieweit sich die Annahmen also bestätigen lassen, ist eine offene Forschungsfrage. Die einzige Untersuchung, die das Integrationspotential des Wortmaterials im Altersvergleich untersucht, ist Experiment 4 von Naveh-Benjamin (2000) bzw. die Replikation dieses Experiments durch Naveh-Benjamin, Hussain, Guez und Bar-On (2003). Hier wurden Wortpaare dargeboten, die entweder aus einer semantischen Kategorie stammen, oder aber keinen Bezug zueinander aufweisen. Im assoziativen Rekognitionstest war bei nicht relatierten Wortpaaren ein signifikanter Altersunterschied zu erkennen. Bei relatierten Wortpaaren fand sich kein signifikanter Unterschied zwischen den beiden Altersgruppen. Dieser Befund ist insbesondere im Vergleich zu den Experimenten von S. M. Rhodes und Donaldson (2007) interessant, da sie kein EKP-Korrelat von Vertrautheit bei semantisch relatierten Wortpaaren fanden. Bevor dies als Diskrepanz interpretiert wird, muss auf einige Unterschiede zwischen den Studien hingewiesen werden. Zum einen verwendete Naveh-Benjamin doppelt so lange Präsentationszeiten bei älteren Menschen als bei jüngeren Menschen. Es könnte also der Fall sein, dass die Altersunterschiede durch die verschieden langen Lernzeiten schon abgeschwächt werden und ältere Menschen diese Zeit nutzen, um elaborierte Enkodierungsstrategien für das semantisch relationierte Material anzuwenden. Da außerdem bei Rhodes und Donaldson ein anderes Testverfahren verwendet wurde, ist der direkte Vergleich der beiden Studien erschwert.

Das Experiment von Naveh-Benjamin liefert in jedem Fall erste Hinweise darauf, dass das Integrationspotential der beiden Stimuli ein moderierender Faktor im Altersvergleich darstellt. Neben dieser Studie könnte nur der zwischenexperimentelle Vergleich von Effektgrößen aus Experimenten mit eher hohem vs. eher niedrigem Integrationspotential herangezogen werden, um aus der bisherigen Forschung unterstützende Befunde herzuleiten (siehe Abbildung 3.4).

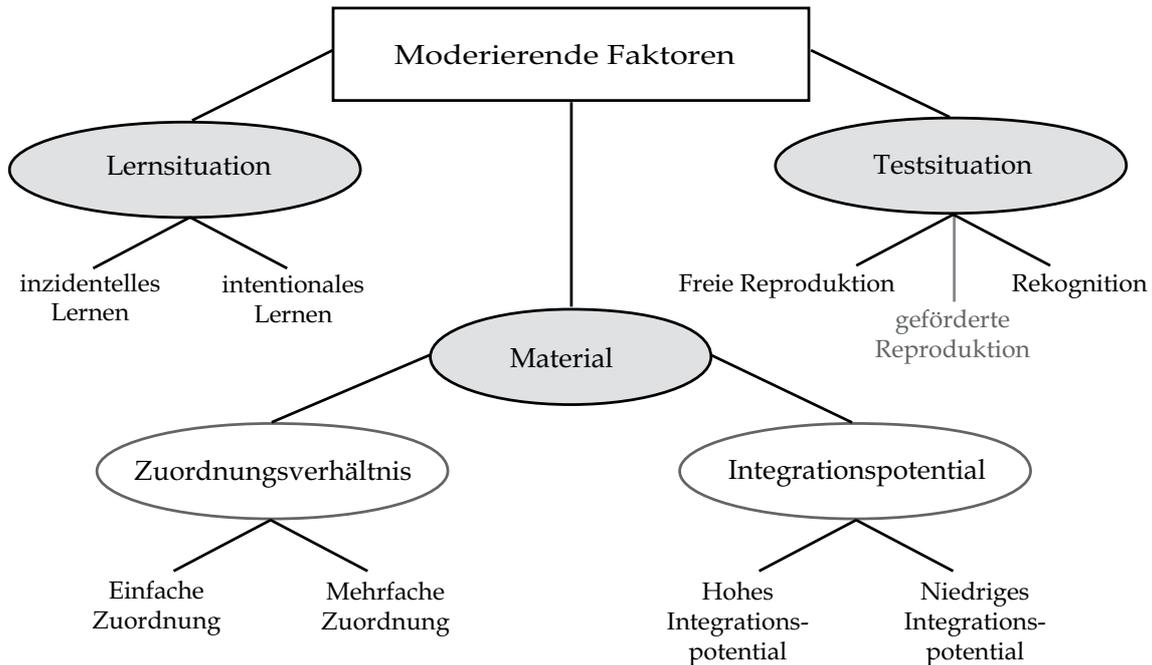


Abbildung 3.5: In dieser Abbildung ist dargestellt, welche Faktoren einen moderierenden Einfluss auf das Assoziationsdefizit haben. Zum einen beeinflussen Charakteristika der Lern- und Testsituation, zum anderen aber auch das Integrationspotential und das Zuordnungsverhältnis der Materialien das Ausmaß des assoziativen Defizits.

Wie in diesem Abschnitt ausgeführt wurde, ist das Integrationspotential als wichtiger Moderator des assoziativen Altersdefizits einzustufen. Die Meta-Analyse von Old und Naveh-Benjamin (2008a) hätte daher sowohl den Faktor Integrationspotential als auch den zuvor diskutierten Faktor Zuordnungsverhältnis aufnehmen sollen, um die Auswirkungen dieser theoretisch fundierten Moderatoren zu untersuchen.

Zusammenfassend können die in diesem Kapitel aufgeführten empirischen Befunde klar das Assoziationsdefizit älterer Menschen unterstützen, die Meta-Analyse von Old und Naveh-Benjamin quantifiziert das Defizit über 90 Experimente hinweg mit einem Altersunterschied von 0.92, wohingegen im Altersvergleich des Itemgedächtnisses der Altersunterschied nur bei 0.72 liegt. Hinsichtlich des Testformats stellten sich Rekognitionstests als besonders sensitiv heraus und sollten aus vielen diskutierten Gründen Reproduktionstests vorgezogen werden. Moderierende Faktoren, die nicht in der Meta-Analyse untersucht wurden, aber durch den Vergleich einzelner Studien und aufgrund ihrer theoretische Verankerung sinnvoll erscheinen, sind das Zuordnungsverhältnis zwischen den Elementen und das Integrationspotential der Komponenten. Abbildung 3.5 zeigt, welche Faktoren einen moderierenden Einfluss auf das assoziative Defizit im Alter haben.

3.2.3 Qualitatives Defizit oder Basismechanismus

Nachdem im letzten Kapitel die empirische Evidenz des Assoziationsdefizits aufgezeigt wurde, sollen in diesem Kapitel die Ursachen dieses Altersdefizits genauer beleuchtet werden. Hierfür kommen im Sinne der theoretischen Perspektiven nach Salthouse (1991) vor allem qualitative Verarbeitungsunterschiede (Kausler, 1994) sowie veränderte kognitive Basismechanismen in Frage (Salthouse, 1995). Auch Naveh-Benjamin und Kollegen diskutieren eine multikausale Verursachung des assoziativen Defizits. Sie sprechen von einem ‚multiprocess account‘, bei dem sowohl strategische als auch automatische Bindungsprozesse zu berücksichtigen sind („strategic binding and automatic binding“, Kilb & Naveh-Benjamin, 2007, S. 1171). In diesem Kapitel sollen qualitative Unterschiede der Verarbeitung fokussiert werden. Konkret soll dabei der Frage nachgegangen werden, ob das Assoziationsdefizit allein auf einem strategischen Defizit älterer Menschen beruht. Im nächsten Kapitel wird dann auf die automatischen Bindungsprozesse bzw. Basismechanismen eingegangen.

Zunächst wird ein kurzer Überblick über Strategien gegeben, die sich auf assoziative Rekognitionsaufgaben und das entsprechende Wortmaterial anwenden lassen. Dann wird spezifiziert, in welchen Phasen der Strategieanwendung es zu Defiziten kommen kann. Schlussendlich werden empirische Studien berichtet, die aufzeigen, welche strategischen Anteile an dem Assoziationsdefizit älterer Menschen beteiligt sind und inwieweit diese das assoziative Defizit älterer Menschen erklären können.

Strategien für assoziative Aufgaben

Schon 1918 wurde untersucht, welche Strategien Versuchspersonen spontan anwenden, um verschiedene Arten von verbalem Material für einen Paar-Assoziationstest zu lernen (Reed, 1918a, 1918b). Als Stimuli dienten dabei zum einen Wortpaare, zum anderen auch Paare von Nonwörtern sowie Kombinationen aus einem bekannten Wort und einem unbekanntem Wort wie z.B. einem Deutsch-Spanischen Wortpaar (Martin, Boersma & Cox, 1965; Montague & Wearing, 1967; Reed, 1918a, 1918b). Im Folgenden sollen die Ergebnisse eingegrenzt werden auf das Stimulusmaterial der Wortpaare, da die anderen Materialarten im Kontext der Assoziationsdefizit-Hypothese eher selten untersucht wurden (siehe aber Shing et al., 2008).

Von Richardson (1998) wurden Strategien, die zum Lernen von Wortpaaren eingesetzt werden, grob in drei Kategorien unterteilt: einfache Wiederholung („rote rehearsal“), verbale Strategien und Strategien der visuellen Vorstellung. Unter verbale Strategien fällt z.B. das Generieren von Sätzen, die beide Wörter enthalten, der Gebrauch von Verbindungswörtern (Mediatoren) sowie die Nutzung von gemeinsamen semantischen Merkmalen der beiden Wörter (Reed, 1918a, 1918b; Richardson, 1998). In nachfolgenden Studien stellte sich heraus, dass verbale Strategien vor allem bei abstraktem Wortmaterial Verwen-

dung finden; bei konkretem Wortmaterial hingegen wird häufiger auf mentale Vorstellung zurückgegriffen (Paivio, Yuille & Smythe, 1966; Richardson, 1998). Nicht nur die Eigenschaften der einzelnen Wörter, sondern auch Merkmale des Wortpaares (räumlicher vs. kategorialer Bezug zwischen den Wörtern) beeinflussen die Wahl der Enkodierungsstrategie (McDougall & Velmans, 1993). Bei einem räumlichen Bezug zwischen den Wörtern wird eher auf eine Vorstellungsstrategie zurückgegriffen, bei kategorialem Bezug werden verbale Mediatoren generiert. Die Effizienz der Strategien ist darüber hinaus abhängig davon, inwieweit die Materialien interaktiv miteinander verknüpft werden (Begg, 1978; McDaniel & Kearney, 1984). Eine hohe Interaktion liegt z.B. vor, wenn die Wörter ‚Hund‘ und ‚Zeitung‘ mit dem Satz ‚Der Hund fängt die Zeitung‘ verbunden werden anstatt mit dem wenig interaktiven Satz ‚Der Hund sitzt neben der Zeitung‘. Analog gilt dies auch für die bildliche Vorstellung.

Bereits Reed (1918a, 1918b) interessierte sich nicht nur dafür, *ob* Strategien verwendet werden, sondern spezifischer, ob diese sowohl während der Enkodierungsphase als auch in der Abrufphase genutzt werden. Warum eine solche spezifische Analyse der vorgestellten Strategien von Bedeutung ist, wird im nächsten Abschnitt erläutert.

Phasen der Strategieanwendung

Um Altersunterschiede in der Verwendung der genannten Strategien untersuchen zu können, muss zunächst festgestellt werden, dass diese nicht nur global analysiert werden können (für einen Überblick siehe Rogers, Hertzog & Fisk, 2000), sondern im Hinblick auf spezifische Phasen des Lern-Testdurchgangs. Eine übersichtliche Phasen-Klassifikation wurde von Dunlosky, Hertzog und Powell-Moman (2005) aufgestellt. Sie unterscheiden insgesamt fünf Phasen der Strategieanwendung, wobei Altersunterschiede in allen Phasen auftreten könnten. Das spontane Generieren einer angemessenen Strategie während der Lernphase kann als erste Phase angesehen werden. Hier zeigen sich Altersunterschiede, falls ältere Menschen spontan seltener Strategien anwenden als jüngere Menschen. Ein solches *Produktionsdefizit* (1) kann aufgehoben werden, indem die Teilnehmer über Strategien in Kenntnis gesetzt werden und die Anwendung von Strategien empfohlen wird. Falls ältere Menschen wirklich nur ein Produktionsdefizit aufweisen, sollten die Altersunterschiede nach einer solchen Instruktion eliminiert sein. Bestehen weiterhin Altersunterschiede könnte ein *Mediationsdefizit* (2) vorliegen. Darunter wird verstanden, dass ältere Menschen eine Strategie nicht anwenden können, weil sie z.B. zu komplex ist und die kognitiven Kapazitäten nicht ausreichen, um in der kurzen Präsentationszeit diese Strategie zu nutzen. Auch bei kleinen Kindern wird ein solches Mediationsdefizit diskutiert, wenn eine Strategie genutzt werden soll, die das Kind überfordert, da es noch nicht die kognitive Reife für diese Strategie erreicht hat (P. H. Miller, 2000; Waters, 2000). Im Falle eines Mediationsdefizits wären die Teilnehmer also nicht in der Lage, die Strategie zu verwenden, obwohl sie auf die Strategie hingewiesen werden. Wird eine Strategie zwar genutzt, führt aber bei älteren Menschen

zu einem geringeren Leistungszuwachs als bei jüngeren Menschen, so spricht man von einem *Nutzungsdefizit* (3). Ein Nutzungsdefizit kann z.B. dadurch entstehen, dass ältere Menschen zwar die gleiche Anzahl von Sätzen generieren, aber seltener interaktive Sätze verwenden als jüngere Menschen. Selbst wenn ältere Menschen spontan hoch interaktive Sätze generieren, d.h. kein Nutzungsdefizit vorliegt, kann es der Fall sein, dass diese Sätze in der Testphase von älteren Menschen seltener abgerufen werden. In diesem Fall liegt ein Abrufdefizit vor (4). Auch der erfolgreiche Abruf des generierten Hinweisreizes stellt die Anwendung der Strategie noch nicht sicher. An dieser Stelle kann ein Dekodierungsproblem (5) auftreten, welches dazu führt, dass der erinnerte Hinweisreiz falsch angewandt wird und daher nicht das richtige Zielwort wiedergegeben wird.

Die empirische Trennung zwischen diesen Defiziten ist anspruchsvoll, insbesondere da die Erfassung der Strategieverwendung über subjektive Selbstberichte erfolgen muss. Diese Methode setzt die Zuverlässigkeit von Selbstberichten voraus, und es besteht die Gefahr, dass allein durch deren Einsatz bereits eine Veränderung der Leistung hervorgerufen wird (siehe Dunlosky et al., 2005). Selbstberichte über die genutzten Strategien sind unerlässlich, um zwischen einem Mediationsdefizit und Nutzungsdefizit zu trennen. Bei einem Mediationsdefizit sollte keine erfolgreiche Verwendung der Strategie berichtet werden, bei einem Nutzungsdefizit dahingegen schon. Für die Trennung von Nutzungsdefizit und Abrufdefizit ist für jedes Item eine zweimalige Strategieerfassung sowohl während der Lernphase als auch während der Testphase erforderlich (Dunlosky & Hertzog, 2001). Darüber hinaus müssen diese Daten über die Strategieverwendung mit Leistungsdaten der Gedächtnisaufgabe in Verbindung gesetzt werden, um zwischen Abruf- und Dekodierungsdefizit zu differenzieren. Dunlosky und Hertzog (1998) diskutieren die Methode der itemspezifischen Strategieerfassung während Lern- und Testphase und weisen auf Vor- sowie Nachteile hin.

Allgemeine empirische Befunde zur Strategienutzung älterer Menschen

Im Folgenden werden die empirischen Befunde zu den strategischen Defiziten älterer Menschen erläutert. Die meisten der folgenden Studien verwendeten Wortpaare als Lernmaterial und erfassten die Gedächtnisleistung über geförderte Reproduktionstests.

Ältere Menschen berichten seltener die spontane Anwendung von Strategien als jüngere Menschen (Dunlosky & Hertzog, 2001; Dunlosky et al., 2005; Hulicka & Grossman, 1967, für einen Überblick siehe Dunlosky et al., 2005). Werden sie dahingegen explizit aufgefordert, Strategien zu verwenden, so steigt die Strategienutzung älterer Menschen signifikant an, dies bestätigt ein Produktionsdefizit (Dunlosky & Hertzog, 1998, 2001; Kausler, 1994; Treat & Reese, 1976, zusammenfassend siehe Kausler, 1994). Allerdings finden sich trotz der Verwendung von Strategien weiterhin Altersunterschiede, so dass die Erklärung eines Produktionsdefizits alleine in keinem Fall ausreicht, um die Altersvarianz aufzuklären.

Im Gegensatz zu der Evidenz eines Produktionsdefizits konnte bisher keine Evidenz für ein Mediationsdefizit gesammelt werden (Dunlosky & Hertzog, 1998; Treat & Reese, 1976). Werden ältere Menschen auf die Verwendung von Strategien hingewiesen, so berichten sie, diese auch anzuwenden. Natürlich sind hier die Grenzen der Selbstberichterstattung zu beachten, die im letzten Abschnitt erwähnte Leistungssteigerung spricht aber für die Validität der Selbstberichte.

In Hinsicht auf ein Nutzungsdefizit liegen unterschiedliche Befunde vor. In manchen Experimenten profitieren beide Altersgruppen gleichermaßen von der Verwendung einer Strategie (Dunlosky et al., 2005; Hulicka & Grossman, 1967), in anderen Studien dahingegen bleibt die Leistungssteigerung der älteren Menschen hinter denen der jüngeren zurück (Dunlosky & Hertzog, 1998; Kliegl et al., 1989). Diese Divergenzen könnten auf unterschiedliche Präsentationszeiten und Unterschiede des Wortmaterials zurückgeführt werden. Bei kurzen Präsentationszeiten und abstraktem Wortmaterial z.B. sollte es älteren Menschen schwerer fallen, eine Strategie effizient anzuwenden. Systematisch wurde dieser Faktor leider bisher nicht untersucht.

Ein Abrufdefizit konnte über die geringere Rate korrekter Strategieerinnerungen festgestellt werden (Dunlosky et al., 2005). Um das Dekodierungsdefizit zu untersuchen, wurden die Fälle herangezogen, in denen eine Strategie in der Lernphase angewendet und im Test erfolgreich erinnert wurde, also kein Abrufdefizit vorlag. Vergleicht man nun die Gedächtnisleistung zwischen den Altersgruppen, so schneiden ältere Menschen weiterhin schlechter ab als jüngere Menschen (Dunlosky et al., 2005). Dies bedeutet, dass es älteren Menschen in einigen Fällen trotz korrekter Erinnerung an den Hinweisreiz nicht gelingt, das Zielwort zu reproduzieren. Dieser Befund wird im Sinne eines Dekodierungsdefizits interpretiert (Dunlosky et al., 2005).

Zusammenfassend zeigen ältere Menschen ein klares Produktionsdefizit sowie in den wenigen Studien, die ein Abruf- und Dekodierungsdefizit untersucht haben, auch diese Defizite. Die Befunde hinsichtlich des Nutzungsdefizits sind unterschiedlich, wobei diese Unterschiede wahrscheinlich auf die Präsentationszeiten oder Abstraktheit des Wortmaterials zurückgeführt werden können (Treat & Reese, 1976). Von diesen Faktoren sollte abhängig sein, wie leicht die Nutzung einer Strategie fällt. Eindeutig stellt sich heraus, dass die Altersunterschiede der Gedächtnisleistung durch Alterseffekte der Strategieanwendung moderiert werden (zusammenfassend siehe Dunlosky et al., 2005; Shing et al., 2008).

Empirische Befunde zu strategischen Altersunterschieden in assoziativen Rekognitionsaufgaben

Die meisten der berichteten Studien verwendeten geförderte Reproduktionstests. Hinsichtlich der anzuwendenden Strategien sind assoziative Rekognitionstests und geförderte Reproduktionstests durchaus vergleichbar. Auch bei neu kombinierten Wortpaaren eines

assoziativen Rekognitionstests müssen die Probanden versuchen, die Wortpaare aus der Lernphase abzurufen („recall-to-reject“), um zu entscheiden, ob die beiden Wörter in dieser Kombination oder in einer anderen Kombination gelernt wurden. Allerdings kann bei der Verwendung von geförderten Reproduktionstests das assoziative Defizit nicht spezifisch untersucht werden, da der Vergleich zwischen assoziativem Test und Itemtest fehlt. Aus den bisher geschilderten Studien kann geschlussfolgert werden, dass strategische Altersunterschiede vorliegen. Inwieweit das assoziative Defizit auf qualitative Verarbeitungsunterschiede zurückzuführen ist bleibt bisher aber offen. Aus diesem Grund werden im Folgenden zwei Studien genauer vorgestellt, die sich mit strategischen Defiziten in assoziativen Rekognitionsaufgaben befassen (Naveh-Benjamin, Brav & Levy, 2007; Shing et al., 2008).

Naveh-Benjamin und Kollegen (Naveh-Benjamin, Brav & Levy, 2007) verwendeten einen Itemtest sowie einen assoziativen Test, um das Assoziationsdefizit zu erfassen. Sie interessierten sich vor allem dafür, ob ein Produktionsdefizit das Assoziationsdefizit älterer Menschen erklären kann. Nach Ansicht der Autoren kann ein Produktionsdefizit sowohl in der Lernphase als auch in der Testphase auftreten. Ein Produktionsdefizit in der Testphase würde darin bestehen, dass ältere Menschen spontan seltener *versuchen*, die in der Lernphase verwendeten Strategien wieder zu erinnern. Inwieweit zusätzlich zu einem Produktionsdefizit in Lern- und Testphase weitere Strategiedefizite auftreten, wurde in dieser Studie außer Acht gelassen. Als abhängige Variable diente allein die Gedächtnisleistung, itemspezifische Strategieranwendungen wurden nicht erfragt. Die Fragestellung führte zu drei experimentellen Gruppen. In der ersten Gruppe wurden keinerlei Hinweise auf Strategien gegeben. Die Probanden wurden lediglich darüber informiert, dass sie Wortpaare für einen späteren Gedächtnistest lernen sollten (intentionales Lernen). Eine zweite Gruppe wurde über Forschungsergebnisse informiert, die zeigen, dass es leichter fällt sich zwei Wörter einzuprägen, wenn man versucht, die beiden Wörter in einen sinnvollen Satz einzubetten. Die Probanden wurden dazu ermutigt, diese Strategie zu nutzen (Aufforderung zur Strategieranwendung in der Lernphase). In der dritten Experimentalgruppe wurden die Probanden ebenfalls ermutigt, die beiden Wörter in einen Satz zu integrieren. Zusätzlich erhielten diese Probanden auch vor Beginn der Testphase den Hinweis, sich während des Tests immer an die generierten Sätze zu erinnern und diese zu nutzen, um ihre Gedächtnisleistung zu verbessern (Aufforderung zur Strategieranwendung in der Lern- und Testphase). Alle drei Gruppen hatten Gelegenheit, sich in einer Übungsphase mit den Lern- und Testbedingungen vertraut zu machen. In dieser Phase konnte auch die Strategie eingeübt werden, und der Versuchsleiter stand danach für Fragen zur Verfügung.

In der ersten Gruppe (intentionales Lernen), die keine weiteren Hinweise zu Strategien erhielt, wurde das klassische Assoziationsdefizit repliziert. In der zweiten Gruppe (Aufforderung zur Strategieranwendung in der Lernphase) steigerte sich die Gedächtnisleistung älterer Menschen. Ältere Menschen profitierten sowohl im Itemtest als auch im assoziati-

ven Test stärker von dem Hinweis auf eine konkrete Strategie als jüngere Menschen. Diese Leistungssteigerung wurde im Sinne eines Produktionsdefizits interpretiert; unterstützt wird dieser Schluss durch die Daten der Nachbefragung. Alle Probanden wurden nach der Testphase gefragt, ob sie Strategien angewendet hatten und falls ja, welche Strategien. In der ersten Probandengruppe gaben alle Studenten an, Strategien genutzt zu haben. Die Mehrheit (63%) berichtete den Gebrauch von Verbindungssätzen, 29% stellten sich die Wörter interaktiv vor, und 8% versuchten Assoziationen zwischen den Wörtern zu nutzen. Bei den älteren Menschen dahingegen berichteten insgesamt nur 11% die Verwendung einer elaborierten Strategie. Sie bildeten Assoziationen zwischen den Wörtern, um ihre Leistung zu steigern. Dieses Bild veränderte sich in der zweiten Experimentalgruppe: Nach der Aufforderung, Strategien zu verwenden, berichteten 100% der älteren Menschen, die vorgeschlagene Satzstrategie genutzt zu haben. Auch in der dritten Experimentalgruppe berichteten 100% der älteren Menschen, sowohl während der Lern- als auch während der Testphase eine elaborierte Strategie genutzt zu haben (88% Bildung von Sätzen; 12% Nutzung von Assoziationen). Auch mit der zusätzlichen Aufforderung, Strategien in Test- und Lernphase zu verwenden (Gruppe 3), sind ältere Menschen in der Lage, ihre Gedächtnisleistung weiter zu verbessern. Sie profitieren von dem zusätzlichen Hinweis, die Strategien auch in der Testphase anzuwenden.

Die Autoren schlussfolgern aus diesen Ergebnissen, dass ein strategisches Produktionsdefizit älterer Menschen an dem Assoziationsdefizit beteiligt ist. Obwohl das assoziative Defizit in dieser Studie sogar fast aufgehoben werden konnte („the associative deficit was almost eliminated under this condition“, Naveh-Benjamin, Brav & Levy, 2007, S. 206), wird auf die Beteiligung weiterer Faktoren hingewiesen. Die zu Anfang des Kapitels gestellte Frage, ob das Assoziationsdefizit allein durch Altersunterschiede in stragischen Defiziten erklärt werden kann, würden Naveh-Benjamin und Kollegen also negativ beantworten.

Auch Shing et al. (2008) gingen der Frage nach, inwieweit das Assoziationsdefizit allein auf ein Strategiedefizit zurückzuführen ist. In ihrer Lebensspannen-Studie wurden nicht nur Studenten und ältere Menschen untersucht, sondern auch Kinder im Alter von 10-12 Jahren sowie Jugendliche zwischen 13 und 15 Jahren. Das Ziel der Studie war es, die Entwicklung des assoziativen Defizits über die Lebensspanne zu verfolgen und den jeweiligen Einfluss von strategischen Prozessen bzw. assoziativen Gedächtnisprozessen aufzuzeigen. Shing et al. (2008) nehmen an, dass sich das Assoziationsdefizit auf zwei Prozesse zurückführen lässt: Zum einen auf einen strategischen Anteil, der sich durch geeignete Strategievermittlung und durch Strategietraining minimieren lassen sollte; zum anderen auf einen assoziativen Gedächtnisprozess, der von einem Strategietraining unbeeinflusst bleiben sollte. Dies entspricht der mulikausalen Auffassung des assoziativen Defizits von Naveh-Benjamin.

Hinsichtlich der Entwicklung dieser beiden Prozesse über die Lebensspanne stützen sich die Autoren auf neurophysiologische Korrelate und leiten daraus ihre Hypothesen ab. Korrelate der assoziativen Gedächtniskomponente werden in medio-temporalen Gebieten, vor allem im Hippocampus vermutet (Eichenbaum, 2002; Simons & Spiers, 2003); die Entwicklung von strategischen Kompetenzen wird mit der Entwicklung des frontalen Cortex assoziiert (E. K. Miller & Cohen, 2001; Simons & Spiers, 2003). Über die Lebensspanne weisen diese Areale unterschiedliche Entwicklungsverläufe auf. Während die Synaptogenese und Myelinisierung der frontalen Arealen bis in die mittlere Kindheit reicht (Diamond, 2002; Ofen et al., 2007; Paus, 2005; Thompson & Nelson, 2001), ist die Entwicklung der medio-temporalen Gebiete wie des Hippocampus bereits früher abgeschlossen (Ofen et al., 2007; Thompson & Nelson, 2001). Im Sinne einer ‚first-in-last-out‘-Hypothese (Raz, 2000) zeigt sich im höheren Alter eine frühe Volumenabnahme im frontalen Cortex, wohingegen erst später ein Volumenabbau im Hippocampus zu verzeichnen ist (Buckner, Head & Lustig, 2006; Raz, 2000, 2005). Aus diesen Forschungsergebnissen leiten Shing und Kollegen die Hypothesen ab, dass bei 10-jährigen Kindern die assoziativen Gedächtnisprozesse bereits weit entwickelt sein sollten, wohingegen der frontale Cortex noch nicht voll ausgereift ist. Kinder in diesem Alter sollten ein strategisches Produktionsdefizit aufweisen. Durch ein Strategietraining sollte die vorhandene Plastizität zu einer Leistungsverbesserung führen, die sich den Jugendlichen annähert. Bei älteren Menschen dahingegen sollte die Volumenreduktion im Hippocampus sowie im frontalen Cortex dazu führen, dass sie ein Produktionsdefizit zeigen, welches durch Training vermindert werden kann. Allerdings sollte sich die Post-Trainingsleistung älterer weiterhin klar von der Leistung jüngerer Erwachsener unterscheiden, da die defizitären assoziative Gedächtnisprozesse über das Strategiedefizit hinaus zu stabilen Altersunterschieden führen sollten.

Shing et al. (2008) manipulierten neben dem Alter zwei Faktoren: Zum einen die Enkodierungsstrategien, zum anderen die Art des Wortmaterials. Der Faktor Wortmaterial hatte zwei Stufen: Entweder mussten die Teilnehmern zwei deutsche Wörter lernen, die keine besondere Assoziation zueinander aufwiesen. Oder es wurden Wortpaare aus deutsch-malaiischen Wörtern verwendet, die eine höhere Anforderung an die assoziativen Gedächtnisprozesse der Probanden stellten. Jeder Proband lernte sowohl Listen mit deutsch-deutschen Wortpaaren als auch Listen mit deutsch-malaiischen Wortpaaren. Der Faktor der Enkodierungsstrategien wurde als Messwiederholungsfaktor über drei Zeitpunkte manipuliert. Die Messzeitpunkte unterschieden sich wie folgt: (1) In einer ersten Sitzung wurde den Versuchspersonen mitgeteilt, dass sie im Folgenden Wortpaare präsentiert bekommen und sie jeweils versuchen sollen, sich vor allem die einzelnen Wörter einzuprägen. Die Instruktionen waren also intentional in Hinsicht auf die einzelnen Wörter, aber inzidentell im Hinblick auf die Wortpaare (inzidentelle Paar-Instruktion). (2) In der zweiten Sitzung wurden die Probanden instruiert, sich nun insbesondere die Kombination der Wörter zu merken (intentionale Paar-Instruktion). Bei beiden Sitzungen wurden nach der Lernphase sowohl ein Item- als auch ein assoziativer Rekognitionstest durchgeführt.

Nach diesen ersten beiden Testzeitpunkten folgte eine Sitzung, in der den Teilnehmern effiziente Lernstrategien vermittelt wurden: zum einen die Strategie der interaktiven Vorstellung, zum anderen die Strategie, Verbindungswörter zu generieren. Diese Strategien wurden zunächst in Kleingruppen geübt, danach folgte ein Einzeltraining.

(3) Nach dem Strategietraining fand eine dritte und vierte Lern-Test-Phase statt. Wie nach den ersten beiden Sitzungen wurde auch hier ein Item- und assoziativer Gedächtnistest durchgeführt. Die Ergebnisse dieser letzten Sitzungen wurden aber zusammengefasst und gemeinsam ausgewertet (Test nach Training).

Nur für die schwierigen deutsch-malaiischen Wörter fanden weitere Übungssitzungen statt (Shing et al., 2008, Exp. 2). Dabei wurde die Gedächtnisleistung einmal in der ersten zusätzlichen Trainingssitzung erfasst (4, Test nach erster Übung) und in der sechsten zusätzlichen Trainingssitzung am Ende des gesamten Trainingsblocks (5, Test nach letzter Übung).

Die Analysen zeigten, dass bei deutschen Wortpaaren der Leistungszuwachs von inzidental zu intentionaler Lernbedingung (Sitzung 1, 2) für Studenten und junge Erwachsene signifikant größer war als für Kinder und ältere Menschen. Zwischen dem Zuwachs bei Kindern und älteren Menschen war kein signifikanter Unterschied zu verzeichnen. Dies änderte sich nach der Strategieinstruktion (Sitzung 3). Hier zeigte sich, dass Kinder zu einer stärkeren Leistungsverbesserung in der Lage waren als ältere Menschen. Dies kann als erster Hinweis für ein Produktionsdefizit in beiden Altersgruppen angesehen werden. Sowohl Kinder als auch ältere Probanden verbesserten ihre Leistung, wenn sie explizit auf die Verwendung von Strategien hingewiesen wurden. Allerdings profitierten ältere Menschen (evtl. aufgrund eines zusätzlichen Defizits in assoziativen Gedächtnisprozessen) weniger als Kinder. Diese Interpretation wird weiter durch die Analysen der schwierigen deutsch-malaiischen Wortpaare unterstützt. Kinder profitierten von den zusätzlichen fünf Trainingssitzungen (Sitzung 4, 5) und konnten ihre Leistung weiter steigern. Ältere Menschen dahingegen scheinen bereits vor den zusätzlichen Übungseinheiten ein Leistungslimit erreicht zu haben und zeigten keine weitere Leistungssteigerung.

Wichtig ist also der Befund, dass Kinder in Abhängigkeit von den Enkodierungsstrategien und über die Trainingssitzungen hinweg ihre Leistung verbessern, wohingegen bei älteren Menschen ein klares Plateau erreicht wird (siehe Abbildung 3.6). Dieses Leistungsmuster wird von den Autoren dahingehend interpretiert, dass ältere Menschen zunächst ihr Produktionsdefizit überwinden, dann aber aufgrund der defizitären assoziativen Gedächtnisprozesse an ihre Leistungsgrenzen stoßen und keine weitere Steigerung möglich ist. Bei Kindern dahingegen ist es lediglich wichtig, das Produktionsdefizit zu überwinden. Da die assoziativen Gedächtnisstrukturen schon weit entwickelt sind, zeigt sich ein stetiger Anstieg, wie er auch bei Jugendlichen zu verzeichnen ist. An dieser Stelle wurden nur die für dieses Kapitel relevanten Ergebnisse aus der komplexen Studie von Shing et al. (2008) berichtet, hinsichtlich der weiteren Ergebnisse sei der interessierte Leser auf den Originalartikel verwiesen.

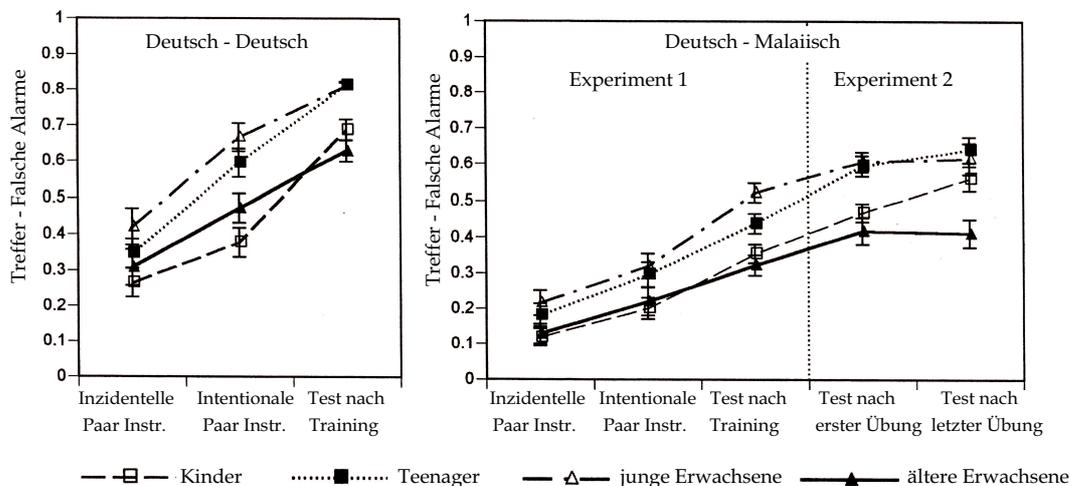


Abbildung 3.6: In dieser Abbildung sind die Ergebnisse der Lebensspannenstudie von Shing et al. (2008) dargestellt. In der linken Graphik sind die Gedächtnisleistungen für die deutsch-deutschen Wortpaare (drei Testzeitpunkte) abgebildet, in der rechten Graphik die Leistungen für die deutsch-malaiischen Wortkombinationen (fünf Testzeitpunkte). Die mittleren Leistungen sind jeweils getrennt für die vier Altersgruppen abgebildet. [Abbildung aus Shing et al. (2008), S. 504]

Sowohl die Studie von Naveh-Benjamin, Brav und Levy (2007) als auch die Experimente von Shing et al. (2008) kommen zu dem Schluss, dass das assoziative Defizit älterer Menschen nicht allein auf strategischen Defiziten beruht. Zwar trägt ein strategisches Defizit zu dem assoziativen Altersdefizit bei, es kann die Alterseffekte aber nicht vollständig erklären. In Anbetracht der Studien von Dunlosky und Hertzog (1998, 2001) sowie Dunlosky et al. (2005) muss allerdings darauf hingewiesen werden, dass weder Naveh-Benjamin, Brav und Levy (2007) noch Shing et al. (2008) alle Varianten strategischer Defizite untersucht haben. Auch wurden die Strategien gar nicht bzw. nicht itemspezifisch erfasst. Daher kann argumentiert werden, dass das Fazit aus den beiden assoziativen Rekognitionsstudien übereilt ist; ein Argument, dass nur durch weitere empirische Studien entkräftet werden kann. Auf der anderen Seite fügen sich die Befunde von Shing et al. (2008) und Naveh-Benjamin, Brav und Levy (2007) gut in das Gesamtbild der strategischen Altersstudien ein. Denn wie schon Salthouse (1991) schlussfolgerte, spielen strategische Defizite zwar eine Rolle, sie sind aber nicht in der Lage, die gesamte Altersvarianz aufzuklären. Das nächste Kapitel widmet sich der verbleibenden Altersvarianz. Es wird erläutert, wie die assoziativen Gedächtnisprozesse mit anderen kognitiven Basismechanismen zusammenhängen. Auch wird der Frage nachgegangen, ob es gerechtfertigt ist, assoziative Gedächtnisprozesse als Basisprozesse anzusehen.



Abbildung 3.7: Diese Abbildung wurde bereits in Kapitel 2 präsentiert, um zu veranschaulichen, wie die Forschungsfrage in das Feld der kognitiven Altersforschung eingeordnet werden kann. Hervorgehoben sind die für das folgende Kapitel relevanten kognitiven Basismechanismen, wobei es zu klären gilt, inwieweit das assoziative Defizit als eigenständiger Basismechanismus angesehen werden kann.

3.2.4 Das assoziative Defizit in Verbindung zu anderen Basismechanismen

Nach den Ausführungen des letzten Kapitels kann die Schlussfolgerung gezogen werden, dass das Assoziationsdefizit nicht allein auf einem strategischen Defizit beruht. Um die Ursachen des Altersunterschieds aufzudecken, muss die theoretische Perspektive der qualitativen Verarbeitungsunterschiede (Salthouse, 1991, siehe Kap. 2) also ergänzt werden um die Perspektive der quantitativ reduzierten Basismechanismen. In Bezug auf diesen Ansatz gilt es zu klären, ob das Assoziationsdefizit auf bekannte Basismechanismen wie reduzierte Verarbeitungsgeschwindigkeit, Inhibitionsdefizit oder Aufmerksamkeitsdefizit zurückgeht (siehe Kap. 2, siehe Abbildung 3.7), oder ob es ein spezifisches assoziatives Gedächtnisdefizit gibt. Um eine solche Abgrenzung vorzunehmen, scheint es notwendig, kurz auf die alternativen Basismechanismen einzugehen. Zunächst werden also die traditionell diskutierten Basismechanismen vorgestellt, dann wird herausgearbeitet, inwieweit das assoziative Defizit in die Reihe dieser Basismechanismen aufgenommen werden kann.¹³

¹³An dieser Stelle sei angemerkt, dass die Theorien des Inhibitionsdefizits, der kognitiven Verlangsamung und der reduzierten Aufmerksamkeit versuchen, dieselben Altersphänomene über unterschiedliche Basismechanismen zu erklären. Sie stehen sich in der Forschung als konkurrierende Ansätze gegenüber. Trotzdem weisen sie zum Teil ähnliche Argumentationsmuster auf und schließen sich nicht grundsätzlich gegenseitig aus.

Kognitive Verlangsamung

Die Hypothese, dass eine Verlangsamung der Verarbeitungsgeschwindigkeit Auswirkungen auf kognitive Leistungen hat, wird von mehreren Autoren vertreten (Cerella, 1985; Myerson, Hale, Wagstaff, Poon & Smith, 1990; Salthouse, 1996). Im Folgenden werden diese Ansätze exemplarisch an der Theorie von Salthouse (1996) erläutert. Salthouse definiert Verarbeitungsgeschwindigkeit als „the speed with which many cognitive operations can be executed“ (Salthouse, 1996, S. 403). Für den Zusammenhang von Geschwindigkeit und Kognition werden zweierlei Mechanismen postuliert: (1) Der Mechanismus der begrenzten Zeit („limited time mechanism“) und (2) der Mechanismus der Simultanität („simultaneity mechanism“).

Der Mechanismus der begrenzten Zeit (1) greift vor allem bei Aufgaben, die eine externe zeitliche Begrenzung aufweisen. Hierunter fallen z.B. Gedächtnisaufgaben mit einer festgelegten Enkodierungszeit. Nach Salthouse führt eine verlangsamte Verarbeitungsgeschwindigkeit dazu, dass bereits für Operationen wie das Lesen der Wörter mehr Zeit benötigt wird. Daher steht für höhere Prozesse wie die Anwendung von elaborierten Enkodierungsstrategien oder das Verknüpfen mit vorangegangenen Zielwörtern *weniger* Zeit zur Verfügung, obwohl diese Prozesse bei älteren Menschen sogar *mehr* Zeit benötigen würden. Folglich resultieren bei älteren Menschen Leistungseinbußen, die durch den Mechanismus der begrenzten Zeit erklärt werden können.

Die erfolgreiche Bearbeitung kognitiver Aufgaben verlangt oftmals auch die simultane Aufrechterhaltung und Bereitstellung mehrerer Zwischenergebnisse, unabhängig von zeitlichen Begrenzungen. Hier greift der Mechanismus der Simultanität (2). Im Verlauf eines Problemlöseprozesses z.B. sind die Ergebnisse zeitlich früherer Operationen für spätere Verarbeitungsprozesse von Bedeutung und müssen für diese verfügbar sein. Wenn nicht alle Zwischenergebnisse simultan aufrechterhalten werden können, kann eine komplexe Aufgabe nicht bewältigt werden, da relevante Teilaspekte vernachlässigt wurden. Die Wahrscheinlichkeit eines solchen Zerfalls vorgeschalteter Informationen steigt bei Menschen mit einer langsamen Verarbeitungsgeschwindigkeit an.

Die zentrale Annahme besteht also darin, dass die Verfügbarkeit von Information als Funktion der Zeit abnimmt. Werden mentale Operationen langsam durchgeführt, ist es wahrscheinlich, dass Zwischenergebnisse bereits nicht mehr zur Verfügung stehen, wenn sie benötigt werden.

Im Kern führt die Theorie von Salthouse kognitive Leistungsunterschiede zwischen Altersgruppen auf eine Veränderung der Informationsgeschwindigkeit zurück. Salthouse (1996) sieht die Verarbeitungsgeschwindigkeit zwar nicht als alleinige Ursache für Altersvarianz an, geht aber davon aus, dass die Geschwindigkeit der Informationsverarbeitung für den Großteil der Altersunterschiede verantwortlich ist: „A slower speed [...] is not assumed to be the exclusive source of age-related differences, because other age-related influences are also postulated to exist. Nevertheless, a reduction with increased age in the

speed [...] is hypothesized to be a major contributor to the adult age differences in many measures of cognition“ (Salthouse, 1996, S. 404). Die empirischen Arbeiten von Salthouse und anderen Autoren zeigen, dass die Verarbeitungsgeschwindigkeit einen großen Anteil der Varianz zwischen Altersgruppen in unterschiedlichen kognitiven Aufgaben aufklären kann (z.B. Verhaeghen & Salthouse, 1997).

Theorien der reduzierten Aufmerksamkeitsressourcen

Craik (Craik, 1986; Craik & Byrd, 1982) zieht begrenzte Aufmerksamkeitsressourcen und reduzierte selbstinitiierte Prozesse als Erklärungsvariablen für Altersunterschiede in kognitiven Leistungen heran. Dieser Ansatz lässt sich durch die Interaktion von internen und externen Faktoren beschreiben: Externe Faktoren sind z.B. Hinweise auf die zu erinnernde Information, während man unter internen Faktoren Aufmerksamkeitsressourcen versteht. Craik und Byrd (1982) gehen davon aus, dass Aufmerksamkeitsressourcen mit zunehmendem Alter abnehmen. Damit einher geht u.a. eine Veränderung der Enkodierungs- und Abrufstrategien in Gedächtnisaufgaben, da diese ressourcenaufwendig sind. In der Lernphase kann tief elaboriert werden, wenn ausreichend Aufmerksamkeitsressourcen vorhanden sind. Ist dies nicht der Fall, muss auf weniger kapazitätsfordernde Strategien wie einfaches Wiederholen ausgewichen werden. Diese unterschiedlichen Enkodierungsstrategien wirken sich wiederum auf die Erinnerungsleistung aus.

Im Hinblick auf externe Faktoren können Gedächtnisaufgaben u.a. danach einteilt werden, wie stark sie Hinweise auf das zu erinnernde Material geben: In freien Reproduktionsaufgaben sind keine Hilfestellungen vorhanden, in geförderten Reproduktionsaufgaben dahingegen schon. Je weniger externe Hinweise gegeben werden, desto wichtiger werden selbstinitiierte Prozesse (z.B. selbstinitiierte Such- und Abrufstrategien) für die erfolgreiche Erinnerung. Craik (1986) postuliert, dass diese selbstinitiierten Prozesse für die schlechtere Leistung älterer Menschen verantwortlich sind, da sie (genau wie Enkodierungsstrategien) Aufmerksamkeitsressourcen erfordern.

Craik und McDowd (1987) untersuchten diese Vorhersagen mit Reproduktions- und Rekognitionsaufgaben unter Berücksichtigung der Schwierigkeit der Aufgaben. Die signifikante Interaktion zwischen Alter und Aufgabenart bei einer vergleichbaren Schwierigkeit der Aufgaben unterstützt die These, dass nicht die ansteigende Schwierigkeit, sondern der stärkere Rückgriff auf Ressourcen ausschlaggebend für die Altersunterschiede ist. Es ist nicht die relative Schwierigkeit, die Älteren Probleme macht, sondern die qualitative Art der Reproduktionsaufgabe, die im Vergleich zur Rekognitionsaufgabe mehr selbstinitiierte Prozesse verlangt. Die Autoren schließen daraus, dass Reproduktionsaufgaben mehr Prozessressourcen verlangen als Rekognitionsaufgaben und dass ältere Menschen aufgrund ihrer geringeren Aufmerksamkeitsressourcen bei diesen Aufgaben schlechter abschneiden.

Der Ansatz von Craik (Craik, 1986; Craik & Byrd, 1982) lässt sich wie folgt zusammenfassen: Mit zunehmendem Alter nehmen Aufmerksamkeitsressourcen sowie selbstinitiierte

Prozesse ab. Eine Leistungsver schlechterung älterer Menschen tritt insbesondere dann auf, wenn eine Aufgabe ressourcenfordernd ist. Dies erklärt bspw. warum Altersunterschiede in Rekognitionsaufgaben geringer ausfallen als in freien Reproduktionsaufgaben.

Theorie des Inhibitionsdefizits

Die Inhibitionsdefizit-Theorie (Hasher & Zacks, 1988) führt Leistungsdefizite älterer Menschen auf ineffiziente inhibitorische Mechanismen zurück. Inhibitorische Mechanismen unterliegen der selektiven Aufmerksamkeit und hemmen den Zugang irrelevanter Information zum Arbeitsgedächtnis. Im Gegensatz dazu werden relevante Inhalte über exzitatorische Mechanismen aktiviert. Diese beiden Prozesse stellen sicher, dass nur Informationen verarbeitet werden, die für die augenblickliche Aufgabe oder Situation von Bedeutung sind. Gelangt irrelevante Information in das Arbeitsgedächtnis, so interferiert diese mit relevanter Information. Dies kann zu einer geringeren Kapazität für relevante Information führen. Unterschiede in kognitiven Aufgaben (z.B. Textverstehen, Gedächtnisspannenaufgaben, Abruf von Information aus dem Langzeitgedächtnis) werden von Hasher und Kollegen (Hasher & Zacks, 1988; Hasher et al., 1999; May, Hasher & Kane, 1999) auf die Güte der inhibitorischen Mechanismen zurückgeführt.

Inhibition wird von Hasher et al. (1999) über drei Funktionen spezifiziert: Zugangsinhibition, Löschung und Reaktionshemmung. Die erste Funktion verhindert, dass aufgabenirrelevante Information in das Arbeitsgedächtnis gelangt. Lustig, Hasher und Tonev (2001) sprechen von der Inhibition aktueller Information. Wird dagegen im Arbeitsgedächtnis eine nicht länger relevante Information durch Löschung erfolgreich gehemmt, sprechen die Autoren von Inhibition zurückliegender Information. Die letzte Funktion der Hemmung bezieht sich auf automatische Reaktionen. Ziel einer Aufgabe kann es sein, gelernte Reaktionen gerade nicht auszuführen (z.B. das automatische Lesen von Wörtern in der Stroop-Aufgabe). Auch hier muss auf Inhibitionsmechanismen zurückgegriffen werden. In Gedächtnisexperimenten sind zumeist die ersten beiden Funktionen für eine hohe Gedächtnisleistung erforderlich. Eine weitere Unterscheidung treffen die Autoren hinsichtlich der Quelle der zu hemmenden Information (Hasher et al., 1999; Lustig, Hasher & Tonev, 2001). Externe Ablenkungen entstammen der Umwelt, interne Ablenkungen dem Individuum selbst z.B. in Form von selbstbezogenen Kognitionen.

Zusammenfassend betrachtet konnte die Inhibitionsdefizit-Theorie durch zahlreiche Studien in unterschiedlichen Paradigmen empirisch untermauert werden (Hasher & Zacks, 1988; Hasher et al., 1999; Kane, Hasher, Stoltzfus & Zacks, 1994; Lustig, May & Hasher, 2001; May, Zacks, Hasher & Multhaup, 1999). Auch zeigten die Autoren auf, welche Bedingungen die Inhibitionsleistung moderieren (z.B. Vorhersehbarkeit der ablenkenden Reize oder Zeitpunkt der Testung). Hasher und Zacks (1988) weisen außer auf moderierende Bedingungen auch auf Kompensationsstrategien älterer Menschen hin, die diese möglicherweise entwickeln, um ihre mangelnde Erinnerungsfähigkeit zu verbessern.

Das assoziative Defizit als Basismechanismus

Im Vergleich zu den gerade vorgestellten kognitiven Basismechanismen ist das assoziative Defizit erst in den letzten Jahren in den Fokus der Forschung gerückt, und es liegen zu diesem Thema bisher wenige Forschungsergebnisse vor. Daher werden in diesem Kapitel publizierte Studien (Castel & Craik, 2003; Kilb & Naveh-Benjamin, 2007; Naveh-Benjamin, Guez & Marom, 2003; Naveh-Benjamin, Guez & Shulman, 2004) auch durch aktuelle Konferenzbeiträge ergänzt (Graef & Naveh-Benjamin, 2008; Naveh-Benjamin, 2008a; Naveh-Benjamin, Kilb & Hara, 2007). Die Konferenzbeiträge stellen in Aussicht, dass in den nächsten Jahren weitere Publikationen zu den Ursachen des Assoziationsdefizits erscheinen werden. Es bleibt daher abzuwarten, inwieweit der hier skizzierte Trend durch zukünftige Forschung bestätigt wird.

Bevor zu der Frage, inwieweit das Assoziationsdefizit auf einen eigenständigen assoziativen Basismechanismus zurückgeht, konkrete Studien berichtet werden, sollen drei prinzipielle experimentelle Herangehensweisen an diese Fragestellung vorgestellt werden.

(1) Ein gängiges Verfahren besteht darin, über experimentelle Manipulationen bei jüngeren Menschen die verfügbaren Verarbeitungsressourcen zu reduzieren und zu beobachten, ob jüngere unter diesen Bedingungen ein ähnliches Ergebnismuster zeigen wie ältere Menschen (Craik, 1982, 1983; Castel & Craik, 2003). Häufig werden z.B. jüngere Menschen mit einer Zusatzaufgabe belastet, die eine reduzierte Aufmerksamkeit für die Hauptaufgabe hervorrufen soll. Falls das Ergebnismuster älterer Menschen mit dem der jüngeren in der Zweitaufgaben-Bedingung übereinstimmt, wird daraus geschlussfolgert, dass z.B. reduzierte Aufmerksamkeit ursächlich für den Altersunterschied ist.¹⁴ An diesem Verfahren sind meiner Ansicht nach gleich mehrere Faktoren kritisch hervorzuheben. Zum einen hängen die Ergebnisse stark von der Wahl der Zweitaufgabe und deren Interaktion mit der Hauptaufgabe ab. Zielen beide Aufgaben auf dieselben Ressourcen und Modalitäten ab, so ist ein anderes Ergebnismuster zu erwarten, als wenn die Aufgaben peripher zueinander gestaltet sind. Außerdem muss die Leistung in der Zweitaufgabe berücksichtigt werden, um Kompensationsstrategien auszuschließen. Auch die Schwierigkeit der Zweitaufgabe kann einen bedeutenden Einfluss haben. Als wichtigster Kritikpunkt kann eingewandt werden, dass es fraglich ist, ob solche Zweitaufgaben bei jungen Menschen überhaupt die Prozesse belasten, die bei älteren Menschen reduziert sind. Trotz dieser m.A.n. wichtigen Kritikpunkte wird das Verfahren häufig eingesetzt, oft ohne dass diese Aspekte kritisch diskutiert werden.

(2) Eine ähnliche Logik liegt zugrunde, wenn die dargebotenen Stimuli degradiert werden, um die Wahrnehmungsbedingungen älterer Menschen zu simulieren (Dickinson & Rabbitt, 1991; Lindenberger, Scherer & Baltes, 2001; Murphy, Craik, Li & Schneider, 2000; Spe-

¹⁴Alternativ zu einer Reduktion der Aufmerksamkeitsressourcen kann auch die zur Verfügung stehende Bearbeitungszeit verändert werden, um die Auswirkungen von Verarbeitungsgeschwindigkeit genauer zu untersuchen.

ranza, Daneman & Schneider, 2000). Im Unterschied zu dem gerade beschriebenen Ansatz geht es hier nicht darum, die kognitiven Verarbeitungsprozesse älterer Menschen nachzuzahlen, sondern vergleichbare Wahrnehmungsbedingungen herzustellen. Visuelle Stimuli werden dabei vor allem hinsichtlich des Kontrasts verändert (Dickinson & Rabbitt, 1991), akustische Stimuli werden von einem Hintergrundrauschen begleitet (Murphy et al., 2000; Speranza et al., 2000).

Hintergrund dieses experimentellen Vorgehens sind Studien, die eine hohe Korrelation von kognitiven und sensorischen Leistungen zeigen (Baltes & Lindenberger, 1997; Lindenberger & Baltes, 1994). Für die Ursache dieses korrelativen Zusammenhangs werden unterschiedliche Modelle diskutiert (für einen Überblick siehe K. Z. H. Li & Lindenberger, 2002). Unter anderem wird angenommen, dass sowohl kognitive als auch sensorische Prozesse auf gemeinsame Ressourcen zurückgreifen („shared resource model“). Sobald mit ansteigendem Lebensalter ein höherer Anteil dieser Ressourcen für die Wahrnehmung der relevanten Stimuli aufgewendet werden muss, stehen weniger Ressourcen für kognitive Verarbeitungsprozesse zur Verfügung (K. Z. H. Li & Lindenberger, 2002). Dieser Ansatz wird durch Befunde gestützt, die zeigen, dass auch jüngere Menschen unter erschwerten Wahrnehmungsbedingungen eine verringerte kognitive Leistung aufweisen (Dickinson & Rabbitt, 1991; Murphy et al., 2000; Rabbitt, 1966; Speranza et al., 2000). Für alternative Erklärungsansätze (z.B. „common-cause hypothesis“) sprechen Studien, die ein solches Ergebnismuster nicht replizieren konnten (Lindenberger et al., 2001). Dieser Ansatz wurde nicht im Rahmen der kognitiven Basismechanismen vorgestellt, da kognitive Basismechanismen nicht als Ursache des kognitiven Alterns berücksichtigt werden; Ziel ist es vielmehr, theoretische Erklärungen für die Korrelation von sensorischen und kognitiven Leistungen abzuwägen. Es geht darum, kognitive Basismechanismen sowie kognitive Leistungen mit neurophysiologischen Ursachen in Verbindung zu bringen, die gleichzeitig auch sensomotorische und sensorische Leistungseinbußen verursachen.

Das experimentelle Vorgehen, über degradierte Stimuli die Wahrnehmungsbedingung älterer Menschen zu simulieren, ist m.E. weniger kritisch einzustufen, da sich eine Simulation der Wahrnehmungsbedingungen an psychophysikalischen Richtlinien zur Sehschärfe und Hörleistung im Alter orientieren kann (Schneider & Pichora-Fuller, 2000) und dieses Vorgehen prinzipiell weit weniger komplex ist als eine Simulation kognitiver Verarbeitungsprozesse. Trotzdem können auch diese Ergebnisse durch eine Reihe von Alternativerklärungen kritisiert werden.

(3) Ein letztes Verfahren besteht darin, die Basisfähigkeiten mit gezielten Testverfahren zu erfassen und deren Einfluss auf die abhängige Gedächtnisvariable in einem zweiten Schritt auszupartialisieren. Bleibt nach dieser Kontrolle noch substantielle Altersvarianz bestehen, so wird davon ausgegangen, dass dieser Varianzanteil nicht auf die auspartialisierten Basismechanismen zurückgeht und ggf. einen spezifischen Alterseffekt darstellt (siehe z.B. Salthouse, 1996). Bedeutend für dieses Verfahren ist die reliable und valide

Erfassung der Basismechanismen, denn nur unter dieser Voraussetzung können die Ergebnisse sinnvoll interpretiert werden. Sind diese Bedingungen gegeben, so stellt dieses Verfahren m.E. einen überzeugenden Ansatz dar, den Einfluss von Basismechanismen zu bestimmen, und sollte den anderen beiden Vorgehensweisen vorgezogen werden. Dies ist auch darin begründet, dass es nur mit diesem Ansatz möglich ist, den Einfluss mehrere Basismechanismen gleichzeitig zu untersuchen.

Nachdem vorgestellt wurde, mit welchen Methoden man die Frage nach einem eigenständigen assoziativen Basismechanismus untersuchen kann, wird im Folgenden auf die bisherigen Untersuchungen zu dieser Fragestellung eingegangen. Der empirische Überblick beginnt mit Studien, die die zuerst geschilderte Methodik der Zweitaufgaben bei jüngeren Menschen nutzten. Diese Arbeiten werden trotz der oben geäußerten methodischen Kritik vorgestellt, da sich bisher leider nur wenige Studien überhaupt mit der Frage nach der Stellung des assoziativen Defizits befassen. Die Experimente sollen daher erste Anhaltspunkte geben, die durch Replikationen mit anderen Methoden zu untermauern sind.

Insgesamt wurden zur Untersuchung der Assoziationsdefizit-Hypothese vier Studien mit dieser Methode durchgeführt (Castel & Craik, 2003; Kilb & Naveh-Benjamin, 2007; Naveh-Benjamin, Guez & Marom, 2003; Naveh-Benjamin, Guez & Shulman, 2004). In den drei Studien aus der Arbeitsgruppe von Naveh-Benjamin wurde die Aufmerksamkeit der jüngeren Menschen während der Lernphase geteilt. Es wurden in allen drei Experimenten kontinuierliche Reaktionszeitaufgaben als Zweitaufgaben verwendet. Dabei wurden z.B. Töne unterschiedlicher Frequenzen dargeboten, wobei die Teilnehmer je nach Tonhöhe korrespondierende Tasten drücken mussten (Naveh-Benjamin, Guez & Marom, 2003, Exp. 1 und Exp. 3), oder es wurden visuelle Reize an unterschiedlichen Bildschirmpositionen präsentiert und wieder mussten zu den Präsentationsorten korrespondierende Tasten ausgewählt werden (Naveh-Benjamin, Guez & Marom, 2003, Exp. 2).

Die Zweitaufgabe und die Präsentation des Wortmaterials wurden dabei immer in unterschiedlichen Modalitäten durchgeführt: Wurden also die Zielwörter visuell dargeboten, so wurde als Zweitaufgabe eine akustische Klassifikationsaufgabe verwendet; bei einer akustischen Darbietung des Lernmaterials wurde eine visuelle Zweitaufgabe genutzt. Die geteilte Aufmerksamkeit wurde in einigen Studien innerhalb der Teilnehmer manipuliert (Kilb & Naveh-Benjamin, 2007; Naveh-Benjamin, Guez & Shulman, 2004), in anderen Studien zwischen den Versuchspersonen (Naveh-Benjamin, Guez & Marom, 2003, Exp. 1 und Exp. 3). Die Ergebnisse waren bei allen von Naveh-Benjamin durchgeführten Studien identisch. In den insgesamt zehn Experimenten zeigte sich bei jüngeren Menschen unter geteilter Aufmerksamkeit nie das Muster eines assoziativen Defizits. Die reduzierte Aufmerksamkeit wirkte sich immer gleichermaßen auf die Gedächtnisleistung in den Itemtests und in den assoziativen Tests aus. Das Muster der älteren Menschen konnte also bei jüngeren Menschen unter geteilter Aufmerksamkeit nicht repliziert werden (siehe Abbildung 3.8).

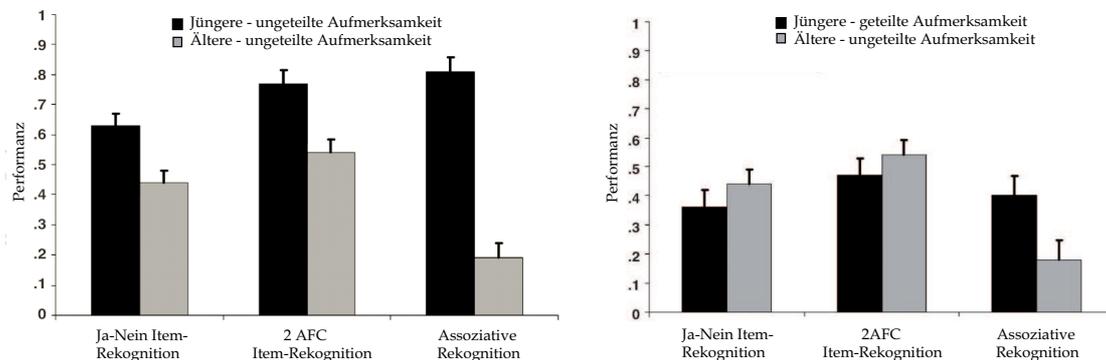


Abbildung 3.8: In beiden Graphiken werden zwei Itemrekognitionstests (Ja-nein- und 2AFC-Test) und ein assoziativer Rekognitionstest einander gegenübergestellt. In der linken Abbildung werden ältere und jüngere Menschen verglichen. In beiden Itemrekognitionstests ist ein Alterseffekt erkennbar, der sich in der assoziativen Testbedingung deutlich verstärkt. Dies zeigt das Assoziationsdefizit älterer Menschen. In der rechten Abbildung ist erneut die Leistung älterer Menschen dargestellt, als Vergleichsgruppe werden nun jüngere Menschen unter geteilter Aufmerksamkeit herangezogen. In der Itemrekognitionbedingung ist kein Altersunterschied mehr erkennbar, ein Effekt zeigt sich aber in der assoziativen Bedingung. [Abbildung aus Naveh-Benjamin, Guez und Shulman (2004, S. 1070)]

Naveh-Benjamin und Kollegen schlussfolgern daraus, dass das Assoziationsdefizit älterer Menschen nicht auf den Basismechanismus der reduzierten Aufmerksamkeit zurückgeführt werden kann.

In der Studie von Kilb und Naveh-Benjamin (2007) wurden nicht nur jüngere Menschen sondern auch ältere Menschen unter geteilter Aufmerksamkeit untersucht. Bei den älteren Menschen unter geteilter Aufmerksamkeit zeigte sich dabei kein verstärktes Assoziationsdefizit im Vergleich zur Bedingung ungeteilter Aufmerksamkeit. Die reduzierte Aufmerksamkeit hat auch bei älteren Menschen einen globalen Effekt auf die Erinnerungsleistung und führt nicht zu einer Verstärkung des spezifischen Defizits. Auch dieser Befund wurde von den Autoren dahingehend interpretiert, dass reduzierte Aufmerksamkeit dem Assoziationsdefizit nicht zugrunde liegt.

Da Naveh-Benjamin und Kollegen verschiedene Zweitaufgaben verwendeten, kann ausgeschlossen werden, dass diese Effekte mit der Wahl einer spezifischen Zweitaufgabe zusammenhängen. Weiterhin kontrollierten sie die Leistung in der Zweitaufgabe, um sicherzugehen, dass diese auch sorgfältig bearbeitet wurde. Nicht untersucht haben die Autoren allerdings, inwieweit eine Zweitaufgabe in derselben Modalität ein anderes Ergebnismuster hervorrufen könnte. Auch wurde nicht erfasst, ob geteilte Aufmerksamkeit während der Testphase ein Assoziationsdefizit bei jungen Menschen verursacht.

Die Bedingung der geteilten Aufmerksamkeit während der Testphase wurde von Castel und Craik (2003) in ihr Experiment aufgenommen. Ihre Studie unterlag denselben Zielset-

zungen wie die Experimente von Naveh-Benjamin. Allerdings verwendeten sie ein etwas anderes Testformat. In Experiment 1 wurde die Aufmerksamkeit nur in der Lernphase geteilt. Im Gegensatz zu den Studien von Naveh-Benjamin zeigten jüngere Menschen in dieser Studie ein Assoziationsdefizit. Sie schnitten in der assoziativen Rekognitionsbedingung signifikant schlechter ab als in der Itemrekognitionsbedingung. Dieses Muster konnte in einem zweiten Experiment repliziert werden. Wird die Aufmerksamkeit zusätzlich während der Testphase vermindert, so verstärkte dies den Befund nicht. Im Unterschied zu den Studien von Naveh-Benjamin finden Castel und Craik (2003) also ein Assoziationsdefizit bei jüngeren Menschen.

Castel und Craik (2003) gelangen allerdings nicht zu der Schlussfolgerung, dass das Assoziationsdefizit älterer Menschen alleine auf reduzierten Aufmerksamkeitsressourcen beruht. In ihrem zweiten Experiment wurde die Zweitaufgabe der jüngeren Menschen so angepasst, dass die Itemgedächtnisleistung in beiden Altersgruppen vergleichbar war. Trotzdem fiel die Leistung in der assoziativen Rekognitionsaufgabe bei älteren signifikant schlechter aus als bei jüngeren Menschen unter geteilter Aufmerksamkeit, ältere Menschen zeigen also ein weitaus größeres Assoziationsdefizit als jüngere Menschen. Außerdem führten unterschiedliche Prozesse zu dem jeweiligen Defizit in der assoziativen Gedächtnisleistung. Während jüngere Menschen eine Reduktion der Trefferraten aufwiesen, führte bei älteren Menschen ein Anstieg in den Raten falscher Alarme zu dem Assoziationsdefizit.

Daher folgern Castel und Craik, dass ihre Ergebnismuster sich zwar von den Ergebnissen aus der Gruppe um Naveh-Benjamin unterscheiden, dass trotz der Differenzen aber dieselben Schlussfolgerungen gezogen werden können und reduzierte Aufmerksamkeitsressourcen nicht die alleinige Ursache für das Assoziationsdefizit älterer Menschen sind. „It seems likely, in fact, that a complete account for age-related memory loss will be multifactorial, reflecting inefficiencies in various types of processing as well as a reduction in attentional resources“ (Castel & Craik, 2003, S. 884). Naveh-Benjamin fasst die Befunde mit folgenden Worten zusammen: „The reported results suggest that depleted attentional resources may result in generally poorer encoding of episodes in the case of both young adults under conditions of divided attention at encoding and older adults. The further differential poorer performance in tests requiring associative information seems to be uniquely related to the effects of age, but not to the effects of divided attention.“ (Naveh-Benjamin, 2006, S. 652).

Nachdem das Assoziationsdefizit nicht auf reduzierte Aufmerksamkeitsressourcen zurückgeführt werden konnte¹⁵, wurde untersucht, inwieweit die veränderten Wahrnehmungsbedingungen älterer Menschen dem Defizit zugrunde liegen. Hinter diesem Ansatz steht der Gedanke, dass ältere Menschen aufgrund von erschwerten Wahrnehmungsbedingungen mehr Kapazität für die Wahrnehmung der Stimuli aufwenden müssen und daher evtl.

¹⁵Zumindest schlussfolgert dies Naveh-Benjamin unter Vernachlässigung der methodischen Kritikpunkte.

Defizite zeigen, die bei jüngeren Menschen nicht zu beobachten sind (K. Z. H. Li & Lindenberger, 2002). In einigen Studien wurde ein solches Vorgehen angewandt und es wurde gezeigt, dass sich erschwerte Wahrnehmungsbedingungen negativ auf die Gedächtnisleistung jüngerer Menschen auswirken (Dickinson & Rabbitt, 1991; Murphy et al., 2000; Rabbitt, 1966; Speranza et al., 2000 siehe aber auch Lindenberger et al., 2001). Graef und Naveh-Benjamin (2008) untersuchten in einem Experiment, inwieweit die degradierte Darbietung von Objekten bei jüngeren Menschen zu einem Assoziationsdefizit führt. Den Probanden wurden Zeichnungen von Objekten gezeigt, die entweder gar nicht, leicht oder moderat degradiert waren. Diese Abstufungen wurden über eine Veränderung der Darbietungsschärfe erzeugt. Die Objekte wurden entweder nur während der Lernphase, nur während der Testphase oder in Lern- und Testphase degradiert präsentiert. Es nahmen sowohl jüngere als auch ältere Menschen an dieser Studie teil, und es wurde ein Item- und ein assoziativer Gedächtnistest durchgeführt. Die Ergebnisse zeigten, dass *ältere* Menschen von der Degradierung kaum beeinflusst werden. Der größte Effekt war bei einer moderaten Degradierung während Lern- und Testphase zu beobachten, wobei sich diese Manipulation auf Item- und assoziative Gedächtnisleistung gleich stark auszuwirken scheint und eher nicht signifikant sein dürfte (da die Autoren leider keinerlei Teststatistiken berichten, ist die Interpretation der Daten erschwert). Ähnlich wie unter reduzierter Aufmerksamkeit führen die erschwerten Wahrnehmungsbedingungen also nicht zu einem verstärkten Assoziationsdefizit älterer Menschen.

Untypischerweise findet sich bei *jungen* Menschen in der Standardbedingung ohne Degradierung ein signifikanter Unterschied zwischen Item- und assoziativer Gedächtnisleistung. Die Leistungsunterschiede sind ähnlich zu den Leistungen der älteren Menschen, und es scheint unwahrscheinlich, dass in der nicht-degradierten Wahrnehmungsbedingung eine signifikante Wechselwirkung von Altersgruppe und Testart, d.h. ein Assoziationsdefizit vorliegt. Jüngere wie ältere Menschen zeigen eine ähnlich starke Diskrepanz zwischen Item- und assoziativer Gedächtnisleistung. Die Leistung jüngerer Menschen weist hier also ein äußerst untypisches Muster auf, das leider von den Autoren nicht weiter erklärt wird.

Neben dem untypischen Effekt in der nicht-degradierten Bedingung hat die leichte und moderate Degradierung keinen großen Effekt auf die Gedächtnisleistung der jüngeren Menschen. Höchstwahrscheinlich ergeben sich keine signifikanten Unterschiede zwischen den Degradierungsstufen.

Das Fazit der Autoren lautet, dass die Degradierung der Stimuli bei älteren Menschen das Assoziationsdefizit nicht weiter verstärkt und dass jüngere Menschen unter diesen Bedingungen kein Assoziationsdefizit zeigen. Leider fehlen, wie erwähnt, die Teststatistiken für diese Ergebnisse. Graef und Naveh-Benjamin ziehen aus diesem Experiment die Schlussfolgerung, dass das Assoziationsdefizit nicht durch verringerte sensorische Funktionen verursacht wird.

Dieses Fazit wird durch zwei weitere Experimente von Naveh-Benjamin, Kilb und Hara (2007) gestützt. Im ersten Experiment wurden Wörter visuell dargeboten. Dabei wurden die Schriftfarbe und die Hintergrundfarbe hinsichtlich des Farbtons, der Sättigung und der Helligkeit so aufeinander abgestimmt, dass entweder eine leichte, mittelschwere oder schwere Lesbarkeit der Stimuli resultierte. In der schweren Bedingung wurden die Wörter z.B. in einem sehr dunklen Blau vor schwarzem Hintergrund dargeboten. Diese Manipulation führte zwar bei jüngeren Menschen zu einer höheren Rate falscher Alarme im assoziativen Gedächtnistest, allerdings wurde kein signifikantes Assoziationsdefizit bei jüngeren Menschen festgestellt. Dieselbe Schlussfolgerung lässt sich aus dem zweiten Experiment ziehen. Hier wurden die Wörter akustisch dargeboten und waren von unterschiedlich lauten Hintergrundgeräuschen begleitet. Die in diesem Vortrag präsentierten Ergebnisse lassen also nicht die Schlussfolgerung zu, dass das Assoziationsdefizit älterer Menschen auf die erschwerten Wahrnehmungsbedingungen zurückzuführen ist.

Zum Abschluss dieses Kapitels soll auf ein Vortrag von Naveh-Benjamin (2008a) mit dem Titel „Assessing potential factors that mediate age-related deficits in episodic memory“ eingegangen werden. Naveh-Benjamin führte hier zunächst an, auf welchen Ursachen das Assoziationsdefizit älterer Menschen beruhen könnte. Reduzierte Aufmerksamkeitsprozesse, verlangsamte Verarbeitungsgeschwindigkeit, mangelnde Inhibitionsfähigkeiten und Abbau sensorischer Funktionen werden als „Hauptkandidaten“ genannt. Aus einem Rückblick auf die veröffentlichte Literatur (s.o.) schließt er, dass das Assoziationsdefizit keinesfalls durch reduzierte Aufmerksamkeitsprozesse erklärt werden kann. Auch sensorische Funktionen können seiner Ansicht nach das Defizit höchstens teilweise erklären. Hinsichtlich der verlangsamten Verarbeitungsgeschwindigkeit berichtete Naveh-Benjamin (leider ohne eine genaue Erläuterung der Experimente), dass auch dieser Basismechanismus das Assoziationsdefizit nicht erklären kann.¹⁶ Lediglich zu den Inhibitionsprozessen scheint noch keine empirische Evidenz zu existieren. Zusammenfassend schlussfolgert Naveh-Benjamin, dass er es für unwahrscheinlich hält, das Assoziationsdefizit älterer Menschen ohne einen spezifischen Einfluss von defizitären assoziativen Gedächtnisprozessen erklären zu können. Auch im persönlichen Gespräch (Naveh-Benjamin, 2008b) bestätigt Naveh-Benjamin, dass er das Assoziationsdefizit nicht als Folge der bisher postulierten Basismechanismen ansieht, sondern einen weiteren grundlegenden assoziativen Gedächtnisprozess für notwendig hält, um das Assoziationsdefizit zu erklären. Diesen assoziativen Gedächtnisprozess würde Naveh-Benjamin auf einer Ebene mit den etablierten Basismechanismen anordnen.

¹⁶Um den Anteil von Verarbeitungsgeschwindigkeit an assoziativen Gedächtnisprozessen zu erfassen, wurden von Salthouse (1995, 1993) zwei Studien durchgeführt. Da die assoziativen Aufgaben, die in diesen Studien verwendet wurden, nicht vergleichbar sind mit assoziativen Rekognitionsaufgaben, wird auf diese Studien nicht weiter eingegangen. Des Weiteren wurde in den Studien von Salthouse auch kein Vergleich zwischen assoziativer Gedächtnisleistung und Item-Gedächtnisleistung hergestellt, so dass sich auch aus diesem Grund die Studien nur schwer mit den Experimenten im Kontext der Assoziationsdefizit-Hypothese vergleichen lassen.

Wie bereits zu Beginn des Kapitels erwähnt, wurde bisher nur in wenigen Studien der Frage nachgegangen, auf welchen Ursachen das Assoziationsdefizit beruht, und die meisten der durchgeführten Studien lassen sich zudem methodisch kritisieren. Wenn man die Befunde daher zunächst vorsichtig als Indikatoren ansieht, so sprechen diese - genau wie die Einschätzung erfahrener Forscher des Gebietes - im Moment dafür, dass ein Teil der Altersvarianz auf einen spezifischen assoziativen Gedächtnisprozess zurückzuführen ist. Naveh-Benjamin postuliert, diesen assoziativen Gedächtnisprozess als weiteren Basismechanismus anzusehen. Sicherlich muss diese Einschätzung noch durch weitere überzeugende empirische Evidenz untermauert werden, diese Arbeit schließt sich trotzdem der Position von Naveh-Benjamin an und geht von defizitären assoziativen Basismechanismen aus.

3.2.5 Das assoziative Defizit aus der Perspektive von Lokalisationsansätzen

Das Assoziationsdefizit wurde in den letzten beiden Kapiteln aus der theoretischen Perspektive qualitativ unterschiedlicher Verarbeitungsprozesse und aus der Perspektive veränderter Basismechanismen betrachtet. Dabei stellte sich heraus, dass sowohl strategische Altersdefizite als auch reduzierte assoziative Gedächtnisprozesse an dem Assoziationsdefizit beteiligt sind und es daher als multikausales Altersphänomen angesehen werden sollte. In einem letzten Schritt soll das Assoziationsdefizit nun aus der Perspektive des Lokalisationsansatzes betrachtet werden. Dabei wird der Frage nachgegangen, ob das Assoziationsdefizit auf Probleme in der Enkodierungsphase oder auf Probleme beim Abruf der Information zurückzuführen ist.

Diese Frage wurde in einigen Arbeiten bereits aufgeworfen (z.B. Cohn et al., 2008; Dew et al., 2007; Naveh-Benjamin, 2000; Naveh-Benjamin, Hussain et al., 2003), allerdings immer mit dem Hinweis, dass dieser Aspekt bisher nicht klar beantwortet werden konnte. So schreiben z.B. Naveh-Benjamin, Hussain et al. (2003, S. 835): „Despite the support for an associative deficit [...] there are some issues and questions left open regarding different aspects of an associative deficit hypothesis. One question is whether such a deficit shown by older people stems mostly from problems in creating new associations during encoding or from problems in retrieving such associations at test.“ Auch Cohn et al. (2008) stellen nach einem Überblick über den aktuellen Forschungsstand fest: „Together the reviewed findings indicate that the age-related deficit in associative memory may be caused as much by faulty retrieval mechanisms as by faulty binding at encoding [...]“ (Cohn et al., 2008, S.94)

An dieser Stelle wird daher der Frage nachgegangen, ob verringerte assoziative Gedächtnisprozesse zu einer mangelnden Bindung der beiden Elemente während der Enkodierungsphase führen, oder ob die Elemente zwar während der Enkodierung miteinander assoziiert werden, es aber nicht gelingt, diese Verbindung während der Testphase abzurufen. Zunächst werden die wenigen Ansätze beschrieben, die bisher zu dieser Fragestellung

vorliegen. Danach wird auf die Grenzen und Probleme dieser Studien eingegangen und alternative Herangehensweisen werden vorgestellt.

Naveh-Benjamin, Hussain et al. (2003) folgern aus den empirischen Befunden, die bisher zur Assoziationsdefizit-Hypothese vorliegen, dass das Defizit sowohl in der Enkodierungsphase als auch in der Testphase zu verorten sei. Als Argumente für diese Schlussfolgerung führen sie an, dass Manipulationen während der Lernphase (z.B. inzidentelle vs. intentionale Lernbedingungen) sowie Manipulationen während der Testphase (z.B. freie Reproduktion vs. Rekognition) einen signifikanten Einfluss auf das Assoziationsdefizit haben. Wie bereits aufgezeigt wurde, muss ein Alterseffekt, der durch Testbedingungen hervorgerufen wird, nicht unbedingt auf ein *assoziatives* Abrufdefizit zurückgehen. Auch ein Einfluss von Enkodierungsmanipulationen auf das Erinnerungsniveau kann nicht als klare Evidenz für ein Enkodierungsdefizit interpretiert werden. Es könnte durchaus der Fall sein, dass strategische Hinweise in der Lernphase zwar die Enkodierungsleistung verbessern, sich aber auch auf die Abrufleistung positiv auswirken, da diese Strategien auch in der Abrufphase genutzt werden. Daher schlussfolgert auch Naveh-Benjamin, dass diese Befunde nur als Hinweise gedeutet werden können und weitere Forschung zu dieser Frage notwendig ist.

In der Studie von Chalfonte und Johnson (1996) wurde ebenfalls diskutiert, ob das Assoziationsdefizit auf Enkodierungs- oder Abrufprobleme zurückzuführen ist. Daher haben Mitchell et al. (2000) in einem Nachfolgeexperiment mit vergleichbaren Materialien und einem ähnlichen experimentellen Ablauf untersucht, ob das Assoziationsdefizit auch bei Kurzzeitgedächtnisaufgaben auftritt. Die zugrundeliegende Idee war, dass ein Assoziationsdefizit bei Kurzzeitgedächtnisaufgaben auf ein Altersdefizit in der Enkodierungsphase hinweisen könnte, da Abrufprozesse in diesen Aufgaben eher zu vernachlässigen sind.

Um diese Hypothese zu testen, wurde den Teilnehmern für eine Sekunde ein Quadrat mit neun Feldern gezeigt, wobei in einem der Felder ein Objekt zu sehen war (vgl. Chalfonte & Johnson, 1996, Abbildung 3.1). Insgesamt wurden pro Durchgang nacheinander drei Objekte in jeweils unterschiedlichen Feldern präsentiert. In der Testphase, die 800ms nach der letzten Präsentation folgte, mussten die Teilnehmer entscheiden, ob das Testobjekt in diesem Durchgang präsentiert wurde, oder ob es sich um ein Distraktorobjekt handelte (Objekttest). Im Kombinationstest mussten sie zusätzlich ihr Urteil über die Position des Objekts abgeben. Die Ergebnisse sind vergleichbar mit den Ergebnissen von Chalfonte und Johnson (1996). Mitchell et al. (2000) finden keinen Altersunterschied im Test der einzelnen Elemente, allerdings einen signifikanten Alterseffekt in der Kombinationsaufgabe (siehe aber auch Brockmole et al., 2008). Aus der Tatsache, dass sich ein Assoziationsdefizit auch bei Kurzzeitgedächtnisaufgaben zeigt, schlussfolgern die Autoren, dass Enkodierungsprozesse am Assoziationsdefizit beteiligt sind. Inwieweit sich der gesamte Alterseffekt auf ein reines Enkodierungsdefizit zurückführen lässt, kann anhand der Daten von Mitchell et al. (2000) nicht beantwortet werden und wird in diesem Kontext als

offene Fragestellung diskutiert. Ein Vorschlag, wie man die Beteiligung von Abrufprozessen untersuchen könnte, wird von den Autoren nicht gemacht.

Konkrete Vorstellungen darüber, wie man die Beteiligung von Enkodierungs- und Abrufprozessen experimentell untersuchen kann, unterbreiten Cohn et al. (2008). Sie postulieren, dass drei unterschiedliche Ursachen für das Assoziationsdefizit älterer Menschen in Frage kommen: (1) Defizitäre Bindungsprozesse während der Enkodierungsphase, (2) eine allgemeine, altersbedingte schlechtere Abrufleistung oder (3) ein spezifisches Defizit bei sogenannten strategischen bzw. aufwendigen Abrufprozessen, wie sie in einem assoziativen Rekognitionstest notwendig sind. Hinsichtlich der Abrufprozesse unterscheiden sie zwischen ‚recall-to-accept‘, einem Gedächtnisprozess der der Annahme alter intakter Items zugrunde liegt und durch Vertrautheit unterstützt wird, und ‚recall-to-reject‘, einem Prozess, der für die Zurückweisung neuer Kombinationen notwendig ist und bewusste Erinnerungsprozesse erfordert. Die Autoren nehmen dabei an, dass Altersunterschiede in den aufwendigeren Prozessen (‚recall-to-reject‘) eher zu Tage treten sollten als bei weniger aufwendigen Prozessen (‚recall-to-accept‘; vgl. selbstinitiierte Prozesse bei Craik, Kap. 3.2.4). Sie postulieren mit dem Prozess ‚recall-to-accept‘ das allgemeine Abrufdefizit zu erfassen, mit ‚recall-to-reject‘ das strategische Abrufdefizit.

Cohn et al. (2008) versuchten über zwei unterschiedliche Aufgaben diese Prozesse zu trennen. Sie verwendeten zum einen ‚assoziative Identifikationsaufgaben‘ (‚associative identification‘) zum anderen ‚assoziative Wiederherstellungsaufgaben‘ (‚associative reinstatement‘). In assoziativen Identifikationsaufgaben werden alte intakte Wortpaare dargeboten sowie neu kombinierte Wortpaare aus zwei gelernten Items. Vergleichbar zu dem traditionellen assoziativen Rekognitionstest werden nur alte Items präsentiert. Die Probanden müssen alte intakte Wortpaare als alt identifizieren und neu kombinierte Wortpaare zurückweisen. Um neu kombinierte Wortpaare zurückzuweisen, ist der Prozess des ‚recall-to-reject‘ notwendig, Itemvertrautheit kann nicht herangezogen werden.

In einem zweiten Test, den die Autoren assoziativen Wiederherstellungstest nennen, werden alte intakte Wortpaare, neu kombinierte Wortpaare sowie Paare aus Target-Distraktor- bzw. Distraktor-Target-Kombinationen dargeboten. In diesem Test werden also auch neue Wörter in die Testphase aufgenommen. Die Aufgabe der Versuchsperson ist es, alle Wortpaare als alt zu beurteilen, die aus zwei alten Wörtern bestehen. Eine „Alt“-Antwort wäre also sowohl auf alte intakte Wortpaare als auch auf neu kombinierte Wortpaare korrekt. Als neu hingegen müssen die Wortpaare zurückgewiesen werden, die ein neues Wort beinhalten, also Target-Distraktor- oder Distraktor-Target-Kombinationen. Die Autoren nehmen an, dass für diese Aufgabe kein Prozess des ‚recall-to-reject‘ notwendig sind, sondern bereits weniger ressourcenfordernde Prozesse (‚recall-to-accept‘) zu einer korrekten Entscheidung führen können. Cohn et al. (2008) stellten nun folgende Hypothesen auf:

- (1) Defizitäre Bindungsprozesse während der Enkodierungsphase sollten zu Alterseffekten in beiden assoziativen Aufgaben führen.
- (2) Ein allgemeines Abrufdefizit sollte sich ebenfalls in beiden Aufgaben niederschlagen.
- (3) Ein strategisches Altersdefizit für aufwendige Abrufprozesse sollte dahingegen nur in den assoziativen Identifikationsaufgaben zu einer schlechteren Leistung führen.

Bevor auf die Ergebnisse eingegangen wird, soll an dieser Stelle bereits angemerkt werden, dass die erste Hypothese m.E. nicht nachvollziehbar ist. Die Autoren begründen nicht überzeugend, warum für die korrekte Bearbeitung der assoziativen Wiederherstellungsaufgabe überhaupt Bindungsprozesse notwendig sein sollten. Meiner Ansicht nach ist die Erinnerung an einzelne Items ausreichend, um alte intakte Wortpaare sowie neu kombinierte Wortpaare von Target-Distraktor-Kombinationen zu unterscheiden. Natürlich *kann* die Erinnerung an die Kombination der Wortpaare dabei helfen, eine Target-Distraktor-Kombination als solche zu identifizieren. Notwendig ist eine Erinnerung an die Kombination aber nicht. Aus diesem Grund kann man m.E. die erste Hypothese nicht über den Vergleich dieser beiden Aufgaben untersucht werden.

Trotz der Vorbehalte soll kurz auf die Ergebnisse eingegangen werden. Die Autoren fanden keinen Altersunterschied in der gerade diskutierten assoziativen Wiederherstellungsaufgabe, während in der assoziativen Identifikationsaufgabe ein Altersunterschied vorhanden war. Sie interpretieren dieses Ergebnis dahingehend, dass weder ein Enkodierungsdefizit noch ein allgemeines Abrufdefizit als potentielle Ursachen für das Assoziationsdefizit in Frage kommen, und schlussfolgern, dass ein spezifisches Abrufdefizit bei ‚recall-to-reject‘ die Ursache des Assoziationsdefizit darstellt. Dieses Ergebnismuster kann aber auch durch eine intakte Itemerkennung älterer Menschen erklärt werden, bei der sich die Senioren auf Vertrautheitsprozesse stützen können und nicht notwendigerweise auf bewusste Erinnerung angewiesen sind.

In einem zweiten Experiment versuchten die Autoren den Alterseffekt in strategischen Abrufprozessen und das allgemeine Abrufdefizit genauer zu untersuchen. Die Ergebnisse aus diesem Experiment konnten ein Altersdefizit in beiden Prozessen aufzeigen, allerdings war der Altersunterschied bei strategischen Abrufprozessen stärker ausgeprägt. In Hinblick auf die Abrufprozesse gehen die Autoren daher zusammenfassend von einem allgemeinen Abrufdefizit älterer Menschen aus (Exp. 1). Dieses Abrufdefizit verstärkt sich aber für aufwendige Prozesse (‚recall-to-reject‘; Exp. 2).

Während von Mitchell et al. (2000) also ein Enkodierungsdefizit unterstützt wird, sehen Cohn et al. (2008) ein (allgemeines wie strategisches) Abrufdefizit als Hauptursache des Assoziationsdefizits an und schließen ein Enkodierungsdefizit aus. Leider diskutieren Cohn et al. (2008) ihre Ergebnisse nicht vergleichend zu den Ergebnissen von Mitchell et al. (2000), daher bleibt es dem Leser überlassen, die unterschiedlichen Resultate zu bewerten.

Meiner Ansicht nach lässt das experimentelle Ergebnis, aufgrund dessen ein Enkodierungsdefizit von Cohn et al. (2008) zurückgewiesen wird, auch alternative Interpretationen zu. Der Nachweis eines Assoziationsdefizits bei Kurzzeitgedächtnisaufgaben dahingegen weist plausibel auf ein Assoziationsdefizit während der Enkodierungsphase hin. Zwar ist nicht geklärt, ob dieses Defizit auf assoziative Basisprozesse oder auf strategische Alterseffekte zurückgeht, dass aber Enkodierungsprozesse eine relevante Rolle im Assoziationsdefizit spielen, kann durch die Arbeiten von Mitchell et al. (2000) belegt werden. Auch in Bezug auf ein Abrufdefizit ist die Studie von Cohn et al. (2008) nicht wirklich überzeugend. Die assoziativen Identifikationsaufgaben und die assoziativen Wiederherstellungsaufgaben unterscheiden sich nicht *nur* in den assoziativen Abrufprozessen, sondern auch in Bezug auf das Itemgedächtnis. Der ausbleibende Altersunterschied bei assoziativen Wiederherstellungsaufgaben kann daher auch durch ein intaktes Itemrekognitionsgedächtnis älterer Menschen erklärt werden und bedarf nicht unbedingt des Rückgriffs auf unterschiedliche Abrufprozesse. Daher fasse ich an dieser Stelle die empirische Evidenz wie folgt zusammen: Durch die Arbeiten von Mitchell et al. (2000) konnte gezeigt werden, dass defizitäre Enkodierungsprozesse an dem Assoziationsdefizit beteiligt sind; inwieweit zusätzlich auch Abrufprozesse zu dem Assoziationsdefizit beitragen, konnte bisher empirisch nicht überzeugend beantwortet werden.

Die Lokalisationsfrage des assoziativen Defizits stellt eine interessante und wichtige Forschungsfrage für zukünftige Forschungsarbeiten dar. Dies kann nicht nur aus Sicht der Grundlagenforschung so eingeschätzt werden, auch für die angewandte Forschung ergeben sich aus einer Klärung dieser Frage wichtige Erkenntnisse. So ist damit z.B. eine direkte Konsequenz für den Ansatzpunkt von Trainings- oder Interventionsstudien gegeben. Die Diskussion über die bisherigen Herangehensweisen an das Problem (Cohn et al., 2008; Naveh-Benjamin, Hussain et al., 2003) haben damit gezeigt, dass ein rein designbasiertes Vorgehen hier an seine Grenzen stößt, da das Dekomponierungsproblem von Enkodierungs- und Abrufprozessen nicht klar beantwortet werden konnte.

In der vorliegenden Arbeit soll daher das Dekomponierungsproblem aufgegriffen und ein Lösungsansatz erarbeitet werden. Damit wird der methodische Weg zur Beantwortung der Lokalisationsfrage des assoziativen Defizits geebnet. Weiterhin wird der Lösungsansatz in einer Altersstudie angewandt. Es kann nicht erwartet werden, dass mit einer einzigen Altersstudie die Lokalisationsfrage des assoziativen Defizits abschließend geklärt werden kann. Dies soll auch nicht als Ziel dieser Arbeit missverstanden werden. Das Ziel der vorliegenden Arbeit ist es, einen Weg aufzuzeigen, wie man die Lokalisationsfrage beantworten und das Dekomponierungsproblem angehen kann.

In diesem Kapitel wurde ein Überblick über das assoziative Defizit, die bisherige Forschung dazu sowie die offenen Forschungsfragen gegeben. Es stellte sich heraus, dass das assoziative Defizit in zahlreichen Studien repliziert werden konnte und als grundlegen-

der Basismechanismus an einer Vielzahl von kognitiven Aufgaben beteiligt ist. Weiterhin konnte aufgezeigt werden, dass es sich um ein multikausales Altersdefizit handelt, an dem sowohl strategische Komponenten als auch assoziative Gedächtnisprozesse beteiligt sind.

Trotz einiger Versuche, das Assoziationsdefizit genauer auf ein Enkodierungs- oder Abrufdefizit einzuschränken, blieb diese Lokalisationsfrage bisher unbeantwortet. Dies ist unter anderem auch dadurch begründet, dass es sich hierbei um ein Dekomponierungsproblem handelt (siehe z.B. Erdfelder, 2000). Im nächsten Kapitel wird daran angeknüpft, und die Dekomponierungsproblematik wird genauer erläutert. Auch werden Lösungsansätze diskutiert, und deren Anwendung auf die vorliegende Lokalisationsfrage des assoziativen Defizits wird aufgezeigt.

Kapitel 4

Methodische Aspekte zur Trennung von Enkodierung und Abruf

Im vorangehenden Kapitel wurde die Lokalisationsfrage des assoziativen Defizits als zentrale Fragestellung dieser Arbeit formuliert. In diesem Kapitel wird zunächst herausgearbeitet, warum es sich bei der Lokalisationsfrage um ein Dekomponierungsproblem handelt und wie dieses gelöst werden kann. Dabei werden die Nachteile designbasierter Ansätze und die Vorteile von mathematischen Modellen verdeutlicht (Kapitel 4.1). In Kapitel 4.2 werden mehrere Modelle angesprochen, die es ermöglichen, Enkodierungs- und Abrufprozesse über getrennte Parameter zu schätzen. Auch wird auf Modelle eingegangen, die assoziative Prozesse in Rekognitionsparadigmen erfassen. Dabei werden zum einen das Enkodierung-Abruf-Modell (Batchelder & Riefer, 1980, 1986), zum anderen Quellengedächtnismodelle (Batchelder & Riefer, 1990; Bayen et al., 1996; Meiser & Bröder, 2002) genauer vorgestellt (Kapitel 4.2.1 und 4.2.2). Aus diesen beiden Modellen bzw. Modellklassen werden zentrale Annahmen übernommen und in ein neues multinomiales Modell zur Untersuchung des assoziativen Defizits integriert. Dieses Modell, das zugrundeliegende empirische Paradigma sowie die spezifischen Modellannahmen und Parameter werden im Kapitel 4.2.3 ausgeführt. Eine empirische Überprüfung des Modells wird dann im empirischen Teil der Arbeit folgen. Einige grundlegende Annahmen und Voraussetzungen für die multinomiale Modellierung werden anhand des Enkodierung-Abruf-Modells ausgeführt, für einen umfassenden Überblick und einen detaillierten Hintergrund zur multinomialen Modellierung wird allerdings auf Erdfelder (2000) verwiesen.

4.1 Die Lokalisationsfrage als Dekomponierungsproblem

Die Lokalisationsfrage des assoziativen Defizits bezieht sich auf die psychologischen Prozesse Enkodierung/Speicherung und Abruf, die sich der direkten Beobachtbarkeit entziehen. Erfasst werden kann lediglich die Antwort der Versuchsperson, d.h. das Endergebnis eines kognitiven Prozesses bzw. das Endergebnis der Verkettung mehrerer kognitiver Prozesse. Offen bleibt, welche Prozesse zu diesem Antwortverhalten geführt haben. Die empirisch erfasste Antwort ist also kein prozessreiner Indikator für ein bestimmtes theoretisches Konstrukt. Damit handelt es sich bei dieser Frage um ein ‚Dekomponierungsproblem‘ oder auch ‚Separierungsproblem‘ (Erdfelder, 2000). Ein solches Dekomponierungsproblem liegt einer Reihe von empirischen Indikatoren zugrunde (zusammenfassend siehe Erdfelder, 2000):

- (1) den Urteilen einer Versuchsperson in einem 2AFC-Rekognitionstest (hier gilt es, Erinnerungsprozesse und Rateprozesse zu separieren),
- (2) den Antworten eines Probanden in Aufgaben zum Rückschaufehler (hier fließen Gedächtnisprozesse, Urteilsprozesse und Rateprozesse zusammen),
- (3) den Reaktionen in prospektiven Gedächtnisaufgaben (automatische und kontrollierte Überwachungsprozesse sowie retrospektive Gedächtniskomponenten müssen in diesem Kontext dekomponiert werden),
- (4) Antworten auf logische Denkprobleme wie z.B. die Wason-Aufgabe (Rateprozesse und Prozesse des logischen Schlussfolgerns können hier das Antwortverhalten bestimmen).

In all diesen beispielhaft genannten Forschungsfeldern liegt das Problem vor, dass die Antworten der Teilnehmer auf ganz unterschiedliche kognitive Prozesse zurückgeführt werden können und es daher von Interesse ist, diese Prozesse getrennt voneinander zu erfassen. Insbesondere für die Anwendung dieser Forschungsparadigmen in der Entwicklungspsychologie ist die Dekomponierungsproblematik von Bedeutung, denn in diesem Bereich liegen oft unterschiedliche Hypothesen für verschiedene spezifische Prozesse vor. So werden z.B. Alterseffekte in Gedächtnisprozessen erwartet, im Hinblick auf Rateprozesse sollten die Gruppen aber vergleichbar sein.

Im Rahmen des Assoziationsdefizits steht man vor der Herausforderung, den vorhandenen Altersunterschied aufgrund des Dekomponierungsproblems nicht genauer identifizieren zu können. Differenzen im Antwortverhalten jüngerer und älterer Menschen könnten aufgrund von Unterschieden in der assoziativen Enkodierung, im assoziativen Abruf oder aufgrund von verschiedenen Ratetendenzen zustande kommen. Welche Prozesse einer Antwort zugrunde liegen und in welchen Prozessen Altersunterschiede auftreten, entzieht sich der direkten Beobachtung.

Prinzipiell sind zwei Herangehensweisen an Dekomponierungsprobleme vorgeschlagen worden: designbasierte und modellbasierte Ansätze (Brainerd, 1985). In designbasierten Ansätzen versucht man, Störeinflüsse auszuschalten oder konstant zu halten, in modellbasierten Ansätzen werden diese in ein stochastisches Modell aufgenommen und zusammen mit den anderen relevanten Prozessen geschätzt.

Als Beispiel für einen *designbasierten Ansatz* kann unter anderem das von Naveh-Benjamin (2000) verwendete Paradigma des Vergleichs von Item- und assoziativem Rekognitionstest angeführt werden. Das Ziel von Naveh-Benjamin (2000) war es zu untersuchen, inwieweit ältere Menschen ein spezifisches Gedächtnisdefizit für assoziative Informationen aufweisen, welches über das Gedächtnisdefizit für Items hinausgeht. Um diese Frage zu klären, muss man die Prozesse der assoziativen Erinnerungsleistung, des Itemgedächtnisses und Rateprozesse separieren. Naveh-Benjamin (2000) wählte daher zwei unterschiedliche Tests aus: einen Test, der vornehmlich Itemerinnerung sowie Rateprozesse erfasst (2AFC-Test), und einen Test für assoziative Gedächtnisprozesse sowie Rateprozesse (assoziativer Rekognitionstest). Unter der Annahme, dass Ratetendenzen zwischen den beiden Tests nicht variieren, führt er Altersunterschiede in den jeweiligen Tests auf Item- bzw. assoziative Gedächtnisleistungen zurück. Als typisches Ergebnismuster findet er einen Alterseffekt in der assoziativen Rekognition, in der Itemrekognition ist kein Altersunterschied festzustellen. Aus diesem Muster schlussfolgert der Autor, dass ältere Menschen ein spezifisches assoziatives Gedächtnisdefizit aufweisen, welches über ein Defizit in der Itemrekognition hinausgeht. Über das Design wird versucht, bestimmte kognitive Prozesse konstant zu halten (in diesem Fall Ratetendenzen), um aus den Ergebnissen Schlussfolgerungen im Hinblick auf andere Prozesse zu ziehen.

Auch das in der Altersforschung typische Prozedere unterschiedlicher Darbietungszeiten für die beiden Altersgruppen (Rabinowitz et al., 1982; für einen Überblick siehe Salthouse, 1982) kann als ein Beispiel des designbasierten Lösungsansatzes angeführt werden. Dies kann ebenfalls anhand eines Experiments von Naveh-Benjamin illustriert werden. Da der Autor Altersunterschiede in assoziativen Gedächtnisprozessen erfassen möchte, die über ein Defizit im Itemgedächtnis hinausgehen, würde es seine Schlussfolgerungen vereinfachen, wenn Altersunterschiede in der Itemerinnerung ausbleiben würden. Daher verwendet Naveh-Benjamin unterschiedliche Präsentationszeiten für jüngere und ältere Menschen. In der Studie von Naveh-Benjamin, Guez und Shulman (2004) z.B. bekamen die jüngeren Menschen die Wortpaare jeweils für 5 Sekunden präsentiert, die älteren Menschen hatten für die Enkodierung 8 Sekunden Zeit. Dieses Vorgehen wird damit gerechtfertigt, dass eine solche Manipulation die ohnehin geringen Altersunterschiede in der Itemrekognition eliminiert und damit den Vergleich von Item- und assoziativer Gedächtnisleistung vereinfacht. Auch in anderen Arbeiten der Altersforschung wird dieser Ansatz verwendet, um die Altersunterschiede in der Itemerinnerung als konfundierte Variable konstant zu halten oder in Bezug auf diese Gedächtniskomponente ein vergleichbares Ausgangsniveau der beiden Altersgruppen herzustellen.

Insbesondere das zweite Beispiel zeigt durch seinen weit verbreiteten kritiklosen Einsatz auf, dass designbasierte Ansätze im Allgemeinen akzeptiert und oftmals auch unkritisch angewandt werden. Problematisch sind an diesem Vorgehen mehrere Aspekte. Zum einen wird ein wichtiges experimentelles Prinzip verletzt: Für die (in den Quasi-Experimenten der Altersforschung nicht randomisierten) Altersgruppen liegen keine vergleichbaren Bearbeitungsbedingungen mehr vor. Es bleibt unklar und unkontrolliert, auf welche Faktoren sich diese Unterschiede auswirken. Es mag möglich sein, dass dieser Faktor wie intendiert nur zu einer besseren Enkodierungsleistung der älteren Menschen führt und damit lediglich die Gedächtnisleistung positiv beeinflusst. Die Enkodierungszeit könnte sich aber auch auf andere Prozesse auswirken und damit nicht *nur* den beabsichtigten Effekt hervorrufen. Daher ist m.E. das Verfahren - trotz der hohen Verbreitung - als problematisch anzusehen.¹⁷

Weitere Kritikpunkte werden u.a. bei Brainerd (1985) diskutiert. Der Autor weist z.B. darauf hin, dass designbasierte Lösungen vom Ansatz her zum Scheitern verurteilt sind, da sie kein zugrundeliegendes Messmodell für die Beziehung von latenten Konstrukten und gemessenen Variablen spezifizieren, geschweige denn dass dieses Messmodell überprüft würde. Brainerd (1985) illustriert dies an einigen Beispielen, die eindrucksvoll aufzeigen, von welcher Bedeutung diese Problematik ist (für Beispiele in der gerontologischen Anwendung siehe auch Bayen, 1990).

Auf weitere konkrete Beispiele sowie auf eine ausführlichere Diskussion der Problematik soll an dieser Stelle verzichtet werden, da im vorangehenden Kapitel (3.2.5) bereits einige designbasierten Lösungsansätze zur Lokalisationsfrage des assoziativen Defizits geschildert wurden, u.a. die Idee, aus einer Leistungssteigerung durch strategische Bearbeitungshinweise auf ein Enkodierungsdefizit zu schließen. Auch hier liegen die Kritikpunkte auf der Hand bzw. wurden bereits diskutiert (siehe Kapitel 3.2.5). Festgehalten werden kann in jedem Fall, dass die Lokalisationsfrage bisher über designbasierte Ansätze nicht beantwortet werden konnte, und auch zukünftige designbasierte Ansätze werden die allgemeine Kritik an diesem Verfahren nur schwer umgehen können.

Die Alternative zu designbasierten Ansätzen besteht in *modellbasierten Lösungsverfahren*. Bei diesem Vorgehen werden alle relevanten, an der Aufgabe beteiligten Prozesse (also auch Rateprozesse oder Fehlereinflüsse) als Parameter in ein entsprechendes stochastisches Modell aufgenommen. Anhand der resultierenden Parameterschätzungen kann dann geschlussfolgert werden, welche Prozesse in welchem Ausmaß an der Aufgabe beteiligt sind und inwieweit spezifische Altersunterschiede vorliegen. Als Beispiel für ein modellbasiertes Lösungsverfahren kann das bereits geschilderte Quellengedächtnismodell von Bayen et al. (1996) angeführt werden. Mit Hilfe dieses multinomialen Modells kön-

¹⁷Der Kritik an diesem konkreten Beispiel der Präsentationszeit kann dadurch begegnet werden, dass die Probanden auf mehrere experimentelle Gruppen aufgeteilt werden: In einer Bedingung müssen beide Altersgruppen die Aufgaben unter identischen Rahmenbedingungen absolvieren. Eine Bedingung sollte die Bearbeitung für ältere erleichtern, eine die für jüngere (siehe z.B. Spaniol & Bayen, 2005).

nen zum einen die Prozesse der Itemerinnerung, Quellenerinnerung und Antworttendenzen dekomponiert werden, zum anderen können Altersunterschiede klar auf diese spezifischen Prozesse attribuiert werden. Auch für das Dekomponierungsproblem der Prozesse Enkodierung, Speicherung und Abruf wurden bereits modellbasierte Lösungen vorgeschlagen und erfolgreich angewendet. In der kognitiven Altersforschung haben diese Ansätze vereinzelt Einzug gefunden. Im nächsten Kapitel sollen daher die bisher existierenden modellbasierten Zugänge zur Trennung der Prozesse Enkodierung, Speicherung und Abruf dargestellt werden. Das Ziel ist dabei keine detaillierte Schilderung jedes einzelnen Modells. Vielmehr soll ein Überblick über die Modelle gegeben werden, und vor allem gilt es zu klären, inwieweit die existierenden Ansätze dazu genutzt werden können, das vorliegende Dekomponierungsproblem der Lokalisationsfrage des assoziativen Defizits zu klären.

4.2 Mathematische Modelle

Nachdem die Lokalisationsfrage als Dekomponierungsproblem eingeordnet wurde und gleichzeitig aufgezeigt werden konnte, dass designbasierte Ansätze ungeeignet zur Lösung dieses Problems sind, soll im Folgenden auf modellbasierte Ansätze und deren Potential eingegangen werden.

Spaniol und Bayen (2004) geben einen Überblick über formale Modelle, die in der Altersforschung Anwendung finden. Besonders relevant für die vorliegende Forschungsfrage ist ihr Kapitel über Modelle zur Trennung von Enkodierungs- und Abrufprozessen. Hier werden Ansätze berichtet, die zur Dekomponierung dieser beiden Prozesse in unterschiedlichen Paradigmen genutzt werden können, sowie die Anwendung dieser Modelle in der Altersforschung. Unabhängig vom gerontologischen Einsatz werden zusammenfassend auch bei Batchelder und Riefer (1999) sowie Erdfelder (2000) Modelle diskutiert, die eine Dekomposition von Enkodierung und Abruf ermöglichen.

Interessant ist, dass sowohl Batchelder und Riefer (1999) als auch Erdfelder (2000) von der Trennung zwischen *Speicherung und Abruf* sprechen, Spaniol und Bayen (2004) dahingegen von *Enkodierung und Abruf*. Die Tatsache, dass einmal Enkodierungs- das andere Mal Speicherprozesse gegenüber Abrufprozessen abgegrenzt werden ist darin begründet, dass die meisten Modelle zwei Parameter schätzen (Batchelder & Riefer, 1980; Chechile & Meyer, 1976; Riefer & Rouder, 1992; siehe aber Bäuml, 1996; Greene, 1996; Howe, 1988; Wilkinson & Koestler, 1983). Diese Modelle trennen zwar Abrufprozesse von Enkodierungs- bzw. Speicherprozessen, die Dekomponierung von Enkodierung und Speicherung wird allerdings nicht vorgenommen. Diese Prozesse können weiterhin nicht separiert werden. Aufgrund dieser Zusammenfassung von Enkodierungs- und Speicherprozessen bleibt die Benennung dieses (Hybrid-)Parameters mehr oder weniger offen.

Trotzdem scheint es kein Zufall zu sein, dass ausgerechnet Spaniol und Bayen (2004) im Kontext der gerontologischen Forschung von Enkodierungsprozessen sprechen. Zwar

wird dies nicht explizit begründet, jedoch ist zu vermuten, dass dies mit der Zielsetzung ihres Artikels zusammenhängt. Sie berichten mathematische Modelle dezidiert im Hinblick auf deren Anwendung in der Altersforschung. In diesem Forschungsfeld hat sich schon relativ früh herausgestellt, dass ältere Menschen (abhängig von der Gedächtnisaufgabe) Enkodierungs- und/oder Abrufdefizite zeigen, Speicherungsdefizite konnten allerdings nicht nachgewiesen werden (Salthouse, 1982, für einen Überblick siehe Bayen, 1990; Salthouse, 1992). Da sich keine empirische Evidenz für eine Speicherdefizithypothese, fand ist für die gerontologische Forschung vor allem die Unterscheidung zwischen Enkodierung und Abruf relevant. Dies spiegelt sich auch darin wider, dass z.B. im Kontext der Assoziationsdefizit-Hypothese die Frage aufgeworfen wurde, ob dieses Defizit auf die Enkodierungs- oder Abrufphase zurückzuführen ist. In keiner mir bekannten Arbeit wurde die Frage nach einem Speicherdefizit gestellt. Dies könnte erklären, warum Spaniol und Bayen (2004) von Enkodierung und im Gegensatz zu anderen Arbeiten nicht von Speicherung sprechen. Letztendlich bezieht sich dies aber nur auf die Bezeichnung des Parameters, denn in beiden Fällen werden Enkodierungs- und Speicherprozesse zusammengefasst.

Im Folgenden wird ein knapper Überblick über Modelle gegeben, die eine Trennung von Enkodierungs- Speicherungs- und Abrufprozessen vorsehen. Dieser Überblick ist dabei an die oben zitierten Arbeiten (Batchelder & Riefer, 1999; Erdfelder, 2000; Spaniol & Bayen, 2004) angelehnt. Danach wird herausgearbeitet, welche Modelle für die Anwendung auf die vorliegende Forschungsfrage besonders relevant sind und inwieweit diese auf das Paradigma von Naveh-Benjamin (2000) angewandt werden können. Detailliert besprochen werden nur die für diese Arbeit und das assoziative Lokalisationsproblem zentralen Modelle.

Die ersten gerontologischen Studien, die eine Trennung von Enkodierungs- und Abrufprozessen vornahm, werden von Howe und Kollegen (Howe, 1988; Howe & Hunter, 1985, 1986) berichtet. Sie wendeten das ‚Zwei-Phasen-Markoff-Ketten-Modell‘ von Greeno (1968) an, welches für ein Paradigma der wiederholten Präsentation von Items sowie des wiederholten Abrufs erstellt wurde. Dieses Modell erlaubt die Schätzung von Enkodierungsparametern (a , a'), einem Speicherungsparameter (f) sowie mehreren Abrufparametern. In der gerontologischen Anwendung konnte anhand dieses Modells die Speicherdefizithypothese genauer untersucht und zurückgewiesen werden. Ebenfalls zur Klasse der Markoff-Modelle gehört der Modellierungsansatz von Wilkinson und Koestler (1983). Dieses Modell kann auf Daten des wiederholten Abrufs nach einmaliger Präsentation angewandt werden. Wie das vorherige Modell werden Enkodierungs-, Speicherungs- sowie Abrufparameter (s , f , r) geschätzt. Auch hier konnte in der gerontologischen Anwendung kein Speicherdefizit nachgewiesen werden. Ein weiteres Modell für dasselbe Paradigma stammt von Howe und Brainerd (1989): Hier werden die Prozesse des Spurenerfalls, der Spure reintegration, des Abruflernens und Abrufvergessens voneinander getrennt. Nicht nur für den wiederholten Abruf, sondern auch für den einmaligen Abruf existieren Markoff-

Modelle, die einzelne Parameterschätzungen für Enkodierung, Speicherung sowie Abruf ermöglichen (Bäuml, 1996). Das Modell von Bäuml (1996) bezieht sich allerdings nicht - wie die bisher genannten Modelle - auf den Abruf von einzelnen Items, sondern auf den Abruf von assoziierten Wortpaaren (siehe auch Batchelder & Riefer, 1980, 1986).

Das wahrscheinlich bekannteste Modell zur Trennung von Enkodierungs- und Abrufprozessen ist das multinomiale Modell von Batchelder und Riefer (1980, 1986). Dieses Modell kann auf die freie Reproduktion von semantisch assoziierten Wortpaaren angewandt werden und trennt die Prozesse der Enkodierung/Speicherung der Wortpaare als Cluster und des Abrufs solcher Cluster. Auch dieses Modell wurde von Bayen (1990) sowie Riefer und Batchelder (1991) für den Vergleich von jüngeren und älteren Menschen herangezogen. Ebenfalls zur Klasse der multinomialen Modelle ist das Modell von Riefer und Rouder (1992) zu zählen. Der Modellauswertung liegen in diesem Fall Wortpaare zugrunde, die nicht aufgrund semantischer Gemeinsamkeiten zu assoziieren sind, sondern zufällig zusammengestellt werden. Zunächst folgt ein freier Reproduktionstest für das Wortpaar, danach ein geförderter Reproduktionstest, bei dem das erste Wort als Hinweisreiz vorgegeben wird. Separiert werden die Prozesse Cluster-Enkodierung, Cluster-Abruf sowie Enkodierung/Abruf von Einzelwörtern. Dabei ähnelt das Modell dem zuletzt geschilderten Modell von Batchelder und Riefer (1980) dahingehend, dass die Trennung von Enkodierung und Abruf sich bei Riefer und Rouder (1992) ebenfalls auf die *Assoziation* von Stimulus- und Response-Wort bezieht. Der multinomiale Ansatz von Riefer und Rouder (1992) wurde m.W. bisher noch nicht im gerontologischen Kontext genutzt. Rouder und Batchelder (1998) stellen eine ganze Klasse von Modellen vor, die für das Paradigma des freien vs. geförderten Reproduktionsvergleichs die Schätzung von Enkodierungs/Speicherungs- und Abrufparametern erlauben, aber bisher auch noch nicht auf entwicklungspsychologische Daten angewandt wurden.

Ein weiteres Modell von Chechile und Meyer (1976) wurde vor allem für Daten, die aus dem ‚Brown-Peterson-Paradigma‘ stammen, entwickelt. Dieses Modell kann aber auch zur Analyse der episodischen Gedächtnisleistung herangezogen werden (als Überblick siehe Chechile, 2004). Notwendig für die Parameterschätzungen ist hier der Vergleich der Gedächtnisleistung in der freien Reproduktion und Rekognition. Diese beiden Testformate liegen auch dem Modell von Riefer und Batchelder (1995) zugrunde, welches vor allem zur genaueren Analyse des ‚Recognition-Failure-Paradigmas‘ genutzt wurde, bisher nicht im Altersvergleich. Eines der wenigen Modelle, das zur Analyse von reinen Rekognitionsdaten verwendet werden kann, ist das ‚Stimulus-Sampling-Modell‘ von Glanzer und Bowles (1976), welches von Bowles und Poon (1982) für den Altersvergleich von 2AFC-Rekognitionsexperimenten genutzt wurde.

Die genannten Modelle befassen sich alle mit dem Problem der Trennung von Enkodierungs- Speicherungs- und Abrufprozessen. Häufig ist es dabei - wie bereits erwähnt -

der Fall, dass Enkodierungs- und Speicherprozesse in einem Parameter erfasst werden und gegenüber Abrufprozessen kontrastiert werden (z.B. Batchelder & Riefer, 1980; Chechile & Meyer, 1976; Riefer & Rouder, 1992). Wenige Modell ermöglichen die Schätzung von Parametern für alle drei kognitiven Teilprozesse (z.B. Howe, 1988; Wilkinson & Koestler, 1983). Im Hinblick auf die Alternsforschung ist erneut festzuhalten, dass die Zusammenfassung von Enkodierungs- und Speicherprozessen als unproblematisch gewertet werden kann, da sich zwar Evidenz für ein Enkodierungs- und Abrufdefizit im Alter findet, Speicherprozesse vom Alterungsprozess aber weitestgehend ausgenommen zu sein scheinen. Modelle, die die Schätzung von Speicherprozessen ermöglichen, konnten einvernehmlich keine Altersunterschiede in diesem Prozess feststellen (Bäuml, 1996; Howe, 1988; Wilkinson & Koestler, 1983, zusammenfassend siehe Bayen, 1990). Diese Befunde sind auch in Einklang mit designbasierten Ansätzen (zusammenfassend siehe Bayen, 1990; Salthouse, 1982) und unterstützen daher die Schlussfolgerung von Salthouse, der im Hinblick auf eine Reihe von Studien zu dem Urteil kommt: „[...] we can be fairly confident, that age differences in the storage state of memory are slight to nonexistent.“ (Salthouse, 1982, S.141)

Neben den modellbasierten Ansätzen zur Trennung von Enkodierungs- und Abrufprozessen gibt es auch mathematische Modelle, die sich mit dem Kontextdefizit bzw. assoziativem Defizit befassen und eine Dekomponierung von (a) Itemerinnerung, (b) Kontexterinnerung, (c) Erinnerung der Assoziation von Item und Kontext sowie (d) Rateprozessen ermöglichen. Auch diese Modellierungsansätze sind für die vorliegende Arbeit relevant. Bereits erläutert wurden das ICE-Modell (Murnane et al., 1999) und das multinomiale Quellengedächtnismodell (Bayen et al., 1996). Auch das bereits erwähnte Markoff-Ketten-Modell von Kliegl und Lindenberger (1993, siehe Kapitel 3.1) kann in diese Reihe eingliedert werden. Nicht genauer erwähnt wurde bisher das ‚Temporal-Context-Modell‘ (Howard & Kahana, 2002), welches sich mit der Bindung des zeitlichen Kontexts an Iteminformationen befasst und für die freie Reproduktion eingesetzt werden kann. Studien, die ihre Datenauswertung auf diesen Modellen basierend vornahmen, konnten Evidenz für ein assoziatives Defizit im Alter liefern, da sich in der spezifisch assoziativen Komponente ein Altersunterschied zeigen ließ (Bayen & Murnane, 1996; Bayen, Phelps & Spaniol, 2000; Howard, Kahana & Wingfield, 2006; Kliegl & Lindenberger, 1993).

Unabhängig von den bisher diskutierten assoziativen Prozessen bzw. der Trennung von Enkodierung und Abruf gibt es natürlich zahlreiche Modelle, die die Schätzung von Vertrautheit und bewusster Erinnerung sowohl für Itemrekognitionsaufgaben als auch für assoziative Rekognitionsaufgaben erlauben (für einen Überblick siehe Macho, 2004; M. R. Healy et al., 2005). Diese wurden auch in der Alternsforschung angewandt und brachten wichtige Erkenntnisse zu Tage (z.B. dass ältere Probanden vor allem eine reduzierte bewusste Erinnerung aufweisen, sich in Vertrautheitsprozessen aber weitestgehend keine Altersunterschiede aufzeigen, siehe Light et al., 2000; Zacks et al., 2000). Allerdings sind diese Modelle - trotz der Anwendung auf das assoziative Rekognitionsparadigma - für

die vorliegende Fragestellung wenig relevant, da sie sich mit anderen spezifischen Prozessen des Gedächtnisses befassen.

In den letzten Abschnitten wurden modellbasierte Lösungsansätze genannt, die sich entweder mit der Dekomponierung von Enkodierung und Abruf befassen oder eine genauere Prozessanalyse der assoziativen Verknüpfung von zentraler Information und Kontext erlauben. Da die Modelle für unterschiedliche Paradigmen geeignet sind und unterschiedliche Arten assoziativer Prozesse modellieren, soll nun eingeschränkt werden, welche Modelle für das vorliegende Forschungsvorhaben zentral sind. Nur diese im Kontext der vorliegenden Arbeit relevanten Modelle sollen dann eingehender erläutert werden.

Um eine Entscheidung darüber zu treffen, welche der berichteten Ansätze einen relevanten Beitrag zur Lösung der vorliegenden Forschungsfrage liefern können, sollen zunächst Kriterien formuliert werden, die zur Erreichung dieses Ziels notwendig sind. Folgende Anforderungen sollte ein formaler Modellierungsansatz, der die Lokalisationsfrage des assoziativen Defizits klären soll, erfüllen:

1. Das Modell sollte sowohl Parameter für die Erinnerung an einzelne Items (Elemente) als auch Parameter für die Assoziation der Elemente beinhalten.
2. Die Erinnerungsleistung an die Assoziation der beiden Elemente sollte nicht nur global erfasst werden, sondern eine spezifische Rückführung auf die Prozesse Enkodierung und Abruf erlauben und deren Dekomponierung ermöglichen.
3. Im Hinblick auf das Paradigma sollte das Modell zur Analyse von Rekognitionsdaten eingesetzt werden können.
4. Das Modell sollte auf verschiedene Materialarten anwendbar sein. Assoziiertes sowie nicht assoziiertes Wortmaterial, visuelle Stimuli sowie die Verbindung von visuellen und verbalen Reizen sollten mit dem Messmodell untersucht werden können.

Filtert man unter diesen vier relevanten Gesichtspunkten die existierenden Modelle, so finden sich sechs Modelle, die sich im weitesten Sinne mit der Assoziation von zwei Elementen befassen und damit der ersten Voraussetzung gerecht werden:

- (1) Das Markoff-Ketten-Modell von Kliegl und Lindenberger (1993) modelliert die Assoziation von Kontext und zentraler Information.
- (2) Das Enkodierung-Abruf-Modell von Batchelder und Riefer (1980) erfasst die Assoziationen semantisch relatierter Wörter.
- (3) Das multinomiale Modell von Riefer und Rouder (1992) kann auf beliebige Assoziationen zwischen zwei Wörtern angewandt werden.

- (4) Das ‚Temporal-Context-Modell‘ von Howard und Kahana (2002) umfasst die Assoziation von Item und zeitlichem Kontext.
- (5) Das ICE-Modell von Bayen, Phelps und Spaniol (2000) sieht die Assoziation von Item und räumlichem Kontext als Ensemble an.
- (6) Das Modell zur Analyse von Quellengedächtnisaufgaben von Bayen et al. (1996) erfasst die Assoziation von Item und Quelle.

Von diesen sechs Modellen enthalten die ersten drei isolierte und unabhängige Parameter für die Erinnerungsleistung an einzelne Items, die getrennt von der Verbindung der beiden Elemente geschätzt werden (Voraussetzung 1). Eine Trennung von Enkodierung und Abruf wird lediglich von den beiden Modellen zwei und drei (Batchelder & Riefer, 1980; Riefer & Rouder, 1992) vorgesehen, nur diese Modelle erfüllen damit die zweite Voraussetzung. Auf alleinige Rekognitionsdaten (ohne den Vergleich zur Reproduktion) sind nur die letzten beiden Modelle, das ICE-Modell und das Quellengedächtnismodell anwendbar (Voraussetzung 3). Eine Analyse von zufällig zusammengestellten Elementkombinationen, die keine besondere Assoziation aufweisen müssen (Voraussetzung 4), ermöglichen das multinomiale Modell von Riefer und Rouder (1992) sowie die Modelle vier bis sechs.

Aus diesem Abgleich zwischen notwendigen Voraussetzungen zur Lösung des vorliegenden Problems und den Eigenschaften der existierenden Modelle kann erschlossen werden, dass keines der Modelle alle Voraussetzungen erfüllt und direkt verwendet werden kann, um die komplexe Fragestellung dieser Arbeit zu beantworten. Deutlich wird aber auch, dass wichtige Teilaspekte bereits modelliert wurden und ein neues Messmodell sich daher an diese bereits existierenden Modelle und ihre zugrundeliegenden theoretischen Annahmen anlehnen sollte. Daher werden im Folgenden zwei Modelle genauer dargestellt. Zum einen das ‚Enkodierung-Abruf-Modell‘¹⁸ (‚pair-clustering model‘), zum anderen die Modelle zur Analyse von Quellengedächtnisexperimenten. Das Enkodierung-Abruf-Modell ist vor allem in Bezug auf die Trennung der Prozesse Enkodierung und Abruf interessant, sowie aufgrund der Tatsache, dass Parameter für die Itemerinnerung unabhängig von den assoziativen Parametern geschätzt werden. Das zweite Modell ist auf ein Paradigma anwendbar, das eine ähnliche Lern- und Testphase aufweist wie die Rekognitionsexperimente zur Assoziationsdefizit-Hypothese. Darüber hinaus wird die Assoziation zweier zufällig kombinierter Elemente erfasst.

4.2.1 Das Enkodierung-Abruf-Modell

Das Enkodierung-Abruf-Modell (‚pair-clustering model‘, Batchelder & Riefer, 1980, 1986) wurde entwickelt, um den Effekt des Abstands semantisch zusammengehöriger Wörter in

¹⁸Die Bezeichnung Enkodierung-Abruf-Modell wird in dieser Arbeit in Anlehnung an Erdfelder (2000) verwendet.

einer Lernliste (‘semantic spacing-effect’) genauer auf spezifische Gedächtnisprozesse zurückzuführen. Diesem Effekt liegt die Beobachtung zugrunde, dass semantisch relatierte Wörter in einer Lernliste tendenziell auch miteinander assoziiert werden und daher in der Testphase oftmals zusammen d.h. direkt nacheinander reproduziert werden. In einigen Studien wurde beobachtet, dass die Wahrscheinlichkeit, zwei Wörter derselben Kategorie auch nacheinander wiederzugeben, mit dem Abstand der beiden Wörter in der Lernliste abnimmt. Mit steigender Anzahl interferierender Items sinkt die Wahrscheinlichkeit einer erfolgreichen Verknüpfung (zusammenfassend siehe Batchelder & Riefer, 1980). Andere Studien dahingegen zeigten ein komplexeres Bild auf, es resultierte ein U-förmiger Zusammenhang zwischen dem Abstand der Wörter in der Lernliste und der Wahrscheinlichkeit, die Wörter gemeinsam zu reproduzieren (zusammenfassend siehe Batchelder & Riefer, 1980).

Batchelder und Riefer (1980) vermuteten, dass der U-förmiger Zusammenhang durch die unterschiedlichen Beiträge von Enkodierungs- und Abrufprozessen zustande kommen könnte: Die Wahrscheinlichkeit der erfolgreichen Enkodierung zweier Wörter als Cluster ist besonders hoch bei geringem Abstand der beiden Einzelitems in der Wortliste und sinkt mit zunehmendem Abstand. Genau entgegengesetzt verhält es sich mit dem Abruf von einem als Cluster enkodierten Wortpaar. Der Erfolg des Abrufs ist umso wahrscheinlicher, je größer der Abstand zwischen den beiden Wörtern in der Wortliste ist.

Die Überprüfung dieser theoretischen Annahmen stellt sich als schwierig heraus, da es sich um ein Dekomponierungsproblem handelt. Werden zwei semantisch assoziierte Wörter direkt nacheinander genannt, so ist klar, dass diese sowohl erfolgreich enkodiert als auch abgerufen werden konnten.¹⁹ Gelingt dies jedoch nicht, ist offen, inwieweit hierfür Enkodierungs- oder Abrufprobleme verantwortlich sind. Auch der U-förmige Zusammenhang zwischen Abstand in der Lernliste und Assoziationserfolg kann anhand der beobachtbaren Daten nicht auf die Prozesse Enkodierung und Abruf zurückgeführt werden. Um dieses Problem zu lösen, entwickelten Batchelder und Riefer daher das Enkodierung-Abruf-Modell.

Dieses Modell ist auf das Paradigma der freien Reproduktion zugeschnitten und erfordert eine Lernliste mit semantisch assoziierten Wörtern sowie Einzelitems ohne assoziiertes Partnerwort. Empirisch beobachtet werden kann, welche Itemklasse (Einzelitem oder Itempaar) vorliegt und inwieweit die Wörter in der Testphase erinnert werden. Für die reproduzierten Wörter werden dabei mehrere Ereigniskategorien unterschieden:

- (1) Beide Items des Paares werden unmittelbar nacheinander reproduziert [$C_{1,1}$],
- (2) beide Items des Paares werden reproduziert, aber nicht nacheinander [$C_{1,2}$],
- (3) nur ein Item des Paares wird reproduziert [$C_{1,3}$],

¹⁹Die Wahrscheinlichkeit, dass ein solches Ereignis rein zufällig eintritt, muss natürlich berücksichtigt werden. Diese kann jedoch als verschwindend gering angesehen werden.

- (4) kein Item des Paares wird reproduziert [$C_{1,4}$],
- (5) ein Einzelitem wird reproduziert [$C_{2,1}$],
- (6) ein Einzelitem wird nicht reproduziert [$C_{2,2}$].

Die sechs Ereigniskategorien wurden in eckigen Klammern mit Abkürzungen versehen. Der erste Index differenziert dabei zwischen Wörtern mit einem semantisch assoziierten Partnerwort und Einzelitems. Der zweite Index nummeriert die Ereignisse willkürlich. Die Datenstruktur in diesem Paradigma erfüllt damit eine der Voraussetzungen, die für die multinomiale Modellierung notwendig ist: das Vorliegen einer endlichen Anzahl diskreter, beobachtbarer, exhaustiver und disjunkter Kategorien, C_1, C_2, \dots, C_j , wobei die Menge der möglichen Beobachtungsergebnisse mit $S = \{C_1, C_2, \dots, C_j\}$ notiert wird.

Batchelder und Riefer (1980) postulieren vier kognitive Prozesse, deren Zusammenwirken erklärt, wie die Ergebniskategorien erreicht werden. Diese Prozesse können dabei in Form eines Verarbeitungsbaummodells visualisiert werden, man spricht dann von einem multinomialen Verarbeitungsbaum-Modell (MVB-Modell). Ein solches Modell mündet in die beobachteten Ereignisse C_{ij} . Seinen Ausgangspunkt nimmt das Modell in der jeweiligen Itemklasse, d.h. es gibt ein Verarbeitungsbaummodell für Itempaare und ein zweites für Einzelitems. Damit liegen in diesem Fall $k = 2$ verschiedene Itemklassen vor, wobei im Allgemeinen $k \in N$ ist. Sobald $k > 1$ ist, spricht man von einem verbundenen multinomialen Modell, wobei k nicht nur für unterschiedliche Itemklassen, sondern auch für unterschiedliche experimentelle Bedingungen oder Personengruppen stehen kann. In Abbildung 4.1 ist das verbundene multinomiale Verarbeitungsbaummodell, das Enkodierung-Abruf-Modell abgebildet. Die empirisch beobachtbaren Datenpunkte sind durch Rechtecke gekennzeichnet, zwischen diesen beobachtbaren Ereignissen sind die latenten kognitiven Prozesse dargestellt (gekennzeichnet durch Rechtecke mit abgerundeten Ecken).

Das Modell enthält folgende theoretische Annahmen: Ein Itempaar kann mit der Wahrscheinlichkeit c als Cluster enkodiert/gespeichert werden. Darauf folgend kann dieses Cluster mit der Wahrscheinlichkeit r auch erfolgreich abgerufen werden. In diesem Fall würde die Person beide Wörter direkt nacheinander nennen, d.h. die Kategorie $C_{1,1}$ wäre zu beobachten. Gelingt es allerdings nicht das Cluster abzurufen ($1 - r$), so kann auch keines der beiden Wörter als einzelnes Wort reproduziert werden, und die Versuchsperson würde daher kein Wort dieses Wortpaares nennen ($C_{1,4}$). Hier wird die theoretische Annahme getroffen, dass der Abruf eines enkodierten und gespeicherten Clusters ein ‚Alles-oder-nichts‘-Prozess ist. Entweder gelingt es, beide Wörter als Cluster zu erinnern, oder aber kein Wort kann erfolgreich reproduziert werden.

Falls die Wörter nicht als Wortpaar enkodiert wurden ($1 - c$), so kann es der Fall sein, dass beide Wörter getrennt voneinander als Einzelwörter enkodiert/gespeichert wurden und auch abgerufen werden können. Dies würde dazu führen, dass im Test beide Items genannt werden, allerdings nicht direkt nacheinander ($C_{1,2}$). Es ist auch denkbar, dass

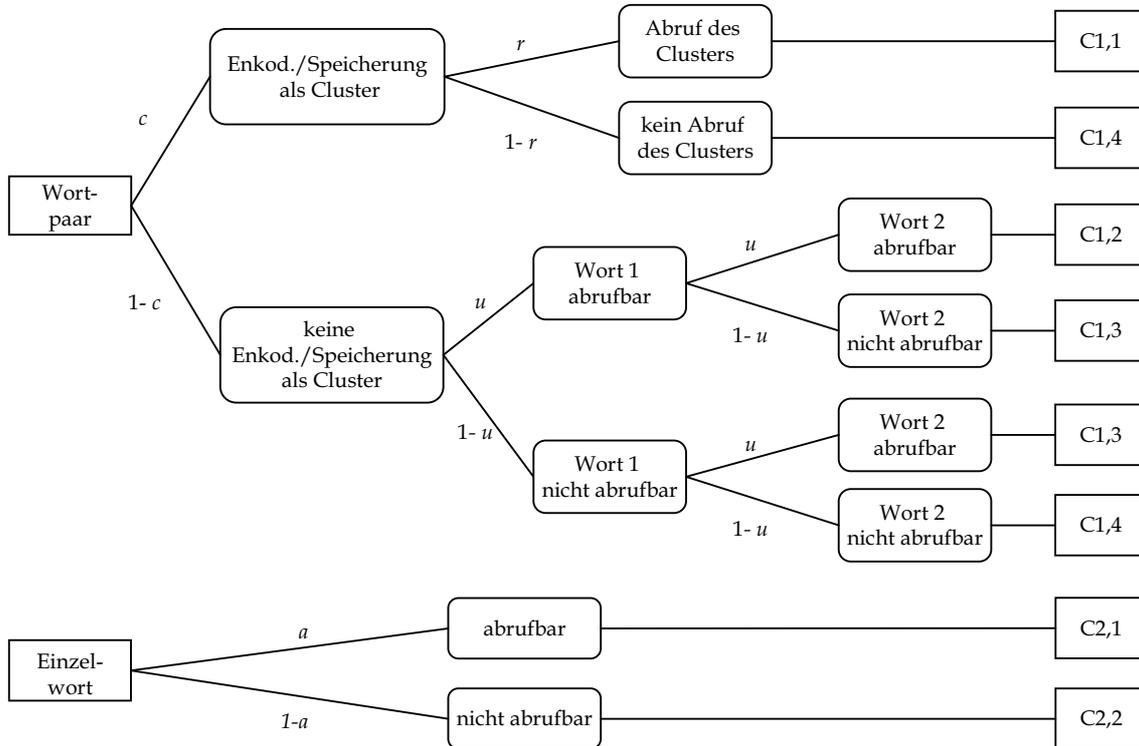


Abbildung 4.1: In dieser Abbildung ist das multinomiale Enkodierung-Abruf-Modell von Batchelder und Riefer (1980) dargestellt. Die genaue Beschreibung der einzelnen Parameter kann dem Text entnommen werden.

nur das erste Wort oder nur das zweite Wort erfolgreich reproduziert wird (u). Diese beiden Fälle würden jeweils zu dem Ereignis $C_{1,3}$ führen. Falls keines der beiden Wörter im Test genannt wird ($C_{1,4}$), so wird angenommen, dass kein Item des Wortpaares enkodiert/gespeichert oder abgerufen wurde ($1 - u$).

Für die Einzelitems wird postuliert, dass diese mit der Wahrscheinlichkeit a abgerufen werden können und damit zu dem Ereignis $C_{2,1}$ führen, oder aber dass dieser Prozess nicht gelingt ($1 - a$) und das Ereignis $C_{2,2}$ (kein Wort wird im Test genannt) tritt ein. Die Parameter a und u für die Reproduktion der Einzelitems sind Hybridparameter. Sie trennen nicht zwischen Enkodierung/Speicherung und Abruf. Da die Einzelitems und Itempaare aus demselben Itempool gezogen wurden und sich in Bezug auf ihre Salienz, Bildhaftigkeit und Bedeutungshaltigkeit nicht systematisch unterscheiden, sollten sich Einzelitems verhalten wie nicht-geclusterte Items aus Itempaaren, so dass die Annahme $a = u$ getroffen werden kann.

Wie bei der Modellbeschreibung klar geworden sein sollte, sind mit einer Modellkonstruktion immer auch theoretische Annahmen verbunden. Zum einen betreffen diese die Auswahl und Beschreibung der postulierten kognitiven Prozesse, zum anderen die Anordnung und Kombination dieser Prozesse. Gleiches gilt für die empirischen Beobachtungsdi-

mensionen. Diese könnten prinzipiell und abhängig vom Forschungsinteresse anders definiert werden.

Auch im Hinblick auf das Ziel der multinomialen Modellierung werden bestimmte Annahmen vorausgesetzt, z.B. dass die Wahrscheinlichkeit eines Beobachtungsergebnisses C_{ij} nur von der Kategoriezugehörigkeit abhängt, nicht aber von der Anzahl der Beobachtungsergebnisse oder von den Ergebnissen anderer Beobachtungen. Definiert man zu dem Stichprobenraum S das Wahrscheinlichkeitsmaß $P: \text{Pot}(s) \rightarrow [0, 1]$ und notiert die unbekannte Wahrscheinlichkeit eines Beobachtungsergebnisses C_j mit $\pi_j := P(\{C_j\})$ (wobei $\sum_{j=1}^J \pi_j = 1$), dann gilt unter der Voraussetzung der Unabhängigkeit und Gleichverteilung der Beobachtungskategorien mit der Wahrscheinlichkeit π_j , für die Verteilung der Zufallsvariablen Y_j ($\sum_j Y_j = N$) die allgemeine multinomiale Verteilung der Art:

$$P^N(y_1, y_2, \dots, y_J; \pi_1, \pi_2, \dots, \pi_J) = \frac{N!}{y_1! \cdot y_2! \cdot \dots \cdot y_J!} \pi_1^{y_1} \cdot \pi_2^{y_2} \cdot \dots \cdot \pi_J^{y_J} = N! \prod_{j=1}^J \frac{\pi_j^{y_j}}{y_j!} \quad (4.1)$$

Die Zufallsvariable Y_j mit $j = 1, \dots, J$ steht dabei für die Häufigkeiten der Beobachtungskategorien C_j bei N Beobachtungen. Gleichung 4.1 ist die Multinomialverteilung für J multinomialverteilte Zufallsvariablen Y_j bei N Beobachtungen und den Kategoriewahrscheinlichkeiten $\boldsymbol{\pi} := (\pi_1, \pi_2, \dots, \pi_J)$. Diese Gleichung gibt also die Wahrscheinlichkeiten an, bei N Beobachtungen genau y_j -mal die Kategorie C_j zu beobachten.

Da das Einkodierung-Abruf-Modell ein verbundenes multinomiales Modell mit $k = 2$ ist, wird in diesem Fall die Kategoriewahrscheinlichkeit mit $\pi_{k,j}$, die beobachtete Kategoriehäufigkeit mit $y_{k,j}$ angegeben. Der Vektor der Kategoriehäufigkeiten über alle Versuchsbedingungen ist dann:

$$\boldsymbol{y} := (y_{1,1}, y_{1,2}, \dots, y_{1,J_1}, y_{2,1}, y_{2,2}, \dots, y_{2,J_2}, \dots, \dots, y_{K,1}, y_{K,2}, \dots, y_{K,J_K}) \quad (4.2)$$

Der Vektor der Kategoriewahrscheinlichkeiten über alle k Bedingungen hinweg ist:

$$\boldsymbol{\pi} := (\pi_{1,1}, \pi_{1,2}, \dots, \pi_{1,J_1}, \pi_{2,1}, \pi_{2,2}, \dots, \pi_{2,J_2}, \dots, \dots, \pi_{K,1}, \pi_{K,2}, \dots, \pi_{K,J_K}) \quad (4.3)$$

Bei gegebener Annahme der Unabhängigkeit der Beobachtungshäufigkeiten in unterschiedlichen Versuchsbedingungen ist die Wahrscheinlichkeitsverteilung des Ereignisses \boldsymbol{y} bei zugrundeliegender Kategoriewahrscheinlichkeit $\boldsymbol{\pi}$:

$$\Pr(\boldsymbol{y}, \boldsymbol{\pi}) = \prod_{k=1}^K P^{N_k}(y_{k,1}, y_{k,2}, \dots, y_{k,J_k}; \pi_{k,1}, \pi_{k,2}, \dots, \pi_{k,J_k}) = \prod_{k=1}^K N_k! \cdot \prod_{j=1}^{J_k} \frac{\pi_{k,j}^{y_{k,j}}}{y_{k,j}!} \quad (4.4)$$

Diese Wahrscheinlichkeitsfunktion für k Bedingungen wird verbundene Multinomialverteilung genannt. Mit diesem Rahmenmodell ist es nun möglich, Wahrscheinlichkeiten der

Beobachtungskategorien zu schätzen oder Annahmen über Gleichheit von Wahrscheinlichkeiten zu testen.

Um aber Kategoriewahrscheinlichkeiten auf die genannten psychologisch interpretierbaren Parameter zurückführen zu können, müssen weitere Annahmen getroffen werden. Über Modellgleichungen gelangt man von den beschriebenen multinomialen Modellen zu parametrisierten multinomialen Modellen. Modellgleichungen spezifizieren die Annahmen des zugrundeliegenden theoretischen Modells über das Entstehen der Kategoriewahrscheinlichkeiten. Mit Hilfe von Parametern werden die Kategoriewahrscheinlichkeiten erklärt. Diese Parameter, die eine psychologische Interpretation erfahren, werden latente Parameter oder Parameter zweiter Ordnung genannt und sind von den Kategoriewahrscheinlichkeiten π_k , den Parametern erster Ordnung, abzugrenzen. Das Modell spezifiziert also einen oder mehrere funktional unabhängige Parameter $\theta_1, \theta_2, \dots, \theta_S, 1 \leq S < J$, wobei die Parameter innerhalb eines Intervalls reeller Zahlen I_s liegen. In vielen Paradigmen, so auch im Fall des Enkodierung-Abruf-Modells, sind die Parameter als Wahrscheinlichkeiten zu interpretieren und auf das Intervall $[0, 1]$ beschränkt. Der Parameterraum ist damit gegeben durch $\Omega = \{\theta = (\theta_1, \dots, \theta_s, \dots, \theta_S) | \theta_s \in I_s, s = 1, \dots, S\}$. Im konkreten Fall des Enkodierung-Abruf-Modells liegen die drei Parameter, $\theta_1 = c, \theta_2 = r, \theta_3 = u$ vor, der Parameterraum ist damit $\Omega = [0, 1]^3$.

Über die parametrisierten Modellgleichungen werden die Wahrscheinlichkeiten von Kategorien in Abhängigkeit der Parameter vorhergesagt. Diese werden mit $p_{j,k}$ Ω notiert. Die Modellgleichungen definieren jeweils die Abbildung des Parameterraums Ω auf die Menge möglicher Kategoriewahrscheinlichkeiten Γ , so dass $f : \Omega \rightarrow \Gamma$ gilt, wobei die Menge der möglichen Kategoriewahrscheinlichkeiten

$$\Gamma = \left\{ p = (p_{1,1}, p_{1,2}, \dots, p_{1,J_1}, p_{2,1}, p_{2,2}, \dots, p_{2,J_2}, \dots, p_{K,1}, p_{K,2}, \dots, p_{K,J_K}) \mid \right. \\ \left. \forall k, j : p_{k,j} \in [0, 1]; \forall k : \sum_{j=1}^{J_K} p_{k,j} = 1 \right\} \quad (4.5)$$

ist.

Modellgleichungen können im Beispiel des Enkodierung-Abruf-Modells aus dem Verarbeitungsbaummodell abgeleitet werden. Die Wahrscheinlichkeit einer Beobachtungskategorie ergibt sich als Summe der Wahrscheinlichkeiten aller „Äste“, die in dieses Ereignis einmünden. Die Wahrscheinlichkeit eines Astes ist dabei das Produkt der Übergangswahrscheinlichkeiten entlang dieses Pfades. Man erhält für das Enkodierung-Abruf-Modell daher folgende Modellgleichungen:

$$\begin{aligned} p(C_{1,1}) &= c \cdot r \\ p(C_{1,2}) &= (1 - c) \cdot u \cdot u = (1 - c) \cdot u^2 \\ p(C_{1,3}) &= (1 - c) \cdot u \cdot (1 - u) + (1 - c) \cdot (1 - u) \cdot u = 2 \cdot (1 - c) \cdot u \cdot (1 - u) \end{aligned}$$

$$\begin{aligned}
p(C_{1,4}) &= c \cdot (1-r) \cdot (1-c) \cdot (1-u) \cdot (1-u) = c \cdot (1-r) + (1-c) \cdot (1-u)^2 \\
p(C_{2,1}) &= u \\
p(C_{2,2}) &= 1-u
\end{aligned}$$

Mit diesen Annahmen kann das zugrundeliegende Modell als parametrisiertes Modell bezeichnet werden. Die Parameter sind als Wahrscheinlichkeiten zu interpretieren.

Bevor ein multinomiales Modell zur Datenanalyse verwendet werden kann, müssen allerdings mehrere Voraussetzungen erfüllt sein. Zum einen muss das Modell identifizierbar sein, zum anderen muss das Modell eine gute Anpassung an die Daten aufweisen, d.h. die Abweichung zwischen den empirischen Häufigkeiten und den aus dem Modell vorhergesagten Häufigkeiten sollten auf dem Zufallsniveau liegen. Im Folgenden wird zunächst das Identifizierbarkeitskriterium erklärt, dann wird auf die Modellanpassung und das genauere Verfahren der Parameterschätzung eingegangen.

Multinomiale Verarbeitungsbaummodelle ordnen über die Modellgleichungen jedem Parametervektor $\Omega = (\theta_1, \theta_2, \dots, \theta_S)$ bestimmte Kategoriewahrscheinlichkeiten $p(C_{kj})$ zu. Damit wird eine Abbildung f des Parameterraums in die Menge der möglichen Vektoren der Kategoriewahrscheinlichkeiten definiert. Wenn die Abbildung f injektiv ist, so spricht man von einem *global identifizierbaren* Modell. Die globale Identifizierbarkeit muss analytisch gezeigt werden. Eine Mindestvoraussetzung für die Identifizierbarkeit ist, dass die Anzahl der Parameter nicht die Anzahl der freien Kategoriewahrscheinlichkeiten ($M = \sum_{k=1}^K (J_k - 1)$) überschreitet. Im Einkodierung-Abruf-Modell liegen $M=4+2$ Beobachtungskategorien vor. Daraus folgt, dass 3+1 Kategorien voneinander unabhängig sind. Da drei Parameter geschätzt werden, weist dieses Modell einen Freiheitsgrad auf, und die Identifizierbarkeit ist damit prinzipiell möglich. Auf den Nachweis des schwächeren Kriteriums der lokalen Identifizierbarkeit wird im Folgenden noch eingegangen.

Für die Bestimmung der Parameterwerte haben Hu und Batchelder (1994) ein Verfahren entwickelt, das auf dem ‚Expectation-Maximization-Algorithmus‘ (EM-Algorithmus) basiert. Dabei wird die Distanzfunktion PD_λ (Potenz-Divergenz-Statistik) zwischen den beobachteten Kategoriehäufigkeiten und den bei Modellgültigkeit vorhergesagten Kategoriehäufigkeiten minimiert. Für $\lambda = 0$ ergibt sich folgende PD_λ -Funktion:

$$PD_{\lambda=0} = G^2 = 2 \sum_{k=1}^K \sum_{j=1}^{J_k} y_{kj} \cdot \ln \left(\frac{y_{kj}}{N_k \cdot p_{kj}(\theta_1, \dots, \theta_S)} \right) \quad (4.6)$$

Wie Gleichung 4.6 zeigt, reduziert sich für den Spezialfall $\lambda = 0$ die Funktion auf die Likelihood-Quotientenstatistik G^2 . Diese Likelihood-Quotientenstatistik kann als Modellanpassungsstatistik genutzt werden. Dabei gilt, dass die Statistik $\chi^2_{(df)}$ verteilt ist, wobei für die Freiheitsgrade $df = M - S = \sum_{k=1}^K (J_k - 1) - S$ gilt. Damit kann die Nullhypothese der Modellgültigkeit überprüft werden.

Im Kontext der Parameterschätzungen kann die lokale Identifizierbarkeit relativ leicht gezeigt werden. Für die lokale Identifizierbarkeit muss die Abbildung f nur für die nähere Umgebung des Parametervektors, der für eine bestimmte Stichprobe zugrunde liegt, injektiv sein. Diesen Nachweis kann man erbringen, indem man den EM-Algorithmus mehrfach mit unterschiedlichen Zufallsstartwerten aktiviert. Resultiert für den minimalen Wert der G^2 -Statistik immer der gleiche Schätzvektor $\Omega = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_S)$, so ist die lokale Identifizierbarkeit wahrscheinlich. Das Computerprogramm *multiTree* bietet eine einfache Möglichkeit, diesen Simulationsnachweis der Identifizierbarkeit (siehe Erdfelder, 2000; Rouder & Batchelder, 1998) durchzuführen.

Wenn das Modell mindestens lokal identifizierbar ist und eine gute Modellanpassung gegeben ist, muss in einem letzten Schritt nachgewiesen werden, dass die Modellparameter auch die Prozesse repräsentieren, die ihrer Beschreibung entsprechen. Es muss also sichergestellt werden, dass z.B. der Enkodierungsparameter auch wirklich die Enkodierung des Wortpaares widerspiegelt. Dieser Nachweis erfolgt über Validierungsexperimente. Diese Experimente manipulieren gezielt einen kognitiven Prozess, indem sie auf empirisch gesicherte Befunde zurückgreifen und diese replizieren. Solche klassischen Manipulationen müssen sich dann selektiv auf den Zielparameter auswirken und möglichst keine Effekte auf andere Parameter zeigen. Unter diesen Voraussetzungen können die Parameter inhaltlich interpretiert werden, Hypothesen können überprüft werden und das Modell kann zur Lösung des Dekomponierungsproblems verwendet werden. Sowohl die Validierung als auch eine gute Modellanpassung sowie lokale Identifizierbarkeit wurden für das Enkodierungs-Abruf-Modell nachgewiesen (siehe dazu Batchelder & Riefer, 1980, 1986, 1999; Bayen, 1990; Erdfelder, 2000; Riefer & Batchelder, 1991)

Für die Anwendung auf das Ausgangsproblem gelang es in der modellbasierten Auswertung, die ursprünglichen Hypothesen von Batchelder und Riefer (1980) zu bestätigen: Je geringer der Abstand zwischen zwei assoziierten Wörtern in der Wortliste war, desto höher fiel der Enkodierungs-/Speicherungsparameter c aus. Mit zunehmendem Abstand hingegen erhöhte sich die Wahrscheinlichkeit des erfolgreichen Abrufs des Clusters (r). Damit konnte der U-förmige Zusammenhang auf das Zusammenwirken von Enkodierungs- und Abrufprozessen zurückgeführt, werden und das Dekomponierungsproblem wurde gelöst.

Auch in der Altersforschung wurde das Modell erfolgreich eingesetzt. Sowohl die Arbeit von Bayen (1990) als auch das Experiment von Riefer und Batchelder (1991) zeigten einen signifikanten Altersunterschied im Abrufparameter r auf. Älteren Menschen gelang es signifikant seltener ein als Cluster enkodiertes und gespeichertes Wortpaar auch abzurufen. Dieser Befund steht im Einklang zu mehreren Alternstheorien (für einen Überblick siehe Bayen, 1990). Das Ergebnismuster im Hinblick auf den Enkodierungsparameter fiel zwischen den Studien unterschiedlich aus: Riefer und Batchelder (1991) konnten keinen signifikanten Unterschied zwischen den Altersgruppen feststellen, bei Bayen (1990) hingegen

wurde auch dieser Altersvergleich signifikant. Diese Differenzen konnten durch das Zusammenspiel eines relativ kleinen Alterseffekts im Enkodierungsparameter und einer größeren Stichprobe und damit auch einer höheren Teststärke in der Studie von Bayen (1990) erklärt werden. Neben dem Abrufdefizit zeigen ältere Menschen in diesem spezifischen Paradigma also auch ein Enkodierungsdefizit. Eine erst kürzlich erschienene Publikation von Bröder, Herwig, Teipel und Fast (2008) nutzte das Enkodierung-Abruf-Modell, um Gedächtnisunterschiede zwischen gesunden älteren Menschen und Probanden mit einer leichten kognitiven Beeinträchtigung („mild cognitive impairment“) zu untersuchen. Die Patientengruppe wies dabei eine besondere Beeinträchtigung beim Abruf von Informationen auf. Die zitierte Arbeit diskutiert qualitativ unterschiedliche Verarbeitungsmechanismen bei normalen und pathologischen Alterungsprozessen (siehe auch Riefer, Knapp, Batchelder, Bamber & Manifold, 2002).

Wie bereits erwähnt wurde, ist das Enkodierung-Abruf-Modell im Rahmen der vorliegenden Forschungsfrage relevant, da es in der Altersforschung und vielen weiteren Studien erfolgreich angewandt wurde und die Trennung von Enkodierungs- und Abrufprozessen ermöglicht. Für das zu entwickelnde Modell sollen daher einige theoretische Annahmen des Enkodierungs-Abruf-Modells übernommen werden. Im Kapitel 4.2.3 wird daher nochmals auf die gerade ausgeführten Modellannahmen verwiesen.

4.2.2 Quellengedächtnis-Modelle

Auch die multinomialen Verarbeitungsbaummodelle zur Analyse von Quellengedächtnisexperimenten sind im Kontext der vorliegenden Forschungsfrage von Bedeutung. Das Quellengedächtnisparadigma weist einige relevante Parallelen zu dem von Naveh-Benjamin (2000) verwendeten Paradigma auf: Zum einen werden in der Lernphase zwei zu assoziierende Stimuli dargeboten (Quelle und zentrale Information bzw. die zwei Wörter eines Wortpaares), zum anderen werden Rekognitionstests zur Erfassung der Gedächtnisleistung verwendet.

Im Gegensatz zu dem geschilderten Enkodierung-Abruf-Modell gibt es mehrere multinomiale Modelle, die zur Analyse von Quellengedächtnisexperimenten eingesetzt werden können (als Überblick siehe Batchelder & Riefer, 1999; Bayen et al., 1996; Bröder & Meiser, 2007; Erdfelder, 2000). Einige dieser Modelle sind zur Analyse eindimensionaler Quellengedächtnisstudien geeignet, andere zur Auswertung von mehrdimensionalen Daten. Diese beiden Formate unterscheiden sich dahingehend, dass in eindimensionalen Quellengedächtnisaufgaben die Attribute *einer* Quellendimension unterschieden werden müssen, z.B. ob Aussagen von einem Mann (Quelle A) oder einer Frau (Quelle B) getroffen wurden. Bei mehrdimensionalen Aufgaben müssen *zwei* Quellendimensionen, z.B. das Geschlecht des Sprechers (männlich vs. weiblich) sowie das Alter der Person (jünger oder älter) unterschieden werden. Es liegt also eine Dimension mit den Ausprägungen A,B vor und eine weitere Dimension mit den Ausprägungen X,Y (siehe auch Kapitel 6.1).

Im Hinblick auf ein Modell zur Lösung des Lokalisationsproblems können aus beiden Modellklassen wichtige Annahmen übernommen werden. Zunächst soll daher auf die eindimensionalen Modelle eingegangen werden, dann wird das Modell zur Analyse mehrdimensionaler Daten diskutiert.

Zur Analyse eindimensionaler Quellengedächtnisaufgaben gilt es, Itemerkennungprozesse, Prozesse der Quellendiskrimination und Rateprozesse für Items sowie Quellen voneinander zu unterscheiden. Spezifischer werden die folgenden kognitiven Prozesse über unterschiedliche Parameter modelliert: die erfolgreiche Erinnerung an ein Item aus der Quelle A (D_a), die Erinnerung an ein Item der Quelle B (D_b) sowie die Entdeckung von Distraktoren als neue Items (D_n). Wird ein Item der Quelle A oder B erfolgreich als altes Item erinnert, so ist es des Weiteren möglich, dass eine korrekte Erinnerung an die Quelle gelingt. Mit der Wahrscheinlichkeit d_a erfolgt dabei die Erinnerung an Quelle A, mit der Wahrscheinlichkeit d_b die Erinnerung an die zweite Quelle B. Falls ein als alt erinnertes Wort keiner Quelle zugeordnet werden kann, so wird mit der Wahrscheinlichkeit a geraten, dass dieses Item aus der Quelle A stammte. Falls es einer Versuchsperson nicht gelingt, ein Wort als alt oder neu zu identifizieren, so wird angenommen, dass mit der Wahrscheinlichkeit b für dieses Item „alt“ geraten wird. Auf diesen Rateprozess folgend wird dann erneut für die Quelle geraten, diesmal mit der Wahrscheinlichkeit g , dass das Item aus Quelle A stammt.

Vergleicht man die Parameteranzahl ($S = 8$) mit den freien Kategoriehäufigkeiten ($M = 6$), so stellt man fest, dass für ein Modell mit all den genannten Parametern $S > M$ gilt und das Modell daher nicht identifizierbar ist. Allerdings können sinnvolle Parameterrestriktionen vorgenommen werden, die zu spezifischeren testbaren Modellen führen (Batchelder & Riefer, 1990; Bayen et al., 1996). Von Bayen et al. (1996) wurde z.B. die Restriktion $D_a = D_b = D_n$ eingeführt. Dies impliziert, dass die Parameter der Itemerkennung für Items aus den beiden Quellen A und B sowie die Distraktorentdeckung sich nicht signifikant unterscheiden. Das Modell wurde über die Gleichsetzung von $d_a = d_b$ weiter dahingehend vereinfacht, dass auch für die Quellendiskrimination die Vergleichbarkeit zwischen den beiden Quellen A und B angenommen wird. Auch die Rateparameter (a, g) wurden mit der Begründung gleichgesetzt, dass die Rateprozesse nicht von der Itemerkennung abhängig sind. Nach diesen Restriktionen resultieren zwei Freiheitsgrade für den Modellanpassungstest, und das Modell ist damit prinzipiell identifizierbar. Da dieses spezifische Modell eine gute Anpassung aufweist, kann davon ausgegangen werden, dass die Restriktionen nicht nur aus theoretischer Hinsicht gerechtfertigt werden können, sondern auch mit den empirischen Daten gut vereinbar sind.

Anhand dieser identifizierbaren und testbaren Vier-Parameter-Version des Modells kann sehr gut aufgezeigt werden, dass trotz einer vergleichbaren Parameteranzahl sowie denselben beobachtbaren Stimulus- sowie Antwortkategorien unterschiedliche zugrundeliegende Gedächtnisannahmen modelliert werden können. Batchelder und Riefer (1990)

formulierten ein Ein-Hochschwellen-Modell, Batchelder, Riefer und Hu (1994) ein Niedrigschwellen-Modell und Bayen et al. (1996) ein Zwei-Hochschwellen-Modell. Alle drei Modelle können in einer Vier-Parameter-Version auf dasselbe Paradigma angewandt werden, obwohl sie mit grundsätzlich unterschiedlichen Gedächtnisannahmen verbunden sind. In Ein-Hochschwellen-Modellen wird davon ausgegangen, dass die obere hohe Schwelle des Entscheidungsraums nur von alten Items überschritten werden kann (bzw. die untere Schwelle nur von neuen Items, falls diese Schwelle definiert wird). In Zwei-Hochschwellen-Modellen wird von zwei hohen Schwellen ausgegangen, d.h. nur alte Items können die obere Schwelle überschreiten und damit als alt identifiziert werden. Gleichzeitig können nur neue Items die untere Schwelle unterschreiten und als neu entdeckt werden. Kann eine Schwelle prinzipiell von alten sowie neuen Items überschritten werden, so heißen die entsprechenden Modelle Niedrigschwellen-Modelle.

Für jedes der genannten Schwellenmodelle kann ein multinomiales Verarbeitungsbaummodell konstruiert werden, das die jeweiligen theoretischen Annahmen repräsentiert (siehe Abbildung 4.2). Bayen et al. (1996) evaluierten diese drei Modellvarianten, indem sie sowohl eine Manipulation der Itemerinnerung einführten (semantische Ähnlichkeit von Distraktoren und Zielwörtern) als auch eine Manipulation der Quellendiskrimination (perzeptuelle Ähnlichkeit der beiden Quellen). In einem validen Messmodell sollte sich die erste Manipulation in den Parametern der Itemerinnerung (D) niederschlagen, die zweite Manipulation sollte sich in den Parametern der Quellendiskrimination (d) zeigen. Beide Manipulationen wurden in mehreren Stufen durchgeführt. Die Evaluationsergebnisse dieser Arbeit zeigten, dass lediglich das Zwei-Hochschwellen-Modell in der Lage war, korrekte Maße für die Item- und Quellengedächtnisleistung zu schätzen. Damit zeigte das Zwei-Hochschwellen-Modell seine Überlegenheit gegenüber den anderen Modellvarianten.

Im Hinblick auf ein neues Messmodell zur Beantwortung der Lokalisationsfrage des assoziativen Defizits kann daher eine Zwei-Hochschwellen-Annahme der Itemerinnerung sowohl aufgrund theoretischer als auch empirischer Befunde als sinnvoll angesehen werden.

Die Zwei-Hochschwellen-Annahme wurde auch von den mehrdimensionalen Quellengedächtnismodellen übernommen (siehe Abbildung 4.3). Diese Modelle unterscheiden sich weder im Hinblick auf den Itemerkennungparameter noch in Bezug auf den Distraktorentdeckungsparameter von den eindimensionalen Modellen. Im Falle einer korrekten Erinnerung an ein altes Item muss allerdings eine Quellenerinnerung in Bezug auf zwei Quellen erfolgen. Das heißt die Probanden können mit der Wahrscheinlichkeit d^1 die erste Quellendimension (A,B) korrekt erinnern. Mit der Wahrscheinlichkeit d^2 gelingt dies auch für die zweite Quellendimension (X,Y). Natürlich ist es möglich, dass nur die Erinnerung an eine Quellendimension gelingt, bzw. dass trotz korrekter „Alt“-Rekognition des Items keine der beiden Quellenzugehörigkeiten erinnert werden kann. In letzterem Fall, sowie bei neuen Items, die nicht als Distraktoren erkannt wurden, müssen die Probanden bezüglich der Quellendimensionen raten.

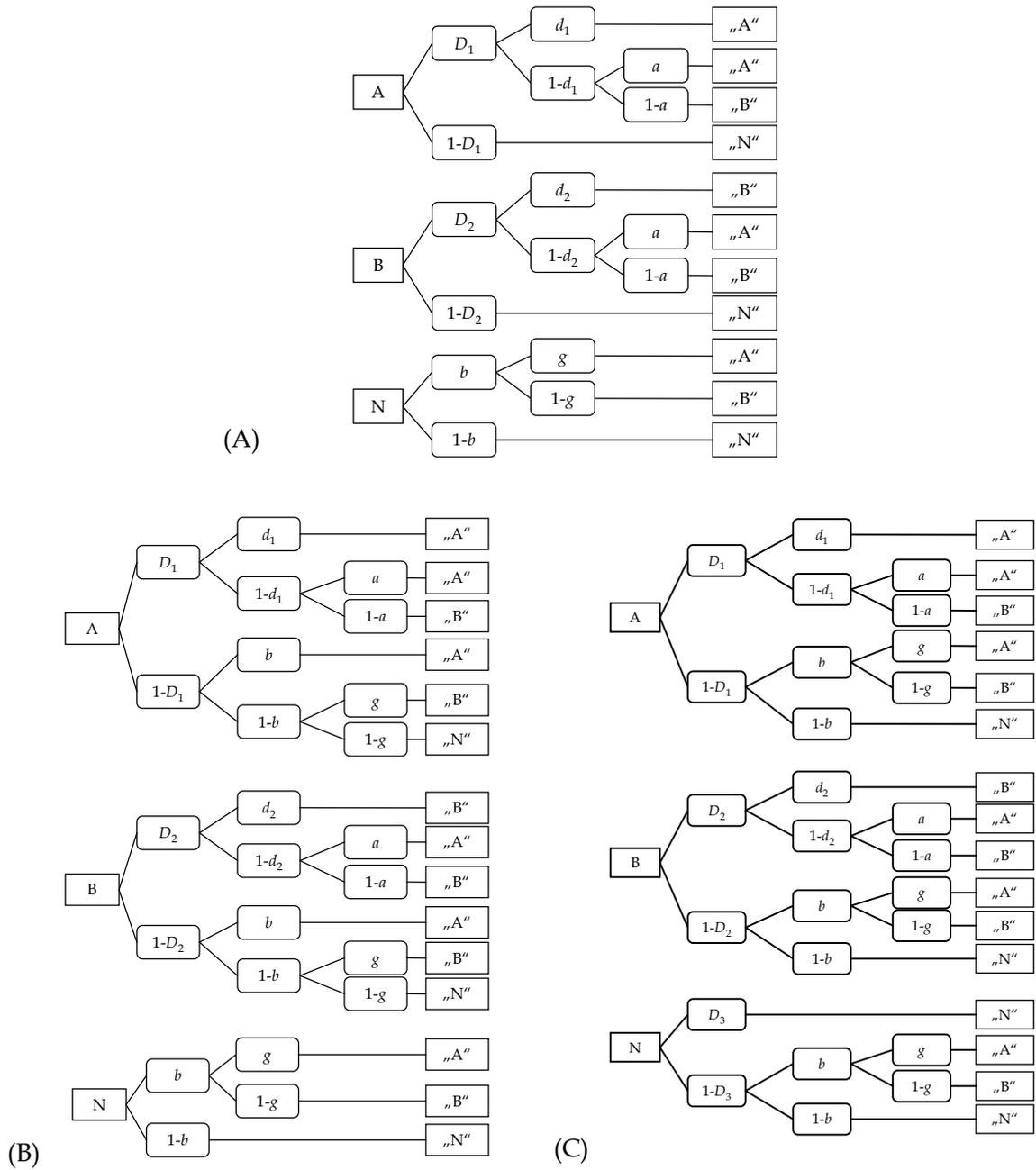


Abbildung 4.2: In dieser Grafik sind die drei im Text besprochenen Quellengedächtnismodelle dargestellt. Abbildungsteil (A) zeigt das Niedrigschwellen-Modell, in Teil (B) ist das Ein-Hochschwellen-Modell abgebildet, in Teil (C) das Zwei-Hochschwellen-Modell.

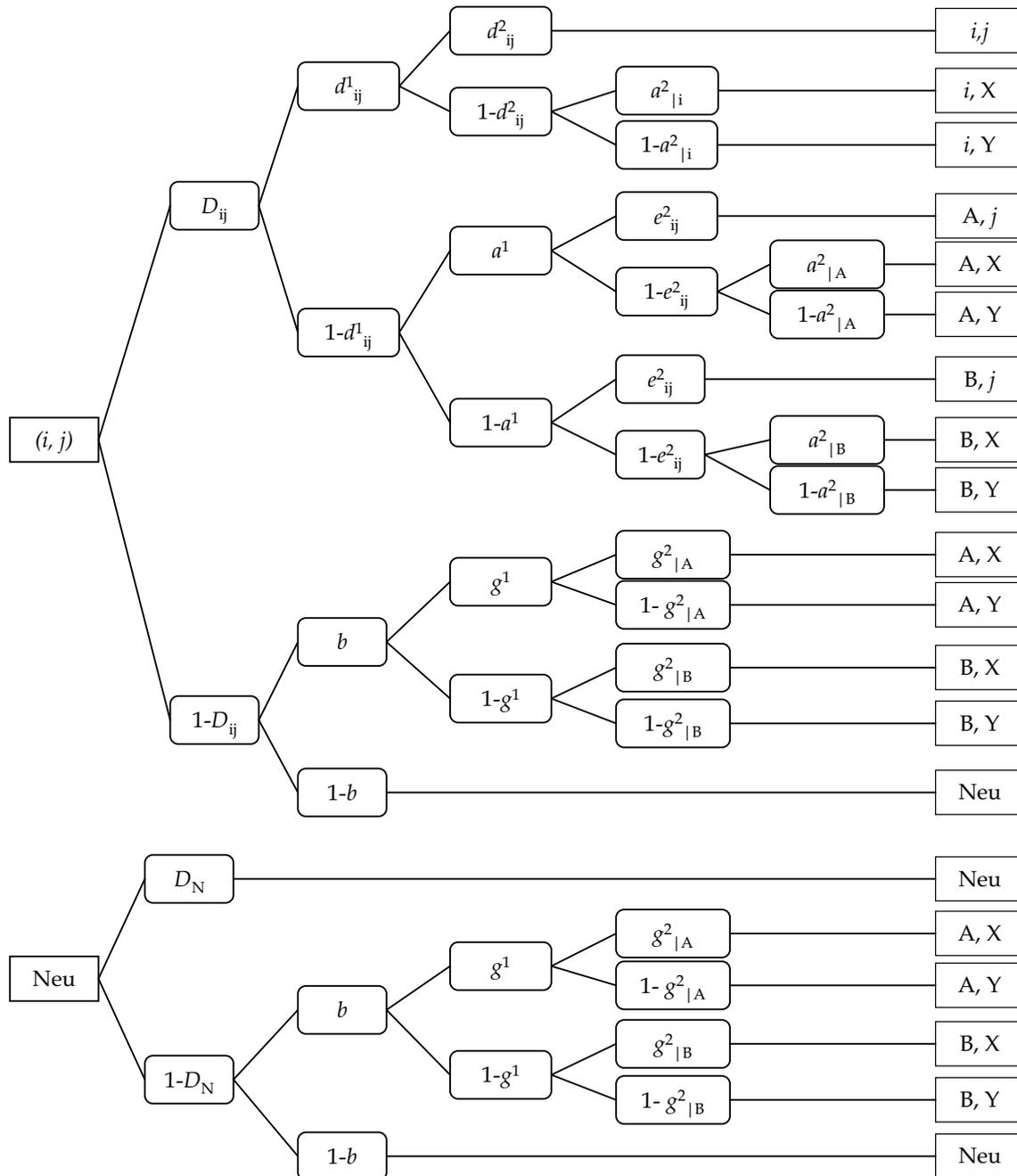


Abbildung 4.3: In dieser Grafik ist das zweidimensionale Quellengedächtnismodell nach Meiser und Brüder (2002) dargestellt. Die genaue Parameterbeschreibung kann dem Text entnommen werden.

Im Hinblick auf die Rateprozesse wird zunächst unterschieden, ob die Quellen für erinnerte Zielwörter ($a^1, a_{|A}^2, a_{|B}^2, a_{|j}^2$) oder für „alt“ geratene Zielwörter bzw. Distraktoren geraten werden ($g^1, g_{|A}^2, g_{|B}^2$). In beiden Fällen gibt es Rateparameter für die erste Quellendimension (a^1, g^1) und Rateparameter für die zweite Quellendimension ($a_{|A}^2, a_{|B}^2$). Die Parameter a^1 sowie g^1 stehen dabei für die Wahrscheinlichkeit, „A“ zu raten, mit der entsprechenden Gegenwahrscheinlichkeit würde „B“ geantwortet werden. Die Rateprozesse für die zweite Quellendimension werden in Abhängigkeit von der ersten Quellendimension formuliert. Die Parameter $a_{|A}^2$ bzw. $g_{|A}^2$ stehen für die Wahrscheinlichkeit, auf der zweiten Dimension „X“ zu raten, nachdem Quelle „A“ geraten wurde; $a_{|B}^2$ bzw. $g_{|B}^2$ stehen für diesen Prozess unter der Voraussetzung, dass für die erste Dimension „B“ geantwortet wurde. Die Rateparameter für die Quellendimensionen sind also zunächst abhängig von der Itemerkennung, des Weiteren werden die Rateparameter für die zweite Dimension konditional zu der ersten Dimension formuliert.

Auch dieses multinomiale Modell für gekreuzte Quellen ist global identifizierbar (Meiser & Bröder, 2002), zeigte eine gute Modellanpassung (Meiser & Bröder, 2002; Meiser, Sattler & von Hecker, 2007) und konnte zur Beantwortung von verschiedenen Forschungsfragen erfolgreich genutzt werden (Meiser & Bröder, 2002; Meiser et al., 2007).

In Anlehnung an diese beiden geschilderten Modelle soll im Folgenden ein neues Modell postuliert werden, welches es ermöglicht, aus Daten, die im Rekognitionsparadigma erhoben wurden, die Prozesse der assoziativen Enkodierung, des assoziativen Abrufs, der Itemerkennung sowie Rateprozesse zu dekomponieren. Auch dieses neue Modell soll als multinomiales Modell konzipiert werden.

4.2.3 Das Modell zur Assoziationsdefizit-Hypothese

Mit dem Ziel, die Prozesse der assoziativen Enkodierung und des assoziativen Abrufs voneinander zu trennen, das damit verbundene Dekomponierungsproblem zu lösen und letztendlich die Lokalisationsfrage des assoziativen Defizits zu beantworten, wird in diesem Kapitel ein neues multinomiales Modell vorgestellt. Zunächst sollen noch einmal kurz die Anforderungen aufgegriffen werden, die ein solches Modell erfüllen muss. Dabei wird auf die in Kapitel 4.2 formulierten Kriterien zurückgegriffen:

- (1) Schätzung von Parametern für die Erinnerung an einzelne Items (Elemente) und Parametern für die Assoziation der Elemente,
- (2) Trennung von Enkodierung und Abruf der Assoziation,
- (3) Analyse von Rekognitionsdaten,
- (4) keine spezifischen Anforderungen an das Stimulusmaterial.

Versucht man diese Kriterien bzw. die postulierten Prozesse in Parameter zu fassen, so muss ein Modell im Groben folgende Parameter bzw. Parameterklassen aufweisen:

(a) einen Enkodierungsparameter der Assoziation zweier Elemente, (b) einen Parameter für den Abruf der Assoziation, (c) mindestens einen Itemgedächtnisparameter, (d) Distraktorentdeckungsparameter sowie (e) Rateparameter. Im Vergleich zu diesem - nur sehr grob skizzierten - Parameterraum ergeben sich aus dem bisher genutzten Paradigma relativ wenige freie Kategoriehäufigkeiten. Sowohl im Itemrekognitionstest als auch im assoziativen Rekognitionstest wird lediglich eine Alt-neu-Entscheidung getroffen, damit liegt jeweils nur ein unabhängiger Datenpunkt vor, d.h. $M = 2$. Die grobe Schätzung der Anzahl erforderlichen Parameter übersteigt damit die Anzahl der freien Kategoriehäufigkeiten, es liegt ein überparametrisiertes Modell vor. Zur Lösung dieses Problems kann der von Erdfelder (2000) vorgeschlagene Lösungsweg bei Identifizierbarkeitsproblemen gewählt werden: die Veränderung des Testparadigmas (siehe auch Klauer, Musch & Naumer, 2000). Diese Lösung wird im Folgenden diskutiert.

Der kombinierte Item- und assoziative Rekognitionstest: ein neues Paradigma

Über eine Modifikation des Testparadigmas kann es gelingen, eine größere Datenbasis zu erlangen und damit mehr Freiheitsgrade für eine Modellierung zu erhalten. Erdfelder (2000) beschreibt dieses Verfahren wie folgt: „Die Kunst „adäquater“ Erweiterungen des empirischen Paradigmas besteht gerade darin, zusätzliche unabhängige Beobachtungsbedingungen zu finden, in die jedoch keine zusätzlichen psychologischen Prozesse einfließen, sondern nur diejenigen, die auch schon im Ausgangsparadigma wirksam sind.“ (Erdfelder, 2000, S.211).

Um dieses Ziel zu erreichen, liegt es für den vorliegenden Fall nahe, die beiden bisher unabhängigen Testformate zu kombinieren. Hebt man die Trennung zwischen Item- und assoziativem Test auf, so könnte eine kombinierte Testphase wie folgt aufgebaut sein: Es werden die aus dem 2AFC-Itemrekognitionstest stammenden Target-Distraktor- und Distraktor-Target-Kombinationen dargeboten, gleichzeitig aber auch alte intakte Wortpaare und neu kombinierte Wortpaare aus der bisherigen assoziativen Testphase. Um alle möglichen Kombinationen von Items in das Paradigma aufzunehmen und eine ausbalancierte Datenmatrix zu erreichen, scheint es sinnvoll, auch die Kombination aus zwei Distraktoren zu integrieren.

Während in der Lernphase weiterhin ohne Veränderung Wortpaare dargeboten werden, liegen in der Testphase des modifizierten Paradigmas nun anstatt der bisher jeweils zwei Testkombinationen fünf unterschiedliche Itemkombinationen vor: (1) Target-Distraktor-Kombinationen, (2) Distraktor-Target-Kombinationen, (3) alte intakte Wortpaare, die in dieser Kombination gelernt wurden, (4) neu kombinierte Wortpaare, bestehend aus zwei alten Items, die neu miteinander kombiniert wurden, sowie (5) Distraktor-Distraktor-Kombinationen.

Die Aufgabe des Probanden in diesem Testverfahren ist es, zunächst eine Entscheidung über den Itemzustand zu treffen, d.h. zu beurteilen, ob das linke Wort ein altes Wort ist

und ob dies für das rechte Wort zutrifft. Anders als im 2AFC-Test kann es in diesem Testformat sowohl möglich sein, dass beide Wörter alt sind, es kann aber auch der Fall sein, dass kein Wort alt ist.

Nach der Alt-neu-Entscheidung auf Itemebene muss für zwei als alt beurteilte Items angegeben werden, ob deren Kombination eine alte Kombination darstellt (d.h. die beiden Items wurden in dieser Kombination gelernt) oder ob die beiden Wörter neu miteinander kombiniert wurden. Diese zweite Beurteilungsphase entspricht dem assoziativen Rekognitionsurteil.

Insgesamt geben die Probanden also zwei bzw. drei Antworten: Die Alt-neu-Einschätzung des rechten Wortes, das Alt-neu-Urteil für das linke Wort und gegebenenfalls eine Alt-neu-Entscheidung für die Kombination der Wörter. Im Folgenden werden die Kombinationen der Antworten abgekürzt. Das Kürzel „AAA“ steht für die Kombination einer „Alt“-Antwort auf das erste Wort, einer „Alt“-Antwort auf das zweite Wort und einer „Alt“-Antwort auf die Kombinationsfrage. Werden beide Wörter als alt eingestuft, die Kombination als neu beurteilt, so wird dies durch das Kürzel „AAN“ wiedergegeben. „AN“ bzw. „NA“ steht für eine „Alt“-Antwort auf das erste und „Neu“-Antwort auf das zweite Wort bzw. umgekehrt. Zwei „Neu“-Antworten auf beide Wörter werden mit „NN“ gekennzeichnet.

Zwar wird dieses Paradigma hier als „neues Paradigma“ beschrieben - und daran soll auch festgehalten werden - es muss aber darauf hingewiesen werden, dass Buchler et al. (2008) eben dieses Paradigma auch verwenden und von dieser Autorengruppe das Paradigma erstmals publiziert wurde. Da die Arbeit von Buchler et al. (2008) inhaltlich nur geringe Berührungspunkte mit der vorliegenden Thematik hat, wird auf diese Studie nicht weiter eingegangen. Die parallele Entwicklung dieses Paradigmas von zwei unabhängigen Forschergruppen weist darauf hin, dass dieses neue Paradigma für die Modellierung nützlich ist und es empirisch sinnvoll eingesetzt werden kann.

Durch die Einführung dieses neuen Testparadigmas wird die Trennung des assoziativen Tests und des Itemrekognitionstest aufgehoben, die Lernphase bleibt vergleichbar zu dem bisherigen Verfahren. Da in diesem angereicherten Paradigma fünf unterschiedliche Itemtypen im Test vorliegen und für jeden Itemtyp fünf Antwortmöglichkeiten in Frage kommen (AAA,AAN,AN,NA,NN), entsteht die in Tabelle 4.1 dargestellte Datenmatrix. Für die freien Kategoriehäufigkeiten gilt $M = 20$. Damit sind die notwendigen Freiräume zur Modellierung der zentralen kognitiven Prozesse gegeben.

Mit der Einführung eines neuen Paradigmas ist es wichtig zu überdenken, inwieweit diese Modifikation auch Veränderungen an den beteiligten kognitiven Prozessen hervorruft und welche weiteren Vor- und Nachteile neben der Gewinnung von Freiheitsgraden für die Modellierung zu beachten sind.

Im Hinblick auf die beteiligten kognitiven Prozesse kann festgehalten werden, dass für die Item-Entscheidung Erinnerungsprozesse an das Einzelwort benötigt werden, die

Tabelle 4.1: In dieser Tabelle ist die Datenmatrix des modifizierten Paradigmas dargestellt

Wahre Itemkat.	Antwort der Versuchsperson				
	Altes WP „AAA“	Neu komb. WP „AAN“	Target Dis. „AN“	Dis. Target „NA“	Dis.Dis. „NN“
Altes WP	$Y_{AAA,AAA}$	$Y_{AAA,AAN}$	$Y_{AAA,AN}$	$Y_{AAA,NA}$	$Y_{AAA,NN}$
Neu komb. WP	$Y_{AAN,AAA}$	$Y_{AAN,AAN}$	$Y_{AAN,AN}$	$Y_{AAN,NA}$	$Y_{AAN,NN}$
Target Dis.	$Y_{AN,AAA}$	$Y_{AN,AAN}$	$Y_{AN,AN}$	$Y_{AN,NA}$	$Y_{AN,NN}$
Dis. Target	$Y_{NA,AAA}$	$Y_{NA,AAN}$	$Y_{NA,AN}$	$Y_{NA,NA}$	$Y_{NA,NN}$
Dis. Dis.	$Y_{NN,AAA}$	$Y_{NN,AAN}$	$Y_{NN,AN}$	$Y_{NN,NA}$	$Y_{NN,NN}$

vergleichbar sind zu den Gedächtnisprozessen eines 2AFC-Tests. Meines Erachtens liegen keine Anhaltspunkte vor, die auf qualitativ unterschiedliche Gedächtnisprozesse hindeuten. Veränderungen könnten allerdings in den Rateprozessen zu beobachten sein, da den Probanden andere Antwortmöglichkeiten zur Verfügung stehen und bekannt ist, dass die Antwortoptionen eines Tests Ratetendenzen beeinflussen können (Hicks & Marsh, 1999; siehe auch Kapitel 6.1). Da die Rateparameter aber nicht von zentralem Interesse sind und sich das assoziative Defizit nicht auf diese bezieht, sollte eine Veränderung der Ratetendenzen für das Forschungsvorhaben nicht weiter kritisch sein. Für den assoziativen Rekognitionsprozess kann m.A.n. dieses Fazit ebenfalls beibehalten werden. Hier sollte sich nicht einmal ein Einfluss auf die Rateparameter ergeben, da die Antwortoptionen (altes intaktes oder neu kombiniertes Wortpaar) identisch sind zu der Antwortfassung des assoziativen Rekognitionsparadigmas. Die Gedächtnisprozesse, die an dem assoziativen Defizit älterer Menschen beteiligt sind und die bisher über den Item- bzw. assoziativen Rekognitionstest gemessen wurden, sollten auch dem neuen Paradigma zugrunde liegen.

Gegenüber dem bisherigen Testverfahren weist das neue Paradigma den Vorteil auf, dass Reihenfolgeeffekte bei der Testdurchführung nicht mehr kontrolliert werden müssen. Bisher wurde zwischen den Probanden ausbalanciert, ob zuerst der assoziative Test oder der Itemtest erhoben wurde. Dieses Verfahren wird nun überflüssig. Auch ist man mit dem neuen Paradigma nicht mehr an bisher ungeprüfte Annahmen z.B. über die Vergleichbarkeit der Ratetendenzen in beiden Paradigmen (siehe Kapitel 4.1) gebunden. Diesen Vorteilen stehen m.E. keine Nachteile gegenüber, die durch das neue Paradigma entstehen. Der einzig denkbare Einwand könnte sein, dass die direkte Vergleichbarkeit zu den bisherigen Studien und Experimenten nicht gegeben ist. Allerdings kann dagegen argumentiert werden, dass das assoziative Defizit älterer Menschen nicht von der Verwendung eines bestimmten Testparadigmas abhängen sollte, wenn defizitäre assoziative Gedächtnisprozesse dem Problem zugrunde liegen. Daher kann in dem neuen Paradigma eher die Chance gesehen werden, den defizitären Bindungsmechanismus zu replizieren und dessen Unabhängigkeit von arbiträren Randbedingungen zu sichern. Als kritisch muss diese Argument nicht eingestuft werden.

Die Spezifizierung des Modells zur Assoziationsdefizit-Hypothese

Nachdem das Testparadigma vorgestellt und diskutiert wurde, wird im Folgenden das Modell, d.h. die Parameter und deren Verknüpfungen erläutert. Die Modellstruktur wird in mehreren Abbildungen visualisiert. Da das Modell mit dem Ziel entwickelt wurde, die Assoziationsdefizit-Hypothese genauer zu spezifizieren, wird das Modell im Folgenden auch als ‚ADH-Modell‘ bezeichnet.

Auch bei diesem multinomialen Modell handelt es sich um ein verbundenes Modell, da $k = 5$ verschiedene Itemtypen in der Testphase vorliegen und damit fünf Modellbäume postuliert werden: Ein Baum für alte intakte Wortpaare, einer für neu kombinierte Wortpaare, ein Baum für Target-Distraktor-Kombinationen sowie Distraktor-Target-Verbindungen und ein Baum für die Kombination zweier Distraktoren. In diesem Modell stellen die unterschiedlichen Itemkategorien den Ausgangspunkt der Modellierung dar (vgl. Bayen et al., 1996; Meiser & Bröder, 2002).

Zuerst wird das Baummodell für alte, intakte Wortpaare erläutert, grafisch dargestellt ist dieses in Abbildung 4.4. Für diese Itemkategorie können fünf unterschiedliche Antworten resultieren (AAA,AAN,AN,NA,NN), wobei die „AAA“-Antwort korrekt wäre. Die Antworten der Versuchspersonen sind im Modell jeweils ganz rechts in Rechtecken dargestellt. Die Prozesse, die zu diesen Antwortmustern führen, werden nun erläutert.

In Anlehnung an das Enkodierung-Abruf-Modell wird auch in diesem Modell davon ausgegangen, dass in der Lernphase die beiden Wörter eines Wortpaares mit der Wahrscheinlichkeit c assoziiert und als Assoziation (d.h. Cluster) enkodiert und gespeichert werden. Ein als Assoziation enkodiertes und gespeichertes Wortpaar kann dann mit der Wahrscheinlichkeit r im Test abgerufen werden. Dies führt dazu, dass die Probanden korrekterweise sowohl für das linke Wort als auch für das rechte Wort eine „Alt“-Antwort geben, auch für die Assoziationsfrage ist dies der Fall („AAA“).

Gelingt es nicht, ein assoziiertes Wortpaar abzurufen, so wird analog zum Enkodierung-Abruf-Modell angenommen, dass nach einem Scheitern dieses ‚Alles-oder-nichts‘-Prozesses auch keines der Einzelwörter abgerufen werden kann. Entweder ist das assoziierte Wortpaar als Assoziation zugänglich, oder gar nicht. Schlägt der Abruf des Wortpaares also fehl, so muss der Proband für beide Einzelwörter und für die Assoziation der beiden Wörter raten²⁰. Aufgrund von Rateprozessen kann jede der fünf möglichen Antwortkombinationen vorkommen.

Wurde ein Wortpaar nicht als solches enkodiert und gespeichert ($1 - c$), so kann es dennoch der Fall sein, dass die beiden Einzelwörter getrennt voneinander enkodiert und gespeichert wurden und im Test abrufbar sind (vgl. Enkodierung-Abruf-Modell). Die Wahrscheinlichkeit eines erfolgreichen Abrufs im Test wird für das linke Einzelwort mit w_1 notiert (der Index i steht dabei für die Position des Wortes im Wortpaar), für das

²⁰Die Rateprozesse werden im Folgenden noch genauer ausgeführt.

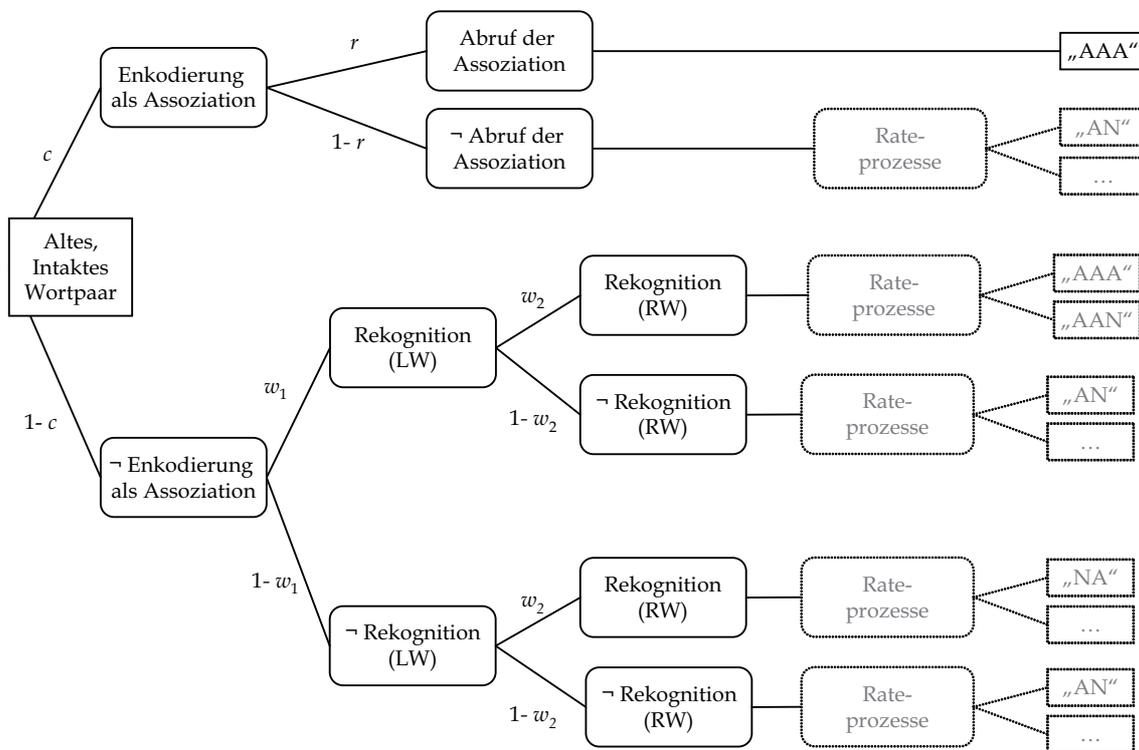


Abbildung 4.4: In dieser Abbildung ist das Verarbeitungsbaummodell für alte, intakte Wortpaare dargestellt. Die genaue Parameterbeschreibung kann dem Text entnommen werden. Das Modell zeigt eine vereinfachte Darstellung, da die Rateparameter nicht genauer spezifiziert werden.

rechte Wort mit w_2 . Die Parameter w_i trennen - vergleichbar zu dem Parameter u im Enkodierung-Abruf-Modell - nicht zwischen den Prozessen Enkodierung, Speicherung und Abruf, sie sind als Hybridparameter zu verstehen. Die Aufteilung auf zwei von der Position abhängige Rekognitionsparameter hängt mit dem Ziel zusammen, ein möglichst breit anwendbares Modell zu postulieren (Kriterium 4): Werden zwei Itemklassen verwendet, von denen bekannt ist, dass sie unterschiedlich gute Rekognitionsleistungen hervorrufen (z.B. Gesichter vs. Namen oder Bilder vs. Wörter), so sollte dieser Gedächtniseffekt auch durch unterschiedliche Itemrekognitionsparameter aufgefangen werden. Des Weiteren ist es möglich, dass bei sehr kurzen Präsentationszeiten von Wortpaaren aufgrund der Leserichtung von links nach rechts, das linke Wort evtl. eine höhere Chance hat, erfolgreich enkodiert und gespeichert zu werden. Dieser Effekt sollte durch unterschiedliche Positionsparameter repräsentiert sein.

Vergleichbar zum Enkodierung-Abruf-Modell ist es möglich, dass die Rekognition beider Wörter gelingt ($w_1 \cdot w_2$), dass nur eines der beiden Wörter rekogniziert werden kann ($w_1 \cdot (1 - w_2)$ bzw. $(1 - w_1) \cdot w_2$) oder aber dass kein Wort erinnert wird ($(1 - w_1) \cdot (1 - w_2)$), wobei die Abfolge der Parameter w_1 und w_2 beliebig gewählt ist und ohne Bedeutungsveränderung auch vertauscht werden könnte. Unabhängig davon, welches dieser Ereignisse eintritt, muss in diesen Fällen für die Assoziationsfrage immer geraten werden, denn die Information über die Kombination der beiden Wörter wurde nicht enkodiert ($1 - c$). Abhängig davon, welche Itemrekognitionsprozesse scheitern, muss für diese Wörter ebenfalls geraten werden.

Damit wurde der erste Verarbeitungsbaum für alte intakte Wortpaare erläutert (siehe Abbildung 4.4). Die unterschiedlichen Rateparameter werden später in diesem Kapitel genauer ausgeführt. An die gerade beschriebene untere Hälfte des ersten Baummodells kann sehr gut angeknüpft werden, um das Modell für die zweite Itemkategorie der neu kombinierten Wortpaare zu erläutern (siehe Abbildung 4.5). Bei neu kombinierten Wortpaaren ist es möglich, dass mit der Wahrscheinlichkeit w_1 das erste Wort als altes Wort erinnert werden kann und dies auch für das zweite Wort gelingt (w_2). Falls eines der neu kombinierten Wörter als Assoziation enkodiert wurde und diese nun auch abgerufen werden kann, so führt dies dazu, dass mit der Wahrscheinlichkeit a die Kombination der beiden Wörter in der Testphase als „Distraktorkombination“ zurückgewiesen wird. Die Probanden erkennen aufgrund des erfolgreichen Abrufs der ursprünglichen Assoziation („recall-to-reject“), dass die vorliegende Kombination nicht der Kombination der Lernphase entspricht.

Der Assoziationsabrufs kann entweder durch das erste Wort, oder durch das zweite Wort initiiert werden, beide Fälle werden mit der Wahrscheinlichkeit a notiert. Sobald mindestens der Abruf einer Assoziation gelingt, wird ein neu kombiniertes Wortpaar korrekterweise auch als neue Kombination beurteilt ($w_1 \cdot w_2 \cdot a$ bzw. $w_1 \cdot w_2 \cdot a^2$). Können zwar beide Wörter korrekt erinnert werden ($w_1 \cdot w_2$), für keines der beiden Wörter kann aber die ursprüngliche Assoziation der Lernphase abgerufen werden ($w_1 \cdot w_2 \cdot (1 - a)^2$), so wird für die Assoziationsfrage geraten.

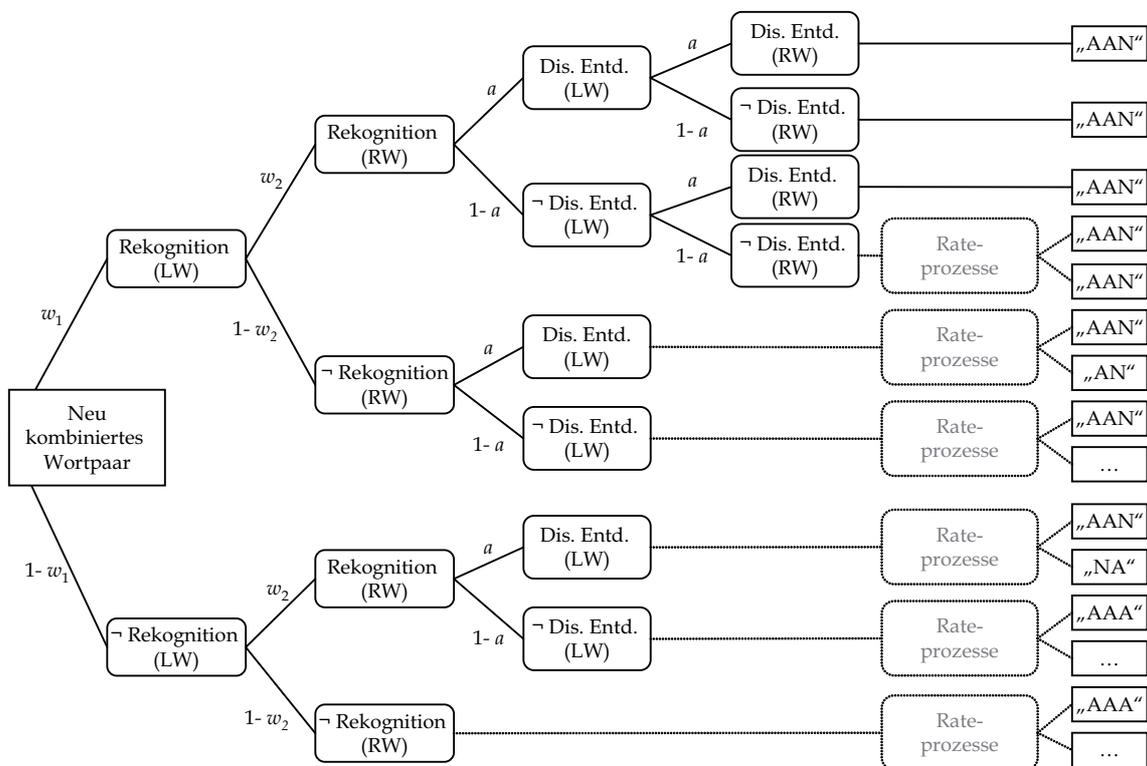


Abbildung 4.5: In dieser Grafik ist das Verarbeitungsbaummodell für neu kombinierte Wortpaare dargestellt. Es handelt sich um eine vereinfachte Abbildung, da die Rateparameter nicht genauer ausdifferenziert sind. Die Parameterbeschreibung kann dem Text entnommen werden.

Auch in diesem Teil des Modells sind alle möglichen Kombinationen der Rekognition der beiden Wörter dargestellt: Die erfolgreiche Rekognition beider Wörter ($w_1 \cdot w_2$); die erfolgreiche Rekognition des ersten linken Wortes, aber nicht des zweiten rechten Wortes ($w_1 \cdot (1 - w_2)$); der umgekehrte Fall, in dem die Rekognition des ersten Wortes misslingt, das zweite aber erinnert werden kann ($(1 - w_1) \cdot w_2$) sowie letztendlich der Fall, dass keines der beiden Wörter erinnert wird ($(1 - w_1) \cdot (1 - w_2)$). Eine Distraktorentdeckung der Assoziation (a) kann immer dann erfolgen, wenn ein Wort erfolgreich erinnert wurde. Falls die Rekognition eines Items oder die Distraktorentdeckung nicht gelingt, so wird für die entsprechenden Fragen geraten.

Aus der Erklärung des zweiten Verarbeitungsbaums, kann direkt auf den dritten Baum für Target-Distraktor-Kombinationen in der Testphase übergegangen werden. Im Wesentlichen entspricht dieser Baum dem gerade geschilderten Verarbeitungsbaum für ein neu kombiniertes Wortpaar, allerdings wird anstatt eines Rekognitionsprozesses für das zweite Wort (w_2) ein Prozess der Distraktorentdeckung (D) postuliert (siehe Abbildung 4.6). Dabei wird vergleichbar zu den berichteten Quellengedächtnismodellen (Bayen et al., 1996; Meiser & Bröder, 2002) eine Zwei-Hochschwellen-Annahme getroffen (siehe Kapitel 4.2.2).

Für Target-Distraktor-Kombinationen wird u.a. dann die korrekte Antwort gegeben, wenn das linke Wort als altes Wort identifiziert wird und das rechte Wort als Distraktor erkannt wird ($w_1 \cdot D$). Ob zusätzlich zu der Itemerinnerung und Distraktorentdeckung auch der Abruf der Assoziation aus der Lernphase erfolgreich ist oder nicht (a bzw. $1 - a$), ändert am Antwortverhalten des Probanden nichts, da aufgrund der Distraktorentdeckung die vorliegende Wortkombination in jedem Fall als neue Kombination eingestuft wird. Der Abruf der Assoziation aus der Lernphase ist nur dann ausschlaggebend, wenn die Distraktorentdeckung nicht gelingt ($w_1 \cdot (1 - D)$). Auch in diesem Modellteil werden alle Kombinationen aus erfolgreicher oder missglückter Rekognition und Distraktorentdeckung modelliert.

Der nächste Modellbaum für die Kombination von Distraktor und Target in der Testphase enthält dieselben Prozesse und Annahmen wie der gerade geschilderte Baum der Target-Distraktor-Kombination. Der einzige Unterschied besteht darin, dass die Position von Target-Wort und Distraktor-Wort vertauscht wurde und damit anstatt w_1 der Parameter w_2 in diesem vierten Baum auftritt.

Bevor auf den letzten Verarbeitungsbaum eingegangen wird, sollen nun die unterschiedlichen Rateprozesse anhand einer vorliegenden Target-Distraktor-Kombination erklärt werden.

Äquivalent zu der Unterscheidung von w_1 und w_2 für die Rekognition des linken und rechten Wortes werden auch die Rateparameter dahingehend differenziert, ob für das linke Wort geraten wird (g_1) oder für das rechte Wort (g_2). Auch hier wurde die Reihenfolge wieder beliebig festgelegt. Diese Differenzierung ist zum einen sinnvoll, wenn wie im oben genannten Fall unterschiedliche Erinnerungsraten an das linke vs. rechte Wort modelliert

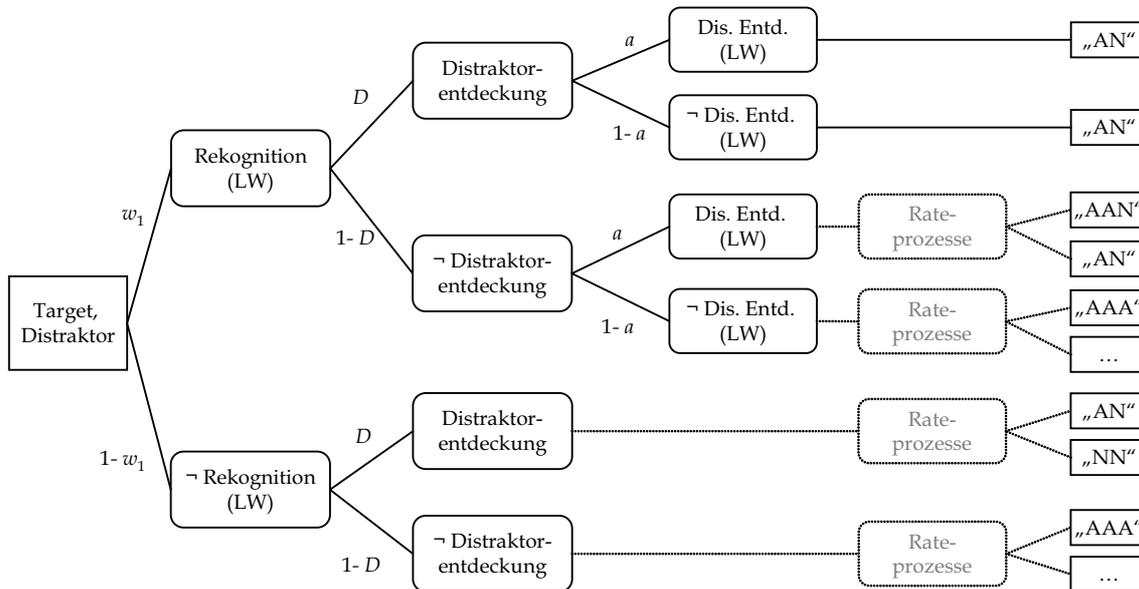


Abbildung 4.6: Hier ist eine vereinfachte Darstellung des Verarbeitungsbaummodells für Target-Distraktor-Kombinationen abgebildet. In der Abbildung fehlt die genaue Darstellung der Rateparameter, diese sind nur vereinfacht abgebildet. Die genaue Beschreibung der Parameter kann dem Text entnommen werden.

werden sollen. Viele Manipulationen, die sich auf die Erinnerungsleistung auswirken, zeigen auch Effekte in den Rateparametern (‘Mirror-Effekt’, siehe Glanzer & Adams, 1985; Glanzer & Bowles, 1976). Zum anderen ist diese Unterscheidung dann notwendig, wenn das Verhältnis von Target-Distraktor- zu Distraktor-Target-Kombinationen nicht gleich ist, denn auch die Basisraten verschiedener Itemklassen im Test zeigen Auswirkungen auf die Rateparameter (siehe z.B. Buchner, Erdfelder & Vaterrodt-Plünnecke, 1995).

Des Weiteren wird unterschieden, ob *nur* für das linke bzw. rechte Wort ($g_{10}; g_{20}$) geraten wird, oder ob auch für das jeweils andere Wort geraten wird ($g_{11}; g_{21}$). Diese Unterscheidung wird in Anlehnung an die unterschiedlichen Rateparameter a sowie g im zweidimensionalen Quellengedächtnismodell (Meiser & Bröder, 2002) getroffen. Diese Differenzierung ist vor allem dann notwendig, wenn die ersten beiden Itemkategorien alter intakter und neu kombinierter Wortpaare sehr viel häufiger getestet werden, als Target-Distraktor- und Distraktor-Target-Kombinationen.²¹ Liegt das Verhältnis $AAA + AAN > AN + NA$ vor, d.h. es werden häufiger zwei alte Wörter getestet als eine Alt-neu-Kombination, so sollten die Probanden für g_{11} und g_{21} liberalere Antworttendenzen entwickeln als in dem Fall $AAA + AAN < AN + NA$.

Damit werden vier Itemrateparameter postuliert, die über g_{ij} notiert werden können.

²¹Diese Unterscheidung mag im ersten Moment sehr konstruiert klingen, da aber die zentralen Parameter c und r nur im ersten Teil des Modells vorkommen, kann es aus Gründen der Teststärke durchaus sinnvoll sein, die Kombination von alten intakten Wortpaaren häufiger zu testen als die anderen Itemkategorien. Um eine dadurch entstehende Anpassung der Rateprozesse zu kontrollieren, ist die Unterscheidung zwischen den Prozessparametern durchaus sinnvoll.

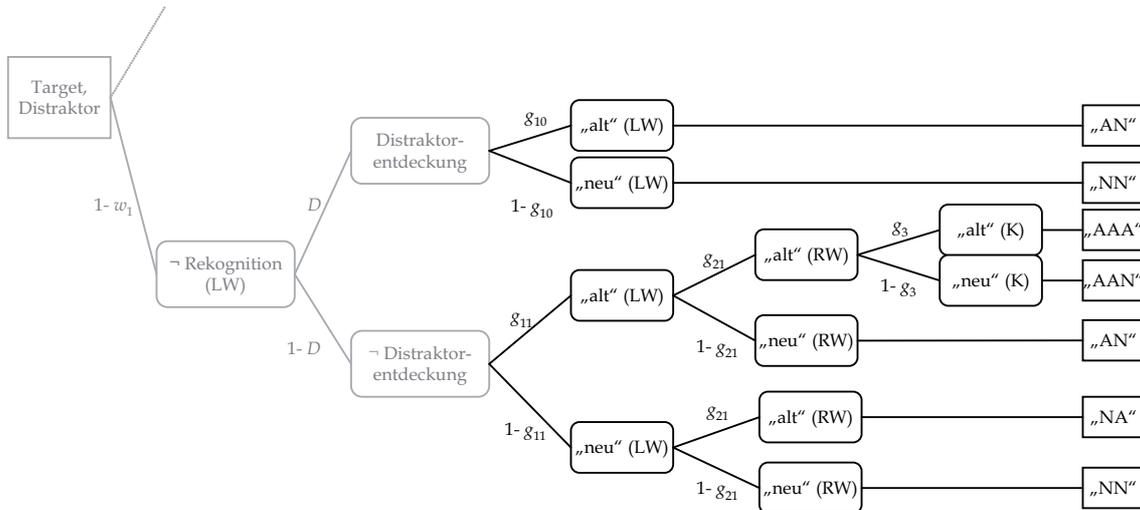


Abbildung 4.7: Diese Grafik zeigt einen Ausschnitt des Verarbeitungsbaummodells für Target-Distraktor-Kombinationen. Im Text werden anhand dieses Modellausschnitts die Rateparameter genauer erläutert.

Der Index i gibt an, für welche Wortposition (links oder rechts) geraten wird ($i = 1, 2$), der Index j kennzeichnet, ob nur für dieses eine Wort geraten wird ($j = 0$) oder ob zusätzlich auch für das andere Wort geraten werden muss ($j = 1$).

Neben den vier Rateparametern für die Items wird zusätzlich ein weiterer Rateparameter für die Kombinationsfrage postuliert. Dieser Parameter g_3 steht für die Wahrscheinlichkeit, die Wortkombination aufgrund von Rateprozessen als alt zu beurteilen.

Anhand des Target-Distraktor-Baums (siehe Abbildung 4.7) werden die unterschiedlichen Rateprozesse innerhalb der Modellstruktur erläutert. Wenn es nicht gelingt, sich an das erste Wort zu erinnern, das zweite Wort aber als Distraktor identifiziert werden kann ($(1 - w_1) \cdot D$), so befindet sich der Proband im Hinblick auf das erste Wort in Unsicherheit. Mit der Wahrscheinlichkeit g_{10} wird dann „alt“ für dieses Item geraten. Der Rateprozess bezieht sich auf das linke Wort ($i = 1$) und da für das andere Wort nicht geraten wird, ist $j = 0$. Kann allerdings weder das erste Wort noch das zweite Wort korrekt klassifiziert werden ($(1 - w_1)(1 - D)$), so muss für beide Wörter geraten werden, es folgt daraus $j = 1$. Der Rateprozess des ersten Wortes wird dabei mit g_{11} notiert, der des zweiten Wortes mit g_{21} . Wurde für beide Wörter „alt“ geraten ($g_{11} \cdot g_{21}$), so wird die Frage nach der Kombination der beiden Wörter gestellt und mit der Wahrscheinlichkeit g_3 wird für die Kombination „alt“ geraten, mit der Gegenwahrscheinlichkeit $(1 - g_3)$ rät der Proband „neu“.

Nachdem nun auch die Rateprozesse erläutert wurden, sind in den folgenden Abbildungen 4.10, 4.11 und 4.9 die vollständigen Baummodelle für alte intakte Wortpaare, neu kombinierte Wortpaare sowie Target-Distraktor-Kombinationen dargestellt. Das Modell ist für Distraktor-Target-Kombinationen bis auf die Indices des Itemrekognitionsparame-

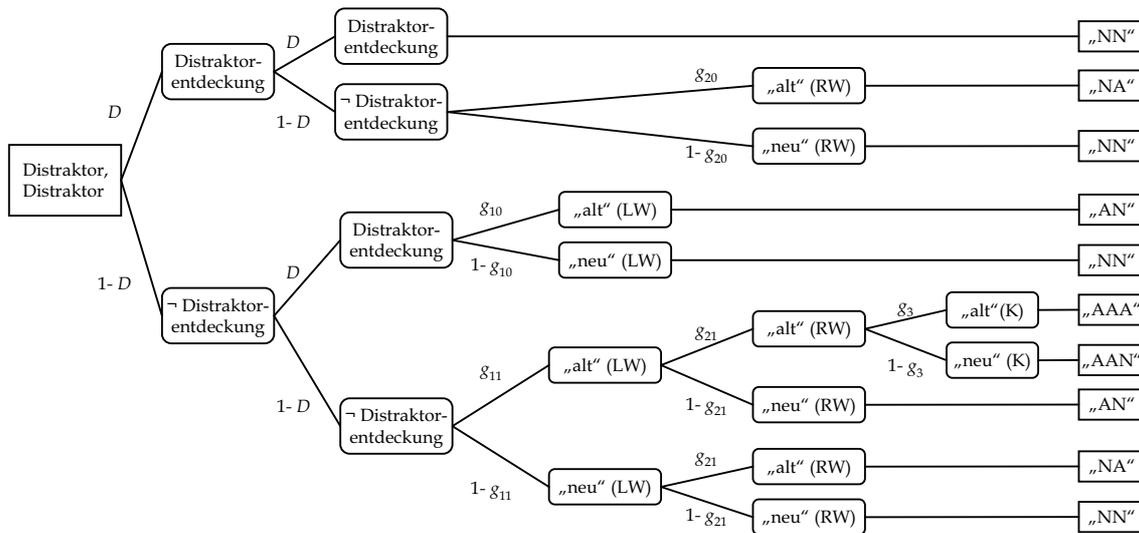


Abbildung 4.8: In dieser Abbildung ist das vollständige Entscheidungsbaummodell für die Kombination von zwei Distraktoren dargestellt.

ters und der Rateparameter identisch ist zu dem Modell für Target-Distraktor-Paare, es kann daher dem Anhang A entnommen werden.

Der fünfte und letzte Baum des Modells beschreibt das Antwortverhalten für die im Test vorliegende Kombination von zwei Distraktoren und enthält Prozesse der Distraktorentdeckung sowie Rateprozesse (siehe Abbildung 4.8). Mit der Wahrscheinlichkeit D wird das erste Wort als Distraktor entdeckt, mit derselben Wahrscheinlichkeit gelingt dies auch für das zweite Wort. In diesem Fall würde der Proband die korrekte Antwort „beide Wörter sind neu (NN)“ geben. Kann nur eines der neuen Wörter als Distraktor entdeckt werden, so setzen Rateprozesse ein, die zu unterschiedlichen Antwortkombinationen führen. Wird das erste Wort zwar entdeckt, nicht aber das zweite ($D \cdot (1 - D)$), so wird mit der Wahrscheinlichkeit g_{20} für das zweite Wort „alt“ geraten und der Proband würde die Antwort „linkes Wort neu, rechtes Wort alt“ geben. Andere Antworten sind natürlich möglich, auch für den Fall, dass keines der beiden Wörter als Distraktor erkannt wird.

Das gesamte Modell besteht damit aus fünf Bäumen ($k = 5$), die jeweils in Abhängigkeit von dem im Test vorliegenden Itemtyp strukturiert sind. Das Modell kann durch den Parameterraum $\Omega = [0, 1]^{11} = (c, r, w_1, w_2, D, a, g_{10}, g_{20}, g_{11}, g_{21}, g_3)$ beschrieben werden. Den $S = 11$ Parametern stehen $M = \sum_{k=1}^K (J_k - 1) = 20$ freie Modellkategorien gegenüber, daraus ergeben sich $df = 9$ Freiheitsgrade für das Modell, und es ist damit prinzipiell identifizierbar. In Tabelle 4.2 sind die Parameter noch einmal aufgelistet, und deren Interpretation wird kurz erläutert. Die eqn-Datei mit den Modellgleichungen kann Anhang B entnommen werden.

Wie für alle multinomialen bzw. mathematischen Modelle gilt dabei auch für das vorliegende Modell, dass die Modellstruktur und die Parameterauswahl auf bestimmten, nicht

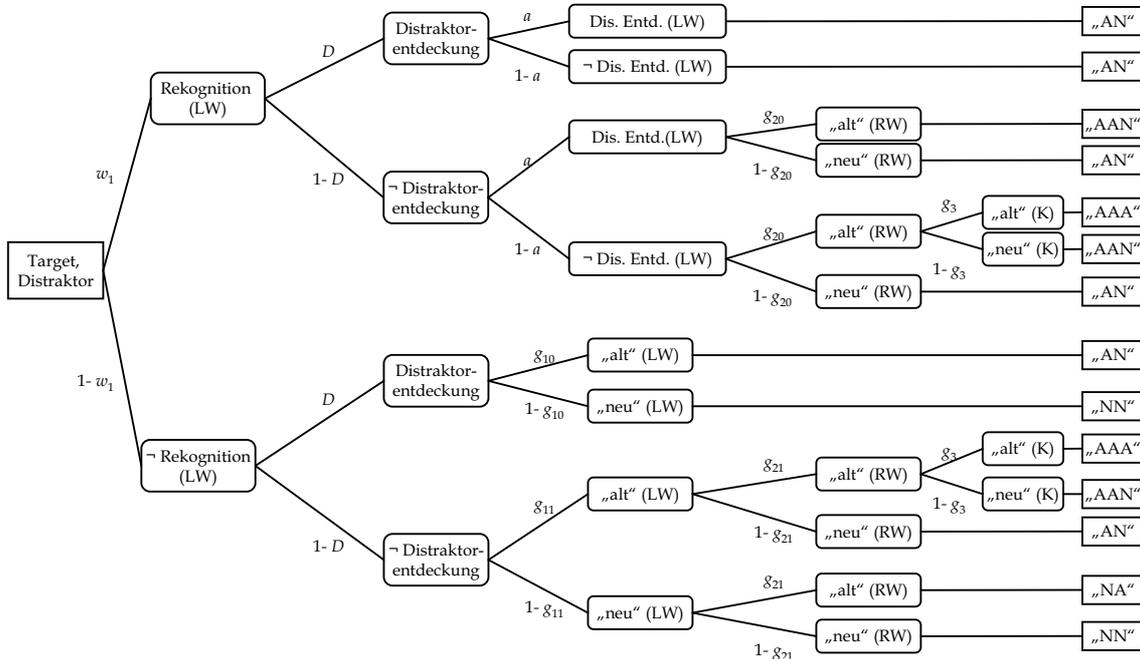


Abbildung 4.9: Diese Grafik zeigt das Verarbeitungsbaummodell für Target-Distraktor Kombinationen.

Tabelle 4.2: Beschreibung der Parameter des ADH-Modells und Erläuterung der Interpretationen

Parameter	Parameterbeschreibung
c	Wahrscheinlichkeit der Enkodierung/Speicherung eines Wortpaares als Assoziation
r	Wahrscheinlichkeit des Abrufs einer Assoziation
w_1	Wahrscheinlichkeit, das linke Item als altes Item zu rekonozieren
w_2	Wahrscheinlichkeit, das rechte Item als altes Item zu erkennen
D	Wahrscheinlichkeit der Distraktorentdeckung eines neuen Wortes
a	Wahrscheinlichkeit der Distraktorentdeckung einer neuen Kombination
g_{10}	Wahrscheinlichkeit, für das linke Wort „alt“ zu raten, gegeben, dass für das rechte Wort kein Rateprozess stattfindet
g_{11}	Wahrscheinlichkeit, für das linke Wort „alt“ zu raten, gegeben, dass für das rechte Wort ebenfalls geraten wird
g_{20}	Wahrscheinlichkeit, für das rechte Wort „alt“ zu raten, gegeben, dass für das linke Wort kein Rateprozess stattfindet
g_{21}	Wahrscheinlichkeit, für das rechte Wort „alt“ zu raten, gegeben, dass für das linke Wort ebenfalls geraten wird
g_3	Wahrscheinlichkeit, für die Kombination der Wörter „alt“ zu raten

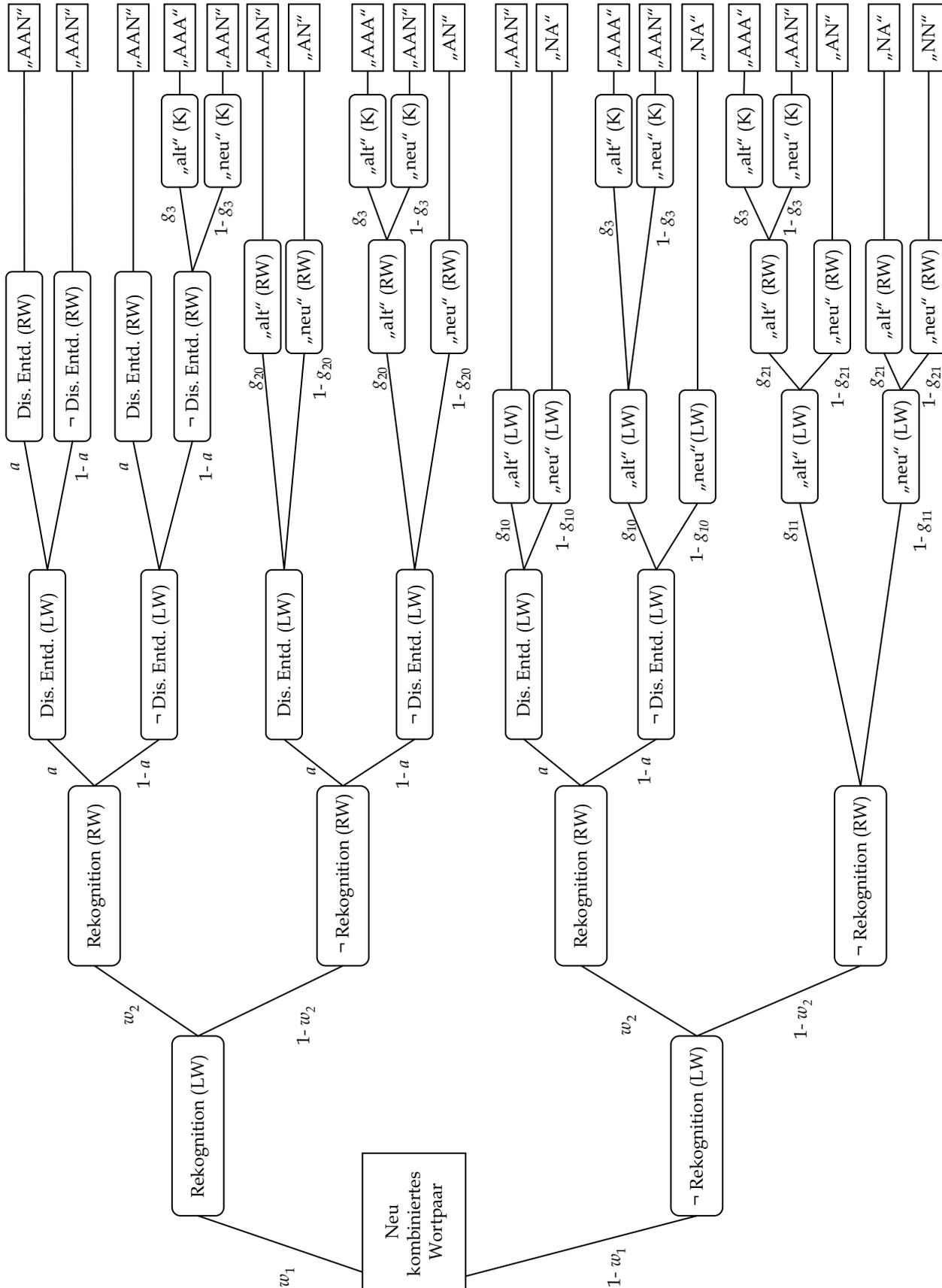


Abbildung 4.11: In dieser Grafik ist das vollständige Verarbeitungsbaummodell für neu kombinierte Wortpaare dargestellt.

unumstrittenen theoretischen Annahmen basieren. Daher ist es wichtig, an dieser Stelle festzuhalten, dass es sich bei dem vorliegenden Modell um ein Messmodell handelt. Das Ziel dieses Modells ist es, das Dekomponierungsproblem des assoziativen Defizits zu lösen und die Schätzung spezifischer kognitiver Prozesse zu ermöglichen. Es wird keinesfalls postuliert, dass die angenommenen Prozesse den realen kognitiven Prozessen genau entsprechen, und sicherlich gibt es auch alternative Beschreibungsmöglichkeiten.

Die Annahmen des 11-Parameter-ADH-Modells können generell in zwei Klassen unterteilt werden: (1) Grundlegende gedächtnispsychologische Annahmen und (2) Annahmen die aufgenommen wurden, um spezifischen Lern- bzw. Testsituationen gerecht zu werden und damit den globalen Einsatz des Modells zu ermöglichen.

In die erste Klasse fällt z.B. die Annahme, dass nach der erfolgreichen Enkodierung des Wortpaares als Assoziation (c), das Wortpaar entweder als Ganzes abgerufen werden kann (r), oder gar nicht. War der Abruf des Wortpaares nicht erfolgreich, so führen nur Rateprozesse zu einer bestimmten Antwortkategorie. Es ist nicht vorgesehen, dass die einzelnen Items einer Assoziation nach gescheitertem Assoziationsabruf zugänglich sind. Diese ‚Alles-oder-nichts‘-Annahme des Wortpaar-Abrufs wurde aus dem Enkodierung-Abruf-Modell übernommen auch die Fragment-Theorie von Jones (Jones, 1984) trifft diese Annahme. Andererseits gibt es empirische Befunde sowie Modelle (Buchler et al., 2008, SAC-Modell, Reder et al., 2000; TODAM, Murdock, 1982) die diese Art der Gedächtnisrepräsentation bzw. des Abrufs nicht unterstützen. Inwieweit es sich um eine für das Messmodell kritische Annahme handelt, sollte anhand der Güte der Modellanpassung sowie dem empirischen Vergleich von alternativen Modellen entschieden werden. Falls ein konkurrierendes Modell keine bessere Datenanpassung aufzeigt und die Parameter des ADH-Modells nicht verzerrt geschätzt werden, kann diese Annahme als unproblematisch eingestuft werden. Dieser Punkt wird nach der empirischen Modellprüfung in der Diskussion (Kapitel 11) noch einmal aufgegriffen.

Eine weitere gedächtnispsychologische Annahme des Modells ist die Konzeption als Zwei-Hochschwellen-Modell. Die in der Psychologie geführte Debatte über Schwellenannahmen soll an dieser Stelle nicht detailliert wiedergegeben werden. Vielmehr wird auf die Arbeit von (Bayen et al., 1996) verwiesen. In diesem Artikel wird der Hintergrund der Diskussion erläutert und es wird, wie im vorangehenden Absatz vorgeschlagen, ein empirischer Vergleich von Quellengedächtnismodellen vorgenommen. Die konkurrierenden Modelle enthalten entweder eine Niedrigschwellen-Annahme, eine Ein-Hochschwellen-Annahme oder eine Zwei-Hochschwellen-Annahme (siehe auch Kapitel 4.2.2). Die Evaluation der Modelle berücksichtigt dabei nicht allein die Güte der Modellanpassung, sondern auch die Effekte gedächtnispsychologischer Manipulationen auf die Modellparameter. Es stellte sich heraus, dass lediglich die Parameter des Zwei-Hochschwellen-Modells als unverzerrte Schätzer der psychologischen Prozesse angesehen werden können. Aus diesem Grund wurde die Zwei-Hochschwellen-Annahme in das ADH-Modell übernommen.

Zu der zweiten Klasse der impliziten Annahmen gehört z.B. die Ausdifferenzierung der Itemrekognitionsparameter aufgrund der Position (w_1 vs. w_2). Diese Annahme wurde vor allem getroffen, um eine möglichst breite Modellanwendung zu ermöglichen. Sobald z.B. unterschiedliche Itemklassen (Gesichter und Namen) genutzt werden, muss die unterschiedliche Erinnerungsleistung an diese beiden Itemklassen auch von unterschiedlichen Parametern aufgefangen werden. Auch bei der Untersuchung von Interferenzeffekten durch sich überlappende Wortpaare in der Lernphase (AB, AC, AD) ist es relevant, die differenzierten Auswirkungen auf das rechts vs. linke Wort zu berücksichtigen. Dies impliziert nicht, dass in einer Standardanwendung, in der das linke und das rechte Wort randomisiert aus einer Menge von Wörtern mit vergleichbaren linguistischen Merkmalen gezogen werden, diese Unterscheidung notwendig ist. Um das Modell aber auf möglichst viele Lern- und Testsituationen anwenden zu können, sollte diese Differenzierung berücksichtigt werden.

Die gerade geführte Argumentation gilt nicht nur für die Positionsunterscheidung der Itemrekognitionsparameter, sondern auch für die Positionsunterscheidung der Rateparameter. Wie unter anderem Glanzer und Adams (1985) sowie Hirshman (1995) zeigen konnten, haben Manipulationen der Gedächtnisstärke auch Auswirkungen auf die Rateprozesse: Für eine Stimulusklasse, die gut erinnert werden kann, werden konservativere Kriterien angelegt, als für eine Stimulusklasse, die eine schlechtere Rekognitionsleistung hervorruft. Um diesem Effekt gerecht zu werden, sollte auch für die Rateparameter eine Positionsdifferenzierung beibehalten werden.

Die Ausdifferenzierung der Rateparameter in Abhängigkeit der vorangegangenen bzw. folgenden Rateprozesse (Index j) ist vor allem dann notwendig, wenn das Verhältnis von alten intakten und neu kombinierten Wortpaaren zu Target-Distraktor- sowie Distraktor-Target-Kombinationen variiert wird. Sobald z.B. ein sehr hoher Anteil an alten intakten Wortpaaren getestet wird (um eine hohe Teststärke für die Parameter c und r zu erreichen), sollte dies zu einer hohen Bereitschaft führen unter Unsicherheit für beide Wörter „alt“ zu raten.

Wie bereits erläutert wurde, ist die Ausdifferenzierung der Parameter vor allem vorgenommen worden, um das Modell möglichst uneingeschränkt einsetzen zu können. Diese Annahmen sollten abhängig von der jeweiligen experimentellen Situation auch restringiert werden können. Eine Veränderung der ‚Alles-oder-nichts‘-Annahme des Wortpaar-Abrufs hingegen oder die Aufhebung der Zwei-Hochschwellen-Annahme würden zu einem alternativen Modell führen. Damit würde sich auch die Frage nach der Identifizierbarkeit, der Modellanpassung und der Parametervalidierung neu stellen.

Im Folgenden wird aufgrund der ausgeführten Argumente an den impliziten Annahmen festgehalten. Grobe Fehlspezifikationen sollten sich in einer schlechten Modellanpassung bemerkbar machen. Auch aus diesem Grund wurde zunächst das Enkodierung-Abruf-Modell sowie die Quellengedächtnismodelle beschrieben. Beide Modelle wiesen über zahlreiche Publikationen hinweg gute Modellanpassungen auf, und zum Teil wurden die

theoretischen Annahmen der Modelle gegenüber konkurrierenden Annahmen empirisch getestet (Bayen et al., 1996). Deshalb lehnt sich das neue Modell im Hinblick auf zentrale theoretische Annahmen an diese Modelle an.

Vergleichbar zu den von Batchelder und Riefer (1990) sowie Bayen et al. (1996) aufgestellten Modellhierarchien für Quellengedächtnismodelle kann auch im Hinblick auf das 11-Parameter-ADH-Modell diskutiert werden, welche Parameterrestriktionen eventuell zu spezifischen Submodellen führen könnten. Die Notwendigkeit zur Parameterrestriktion ist im vorliegenden Fall nicht gegeben, da das Modell identifizierbar und testbar ist. Dennoch sind Modelle mit weniger Parametern natürlich bei einer vergleichbaren Modellanpassung zu bevorzugen, so dass es nahe liegt, Überlegungen der Vereinfachung anzustellen.

Das ADH-Modell als Ausgangsmodell einer Modellhierarchie

Im Folgenden werden spezifische Submodelle des 11-Parameter-ADH-Modells formuliert. Die empirische Überprüfung dieser Modelle wird im folgenden empirischen Teil der Arbeit berichtet.

Die Spezifizierung zweier Itemrekognitionsparameter wurde vor allem damit begründet, dass für unterschiedliche Itemklassen (z.B. Wörter an der linken Position und Bilder an der rechten Position) auch andere Gedächtnisleistungen resultieren könnten. Es ist ein explizites Ziel (siehe Kriterium 4, Kapitel 4.2.3), ein Modell zu formulieren, das offen ist für unterschiedliche Itemklassen und Materialien. Daher sollte in jedem Fall eine Validierung und Spezifizierung der Parameter w_1, w_2 erfolgen. Für die Verwendung von homogenen Itemklassen allerdings ist zu erwarten, dass sich keine starken Positionseffekte ergeben und daher eine Vereinfachung durch die Restriktion $w_1 = w_2$ vorgenommen werden kann. Diese Restriktion würde zu dem Submodell 10 führen (siehe Abbildung 4.12).

Auch die Rateparameter wurden im Hinblick auf die Position spezifiziert ($i = 1, 2$). Mit einer ähnlichen Argumentation wie bei den Itemrekognitionsparametern könnte auch hier die Positionsabhängigkeit aufgegeben werden, d.h. die Restriktion $g_{10} = g_{20}$ sowie $g_{11} = g_{21}$ kann vorgenommen werden. Dies würde zu dem Submodell 9a führen, welches lediglich zwei Prozessparameter für das Raten der Items aufweist.

Genauso wie die Annahme der Positionsunabhängigkeit eingeführt werden kann, gibt es auch Fälle, in denen es sinnvoll erscheint, die Annahme der Prozessabhängigkeit aufzugeben und die Positionsabhängigkeit beizubehalten. Werden zum Beispiel sehr viel mehr Target-Distraktor-Kombinationen als Distraktor-Target-Kombinationen getestet, so sollte dies sicherlich getrennte Itemrateparameter für die Wortposition erfordern. Die Prozessannahme könnte in diesem Fall evtl. aber aufgegeben werden. Dies würde zu dem Modell 9b führen, das lediglich zwei Positionsparameter schätzt.

Auch die Kombination der beiden Restriktionen d.h. die Reduktion auf einen Itemrateparameter durch die Gleichsetzung von $g_{10} = g_{11} = g_{20} = g_{21}$ kann eingeführt werden (siehe Modell 8c).

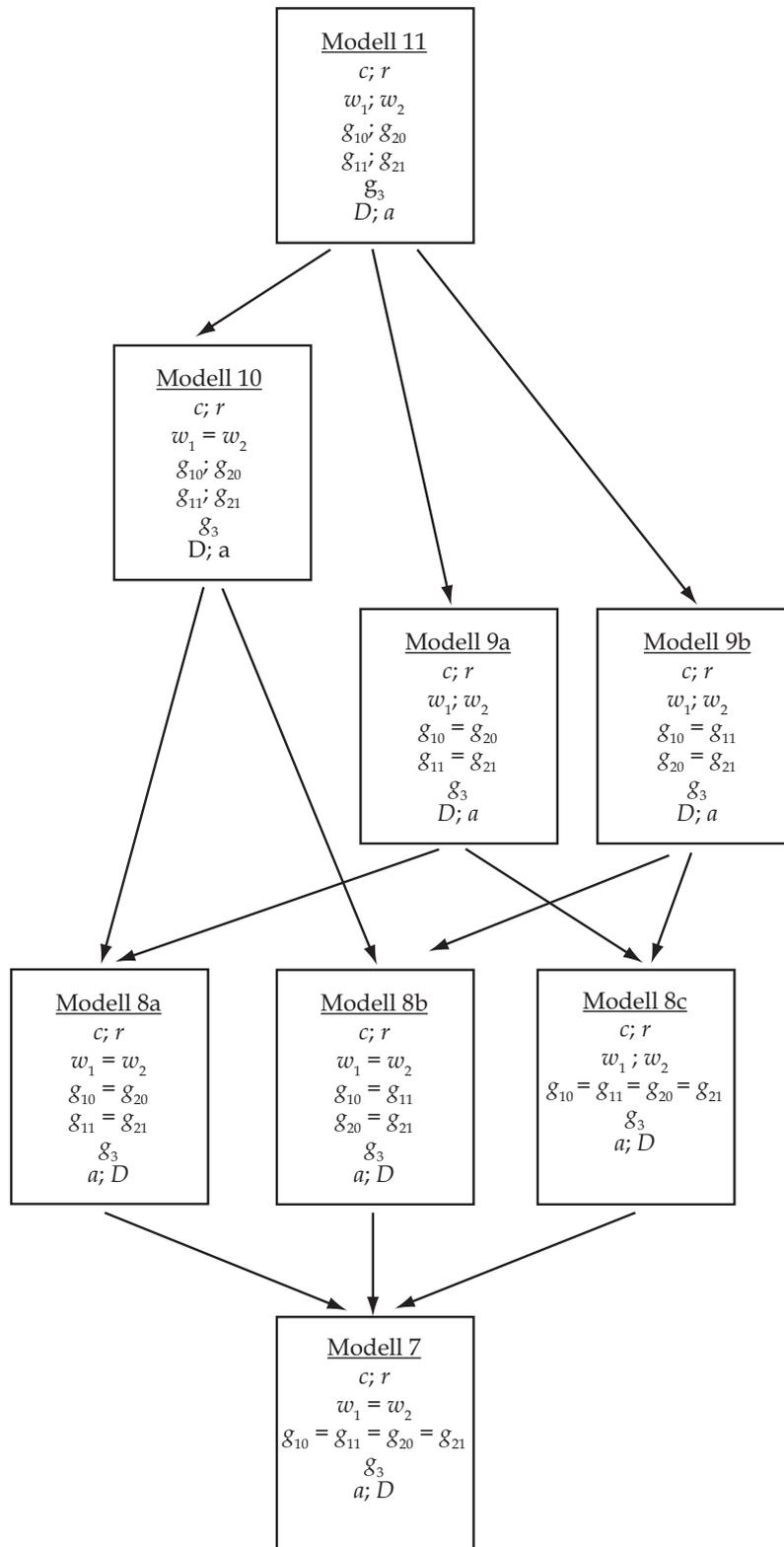


Abbildung 4.12: Diese Abbildung zeigt die 11-Parameter Version des ADH-Modells als Ausgangsmodell einer Modellhierarchie. Es werden spezifischere Submodelle eingeführt, die durch Parameterrestriktionen des Ausgangsmodells erreicht werden können.

Wird das Submodell der zwei Prozessparameter (9a) kombiniert mit der Rekognitionsrestriktion ($w_1 = w_2$), so entsteht dadurch das Modell 8a. Die Kombination der Rekognitionsrestriktion mit dem Submodell zweier Prozessparameter (9b) führt zu Modell 8b. Auch die gleichzeitige Umsetzung des Modells 8c (ein Itemrateparameter) und die Reduktion der Rekognitionsparameter ist möglich. Diese Kombination ist in Modell 7 dargestellt.

Inwieweit diese spezifischen Modelle eine gute Anpassung aufweisen, wird im empirischen Teil der Arbeit berichtet. Da das 11-Parameter-Modell allerdings theoretisch fundierte Annahmen enthält, soll dieses Modell validiert werden. Gelingt die Validierung des globalen Modells, so sind spätere Restriktionen problemlos durchführbar. Allerdings können weitere Parameter zu einem validierten spezifischen Submodell nicht ohne eine erneute Validierung hinzugenommen werden, da mit dieser Veränderung strenggenommen ein neues Modell postuliert wird. Zwar ist es eine größere Herausforderung, das 11-Parameter-Modell zu validieren, es erscheint aber im Hinblick auf eine möglichst breite Modellanwendung sinnvoll, wenn nicht sogar notwendig.

In diesem Kapitel wurde dargestellt, dass es sich bei der Lokalisationsfrage des assoziativen Defizits um ein Dekomponierungsproblem handelt: Anhand der empirisch beobachtbaren Antworten kann nicht geschlussfolgert werden, inwieweit diese auf Enkodierungs- oder Abrufdefizite zurückgehen. Zur Lösung von Dekomponierungsproblemen sind designbasierte Ansätze nur begrenzt nützlich (dies wurde an einzelnen Beispielen verdeutlicht), die adäquate Lösung ist in der mathematischen Modellierung zu sehen. In formalen Modellen können die relevanten kognitiven Prozesse über latente Parameter dekomponiert und geschätzt werden.

Mit dem Ziel, ein multinomiales Modell zur Lösung des Lokalisationsproblems zu entwickeln, wurde zunächst ein neues Paradigma vorgestellt, welches es ermöglicht, genügend unabhängige Datenpunkte zu erfassen, um alle zentralen Prozesse modellieren zu können. Das neue kombinierte Item- und assoziative Rekognitionsparadigma weist darüber hinaus auch einige andere Vorteile auf, die diskutiert wurden.

Die Verarbeitungsbaumstruktur, die Modellparameter und Modellannahmen des neuen ADH-Modells wurden im Kapitel 4.2.3 ausgeführt. Auch wurde eine Modellhierarchie aufgestellt (Kapitel 4.2.3). Diese führt die allgemeine 11-Parameter-Version des Modells als Ausgangsmodell, über Parameterrestriktionen werden spezifischere Submodelle definiert. Im nächsten Kapitel soll nun das allgemeine 11-Parameter-Modell empirisch überprüft werden. Dabei sollen die Identifizierbarkeit und die Modellanpassung evaluiert werden, und die inhaltliche Validierung der einzelnen Parameter wird angestrebt.

Teil II

Empirische Modellvalidierung

Kapitel 5

Methodisches Vorgehen der Modellvalidierung

Im vorangegangenen Teil dieser Arbeit wurde das ADH-Modell vorgestellt, in diesem Teil soll das Modell empirisch überprüft werden. Mehrere Ziele sind hiermit verbunden: (1) Es gilt die Identifizierbarkeit des Modells nachzuweisen, um sicherzustellen, dass eine eindeutige Zuordnung von empirischen Kategoriewahrscheinlichkeiten und Parameterschätzungen gegeben ist. (2) Das Modell sollte als geeignetes Messmodell eine gute Anpassung an die empirischen Daten aufzeigen. Unter der Voraussetzung, dass mindestens eine lokale Identifizierbarkeit vorliegt und eine gute Modellanpassung erreicht wurde, kann man sich den Parametern des Modells genauer zuwenden. Bisher wurden die unterschiedlichen Modellparameter erläutert, und es wurde ihnen eine bestimmte kognitionspsychologische Interpretation zugewiesen. Nun muss gezeigt werden, (3) dass diese inhaltlichen Zuordnungen auch gerechtfertigt sind, die Modellparameter müssen validiert werden.

Sowohl hinsichtlich der Vorgehensweisen als auch im Hinblick auf die Kriterien liegen für alle drei Ziele Standards vor, an denen sich diese Arbeit orientieren wird (für konkrete Bsp. siehe Erdfelder & Buchner, 1998; Klauer & Wegener, 1998; Smith & Bayen, 2004; als Überblick siehe Batchelder & Riefer, 1999; Erdfelder, 2000).

Die Anpassung an die empirischen Daten (1) kann u.a. anhand der Modellanpassungst Statistik G^2 (siehe Kapitel 4.2.1) sowie anhand des Informationsmaßes BIC beurteilt werden (Bayesian Information Criterion, Schwarz, 1978). Beide Statistiken werden von den Computerprogrammen `HmmTree` (Stahl & Klauer, 2007) und `multiTree` (Moshagen, 2009) berechnet. Das Programm `multiTree` bietet auch die Möglichkeit, die lokale Identifizierbarkeit des Modells (2) anhand eines konkreten Datensatzes bzw. anhand von simulierten Daten zu überprüfen. Als Kriterien für die lokale Identifizierbarkeit wird in den folgenden Validierungsexperimenten für jeden Datensatz der Rang der Jacobischen Matrix²² (Bamber

²²Die Jacobische Matrix entspricht der ersten Ableitungen der Modellgleichungen nach den Modellparametern an einer bestimmten Stelle des Parameterraumes

& Santen, 2000) und das Ergebnis der mehrmaligen Aktivierung des EM-Algorithmus mit unterschiedlichen Zufallsstartwerten (siehe 4.2.1) berichtet. Des Weiteren wird in der Zusammenfassung der Validierungsexperimente außerdem auf den Simulationsnachweis der Identifizierbarkeit eingegangen. Um die Parameter zu validieren (3), muss im Sinne der Konstruktvalidität aufgezeigt werden, inwieweit diese für die postulierten kognitiven Prozesse stehen. Dazu müssen experimentelle Manipulationen ausgewählt werden, von denen man aufgrund etablierter Forschungsergebnisse oder allgemein akzeptierter Theorien weiß, dass sie bestimmte kognitionspsychologische Effekte verursachen. Es gilt also für jeden Parameter (bzw. jede Parameterklasse) Manipulationen zu finden, die diesen Parameter gezielt beeinflussen. Prinzipiell ist es dafür nicht notwendig, neue Experimente zu erheben. Es können auch existierende Datensätze zu Validierungszwecken herangezogen werden. Die Einführung eines neuen Paradigmas in dieser Arbeit führt aber dazu, dass leider keinerlei Reanalysen berichtet werden können. So werden also Experimente durchgeführt, die explizit dem Validierungszweck dienen.

Um von einer erfolgreichen Validierung auszugehen, können unterschiedlich strenge Kriterien angelegt werden (Erdfelder, 2000). Man spricht von der Erreichung des schwachen Kriteriums, wenn sich der postulierte Effekt auf den Zielparameter auswirkt. Die Effekte auf andere Parameter werden hierbei ignoriert. Für das starke Kriterium hingegen werden *alle* Parameter des Modells berücksichtigt. Dieses Kriterium ist erfüllt, wenn sich eine Manipulation wie erwünscht im Zielparameter zeigt, darüber hinaus aber die anderen Parameter nicht signifikant beeinflusst. In diesem Kontext werden auch die Konzepte der konvergenten und diskriminanten Validität verwendet: Konvergente Validität wird durch den Effekt auf den Zielparameter sichergestellt, die diskriminante Validität wird durch die Nulleffekte auf die anderen Parameter erreicht (Campbell & Fiske, 1959).

Für die vorliegende Arbeit sollen die strengen Kriterien der Modellvalidierung angewandt werden. In jedem Experiment wird daher erläutert, welche Parameter im Sinne der konvergenten Validität eine Veränderung aufzeigen sollten und welche Parameter unbeeinflusst bleiben sollten, um die diskriminante Validität zu garantieren. Insbesondere im Hinblick auf die Nulleffekte, die häufig zur Kontrolle der diskriminanten Validität postuliert werden, muss sichergestellt werden, dass diese nicht nur eine Folge von mangelnder Teststärke sind. Dieser Aspekt wird in allen Experimenten berücksichtigt.

Im Kontext der folgenden Datenanalysen wird auch auf die unterschiedlichen Submodelle des ADH-Modells eingegangen. Im letzten Kapitel wurde eine Modellhierarchie mit der 11-Parameter-Lösung als Ausgangsmodell dargestellt. Über verschiedene Parameterrestriktionen können die spezifischen Submodelle getestet werden (siehe Kapitel 4.2.3). Anhand der konkreten experimentellen Datensätze soll überprüft werden, inwieweit eine Parameterrestriktion möglich ist, ohne die Modellanpassungsstatistik signifikant zu verschlechtern. Die Validierungsexperimente zielen aber alle auf die Validierung des 11-Parameter-Modells ab.

Tabelle 5.1: Ausblick auf die Ziele und Hypothesen der Validierungsexperimente

	Validierungsexperiment				
	Exp. 1b	Exp. 2		Exp. 3	Exp. 4
Zielparameter	w_1, w_2	c	r	$g_{10}, g_{11},$ g_{20}, g_{21}	$g_{1j},$ g_{2j}
Exp. Manipulation	Präs.zeit kurz/lang	Enkodierung tief/flach	Ret.Intervall 3 min/24 h	Basisrate Testwörter	Aus.matrix lib./kons.
Erw. Effekte					
(a) Konv. Val.	$w_1, w_2,$ c	$c, a,$ w_1, w_2	$r,$ w_1, w_2	$g_{10}, g_{11},$ g_{20}, g_{21}	$g_{1j},$ g_{2j}
(b) Diskr. Val.	$r, g_{ij},$ a, D, g_3	$r, g_{ij},$ D, g_3	$c, g_{ij},$ a, D, g_3	c, r, w_i a, D, g_3	c, r, w_i a, D, g_3

Anmerkung.

Für jedes Experiment sind die Zielparameter der Validierung, die experimentelle Manipulation sowie die erwarteten Effekte im Sinne einer diskriminanten sowie konvergenten Validität genannt.

Bevor in den nächsten Kapiteln auf die Validierungsexperimente im Detail eingegangen wird, soll zur besseren Übersicht zunächst ein kurzer Überblick über die Experimente und deren Ziele gegeben werden.

Aufgrund des neuen Paradigmas stellt sich für das ADH-Modell die Frage nach der Paradigmenunabhängigkeit. In vielen psychologischen Forschungsbereichen können unterschiedliche Testparadigmen verwendet werden, um die Antworten der Teilnehmer zu erfassen. Die die ‚Remember-Know-Prozedur‘ (Gardiner, 1988; Rajaram, 1993; Tulving, 1985) z.B. kann in einem simultanen Testformat durchgeführt werden, d.h. die Antworten „Remember“, „Know“, „New“ werden gleichzeitig nebeneinander dargeboten (‚three-alternative forced-choice test‘; 3AFC-Test). Es ist aber auch möglich, den Test in einem sequentiellen Format durchzuführen. Dies sieht vor, dass die Teilnehmer zunächst eine Alt-neu-Entscheidung treffen müssen und nur nach „Alt“-Urteilen muss in einem zweiten Schritt die Antwort auf die „Remember“- oder „Know“-Frage gegeben werden. Optimalerweise sollte das ADH-Modell auf unterschiedliche Testparadigmen vergleichbar gut anwendbar sein und seine Unabhängigkeit vom Paradigma demonstrieren. Dies kann aber keineswegs als trivial vorausgesetzt werden und soll daher in einem ersten Experiment (Experiment 1a) empirisch untersucht werden.

In Experiment 1b sollen die Itemrekognitionsparameter w_1 und w_2 validiert werden. Dabei sollte sich nicht nur zeigen, dass die Gedächtnisparameter w_1 und w_2 global auf eine Veränderung der Präsentationszeit reagieren, sondern dass sie spezifisch die jeweilige Gedächtnisleistung für das *rechte* respektive *linke* Wort widerspiegeln.

Die zentralen Parameter der assoziativen Enkodierung (c) und des assoziativen Abrufs (r) sollen im zweiten Experiment über ein Vier-Gruppen-Enkodierung-Abruf-Design unabhängig voneinander beeinflusst werden: Eine tiefe assoziative Enkodierung soll über die

Einbettung des Wortpaares in einen Satz gefördert werden, der Vokalvergleich zwischen den beiden Wörtern soll diesen Prozess eher verhindern. Diese beiden Enkodierungsbedingungen sollten nicht nur einen Effekt auf den assoziativen Enkodierungsparameter c haben, auch die Distraktorentdeckung für Assoziationen a sollte in der tiefen Verarbeitungsbedingung erhöht sein. Der Abrufparameter r hingegen sollte auf die Effekte des Testzeitpunkts reagieren, d.h. bei einem sofortigen Test nach drei Minuten sollte der Parameter höher ausfallen als bei einem Test nach 24 Stunden.

In Experiment 3 und 4 ist die Validierung der Rateparameter beabsichtigt. Während in Experiment 3 alle vier Itemrateparameter über den Anteil alter Wörter im Test global beeinflusst werden sollen, wird in Experiment 4 der Frage nachgegangen, ob die Positionsabhängigkeit von Rateparametern für das *rechte* und *linke* Wort gerechtfertigt ist. Dies soll über spezifische Auszahlungsmatrizen erreicht werden.

In Tabelle 5.1 ist für alle vier Experimente aufgezeigt, auf welche Parameter sich die Hypothesen der konvergenten Validität beziehen und welche Parameter im Sinne der diskriminanten Validität unbeeinflusst bleiben sollten. Diese Tabelle wird im Anschluss an die Serie von Validierungsexperimenten in Kapitel 9 erneut dargestellt, mit der Ergänzung welche Hypothesen bestätigt werden konnten und inwieweit die oben genannten Ziele einer guten Modellanpassung, der lokalen Identifizierbarkeit sowie der Parametervalidierung erreicht wurden.

Kapitel 6

Experiment 1

Mit der ersten Datenerhebung dieser Arbeit sind zwei Zielsetzungen verbunden:

(1) Es soll untersucht werden, welches Antwortformat sich am besten für den vorliegenden kombinierten Item- und assoziativen Rekognitionstest eignet (Experiment 1a). Als Alternativen kommen ein sequentieller Ja-nein-Test oder ein einstufiger 5AFC-Test in Frage (‘five-alternative forced-choice test‘; 5AFC-Test). Die Schätzungen der Gedächtnisparameter sollten unabhängig vom Testformat sein und damit die Paradigmenunabhängigkeit dieser Modellparameter bestätigen (Erdfelder, 2000). Bei der Auswahl des Testformats sollten daher Kriterien wie z.B. Verständlichkeit und einfache Handhabung für die Teilnehmer im Vordergrund stehen können.

(2) Das zweite Ziel liegt in der Validierung der Itemparameter (Experiment 1b). Hier gilt es, eine geeignete Manipulation zu finden, die sich gezielt auf die Itemparameter, nicht aber auf andere Parameter auswirkt.

Diese beiden Experimente werden als Experiment 1a und 1b berichtet, da ein Teil der Daten aus Experiment 1b auch in die Auswertung von Experiment 1a einfließt und somit unter zwei unterschiedlichen Gesichtspunkten analysiert wird (siehe Kapitel 6.1.1).

6.1 Experiment 1a

In dieser Arbeit wird mit dem kombinierten Item- und assoziativen Rekognitionstest ein neues Testformat eingeführt. Daher muss zunächst der Frage nachgegangen werden, wie die Antworten der Versuchspersonen in diesem Paradigma erfasst werden können. Für jedes Wortpaar müssen die Teilnehmer drei Urteile abgeben: (1) eine Alt-neu-Entscheidung für das linke Wort, (2) eine Alt-neu-Entscheidung für das rechte Wort und gegebenenfalls (3) eine Alt-neu-Entscheidung für die Wortkombination (die Alt-neu-Entscheidung zur Wortkombination muss nur dann getroffen werden, wenn beide Wörter als alt eingestuft wurden). Es ergeben sich daher mindestens zwei Testmöglichkeiten. Zum einen kann ein sequentieller, dreistufiger Ja-nein-Rekognitionstest durchgeführt werden. In diesem Fall



Abbildung 6.1: Diese Abbildung zeigt die Antwortoptionen des sequentiellen Ja-nein-Formats. Zunächst muss der Teilnehmer beurteilen, ob das linke Wort alt oder neu ist (A), danach muss diese Beurteilung für das rechte Wort abgegeben werden (B). Falls beide Wörter als alt eingestuft wurden, wird auch die Kombination der beiden Wörter getestet (C).

müssen die Teilnehmer eine binäre Ja-nein-Entscheidung zunächst für das linke Wort, dann für das rechte Wort und schließlich für die Kombination abgeben (siehe Abbildung 6.1). Alternativ könnte man auch einen 5AFC-Test durchführen, bei dem die folgenden fünf Antwortoptionen gleichzeitig dargeboten werden: (1) kein Wort ist alt, (2) nur das linke Wort ist alt, (3) nur das rechte Wort ist alt, (4) beide Wörter sind alt, aber neu kombiniert oder (5) beide Wörter sind alt, und es liegt auch eine alte Kombination vor (siehe Abbildung 6.2).

Die Schätzungen der Gedächtnisparameter sollten zwischen den Antwortformaten keine Unterschiede aufweisen, denn die Gedächtnisleistung der Probanden kann nicht von der Antwortfassung abhängen (Erdfelder, 2000; Marsh & Hicks, 1998, siehe nächsten Abschnitt). Es kann aber Gründe geben, eines der beiden Testformate zu präferieren. Ein 5AFC-Test scheint aufgrund der vielen Entscheidungsoptionen z.B. komplexer und insbesondere für ältere Menschen, die weniger Erfahrung mit Testsituationen haben, schwer verständlich. Andererseits können manche Manipulationen wie z.B. eine Manipulation des Zeitdrucks während der Entscheidungsphase nur mit einem 5AFC-Test durchgeführt werden. Auch Reaktionszeitmessungen sind bei einem 5AFC-Test leichter zu interpretieren. Deshalb ist es wichtig, in einem ersten Schritt zu zeigen, dass das Modell nicht an ein

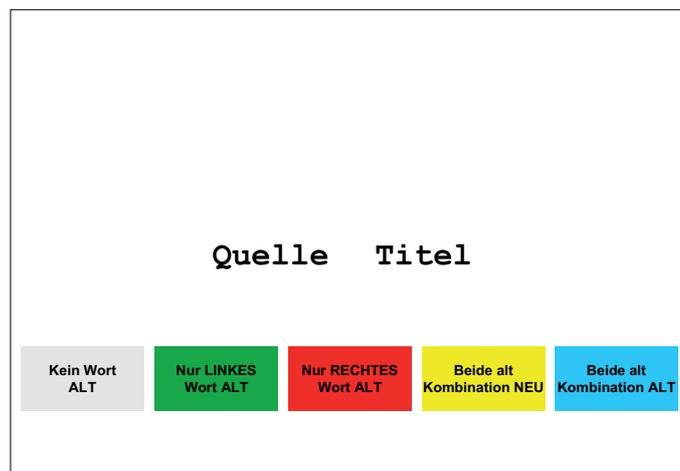


Abbildung 6.2: In dieser Abbildung wird veranschaulicht, wie das 5AFC-Format aussehen könnte. Den Teilnehmern werden alle fünf Antwortmöglichkeiten simultan dargeboten, und ihre Aufgabe ist es, sich für eine Antwortoption zu entscheiden. Sowohl die Anordnung der Antwortoptionen als auch deren Beschreibung sind nur beispielhaft dargestellt.

spezifisches Paradigma gebunden ist und dass beide Antwortformate²³ zu vergleichbaren Parameterschätzungen führen. Dies ist die Zielsetzung von Experiment 1a.

Im Folgenden wird zunächst darauf eingegangen, welche Antwortoptionen in anderen, vergleichbar komplexen Gedächtnisaufgaben Verwendung finden und inwieweit in diesen Paradigmen die Auswirkungen des Antwortformats experimentell untersucht wurden. Danach wird das experimentelle Design und der Ablauf des Experiments 1a erläutert.

Die Bedeutung unterschiedlicher Antwortformate wird in der kognitiven Psychologie vor allem im Kontext der Remember-Know-Prozedur und im Bereich der Quellengedächtnisforschung diskutiert. Im Quellengedächtnisparadigma sind die Antwortoptionen *eindimensionaler* Quellengedächtnisaufgaben weniger komplex als die Optionen *mehrdimensionaler* Verfahren. In eindimensionalen Aufgaben wird nur eine Quellendimension erhoben, z.B. die Position des Stimulus auf dem Bildschirm (oben oder unten). In mehrdimensionalen Quellengedächtnisaufgaben sind mehrere Dimensionen aus der Enkodierungsphase relevant wie z.B. Position und Farbe des Stimulus; d.h. es werden auch mehrere Attribute im Test abgefragt (siehe Kapitel 4.2.2). Zunächst werden an dieser Stelle eindimensionale Aufgaben diskutiert, dann mehrdimensionale. Abschließend wird auf die Befunde aus dem Remember-Know-Paradigma eingegangen.

²³Natürlich sind auch Mischformen dieser beiden Formate denkbar. So könnte zunächst gefragt werden, ob kein Wort, nur ein Wort oder beide Wörter alt sind. In einem zweiten Schritt müsste dann entweder angegeben werden, welches Wort alt ist oder ob beide Wörter auch in dieser Kombination gelernt wurden. Da diese Mischformen aber komplexer sind als der dreistufige Alt-neu-Test und nicht die Vorteile des 5AFC-Tests aufweisen, werden diese alternativen Formate nicht diskutiert.

Eindimensionale Quellengedächtnisaufgaben. Bei eindimensionalen Quellengedächtnisaufgaben muss die Versuchsperson für jedes Item beurteilen, ob es (1) zu Quelle A gehört, (2) Quelle B entstammt oder ob es sich (3) um ein neues Item handelt. Hierbei finden vor allem zwei Aufgabenformate Verwendung: In den meisten Fällen werden die Alternativen im *simultanen Format* präsentiert, d.h. die genannten Alternativen werden in einem 3AFC-Test gleichzeitig dargeboten. Im Gegensatz dazu wird bei der *sequentiellen Variante* zunächst nur das binäre Urteil der Alt-neu-Entscheidung verlangt. Im Falle einer „Alt“-Beurteilung wird dann die zweite Frage nach der Quellenzugehörigkeit (Quelle A oder B) gestellt.²⁴ Diese beiden Antwortalternativen wurden bisher nicht direkt miteinander verglichen.

Kontrastiert wurden 3AFC-Studien und Experimente, bei denen sich bereits die erste Frage des sequentiellen Formats auf eine spezifische Quelle bezieht („Wurde das Item akustisch dargeboten?“). Diese beiden Varianten wurden z.B. im Rahmen des ‚Falschinformationseffekts‘ untersucht. Der Falschinformationseffekt beschreibt das Phänomen, dass die Erinnerung an eine visuell dargebotene Information („Ein rotes Auto hält an einer Ampel“) durch fehlleitende verbale Nachinformationen („Wie schnell war das *blaue* Auto bevor es anhielt?“) beeinflusst wird. Dieser Effekt kann reduziert werden, wenn anstatt eines Rekognitionstests für die visuelle Information („War das Auto rot oder blau?“) ein Quellengedächtnistest verwendet wird („Haben sie ein blaues Auto gesehen oder wurde ein blaues Auto erwähnt?“). In einem solchen Quellengedächtnistest gelingt es den Probanden häufig, die falsche Nachinformation korrekterweise auf die irreführende verbale Frage zurückzuführen. Allerdings erreichten nicht alle Experimente eine vergleichbare Reduktion des Effekts nachträglicher Falschinformationen. Lindsay und Johnson (1989) konnten z.B. eine weitaus stärkere Reduktion des Falschinformationseffekts feststellen, als dies bei Intraub und Hoffman (1992) der Fall war.

Dodson und Johnson (1993) führten diese Divergenz auf die Verwendung von unterschiedlichen Aufgabenformaten zurück. Um diese Hypothese zu überprüfen, manipulierten sie neben anderen Variablen auch das Testformat. Die Ergebnisse konnten die Hypothesen bestätigen, es fand sich ein signifikanter Effekt des Aufgabenformats auf die Häufigkeit der Quellenverwechslung. In der simultanen Version traten signifikant weniger Quellenverwechslungen auf als in der sequentiellen Version. Inwieweit diese Unterschiede allerdings auf eine Veränderung der Antwortkriterien oder auf veränderte Erinnerungsprozesse zurückgeführt werden können, wurde von den Autoren lediglich vermutet, nicht aber weiter überprüft. Aufschluss hierüber kann eine sehr ähnliche Studie von Marsh und Hicks (1998) bzw. die Reanalyse dieser Daten durch Erdfelder (2000) geben.

²⁴In neueren Forschungsansätzen wird auch eine Quellenzuordnung für Items erfragt, die als neu klassifiziert werden (Starns, Hicks, Brown & Martin, 2008). Dies soll an dieser Stelle aber nicht vertieft werden.

Marsh und Hicks (1998) manipulierten, ob eine Quellenentscheidung simultan oder sequentiell abgegeben werden musste. Wie bei Dodson und Johnson (1993) wurde in der ersten Frage des sequentiellen Formats keine Alt-neu-Entscheidung verlangt, sondern es wurde bereits nach einer spezifischen Quelle gefragt. Auch in dieser Studie findet sich ein Einfluss des Aufgabenformats: Das Ja-nein-Frageformat führt generell zu größeren Zustimmungsraten als das simultane Format. Wie Dodson und Johnson (1993) diskutieren auch Marsh und Hicks (1998), ob das Antwortmuster aus einer Verschiebung der Antwortkriterien oder aber aus einer veränderten Gedächtnisleistung resultiert. Eine klare Antwort auf diese Frage wurde durch die Reanalyse von Erdfelder (2000) möglich. Erdfelder (2000) wertete die Daten mit einem multinomialen Modell aus, das die getrennte Schätzung von Rateprozessen und Erinnerungsleistung ermöglicht (Bayen et al., 1996). Die Analyse zeigte einen signifikanten Unterschied in den Ratparametern des Modells, nicht aber in den Quellengedächtnisparametern. Diese Ergebnisse unterstützen also die Hypothese unterschiedlicher Antwortkriterien und sprechen gegen einen alternativen Ansatz intensivierter Gedächtnissuchprozesse.

Die beiden erläuterten Studien sind m.W. die einzigen Studien im Bereich des eindimensionalen Quellengedächtnisses, die sich mit dem Einfluss des Aufgabenformats befassen. In beiden Studien wird aber keine *reine* Manipulation des Aufgabenformats vorgenommen, sondern das Aufgabenformat ist konfundiert mit dem inhaltlichen Fokus der Frage, da in der ersten Frage eine spezifische Quelle in den Vordergrund gerückt wird. Eine reine Manipulation des Antwortformats müsste in der sequentiellen Bedingung zunächst eine Alt-neu-Entscheidung erfragen und dann im Falle einer Alt-Entscheidung die Frage nach den beiden Quellen stellen. Dies berücksichtigen die Autoren auch bei der Interpretation ihrer Ergebnisse. Sie schlussfolgern, dass der veränderte inhaltliche Fokus zu den beobachteten Unterschieden führen könnte und nicht nur das Aufgabenformat an sich.

Zusammenfassend kann festgehalten werden, dass das Antwortformat (bzw. der inhaltliche Fragenfokus) bei eindimensionalen Quellengedächtnisaufgaben keinen Einfluss auf die Gedächtnisleistung hat. Auswirkungen zeigten sich lediglich auf Rateprozesse (Erdfelder, 2000). Der Falschinformationseffekt kann durch das Antwortformat (bzw. den Fragenfokus) moderiert werden, auch dieser Effekt scheint auf einer Veränderung der Antwortkriterien zu beruhen.

Mehrdimensionale Quellengedächtnisaufgaben. Im Gegensatz zu den eindimensionalen sind in mehrdimensionalen Quellengedächtnisaufgaben mehrere Dimensionen der Enkodierungsphase relevant, z.B. Position und Farbe der Stimuli bei der Darbietung. Bei überlagerten Dimensionen (z.B. alle in der oberen Hälfte des Bildschirm gezeigten Wörter sind immer auch in grüner Farbe geschrieben) wird nur eine der beiden Dimensionen getestet (z.B. Ferguson et al., 1992; Geiselman & Crawley, 1983). Dieses Antwortformat ist

daher analog zu eindimensionalen Quellengedächtnisaufgaben. Werden die Dimensionen aber gekreuzt, d.h. sie werden unabhängig voneinander kombiniert (Light & Berger, 1976; Meiser & Bröder, 2002; Schacter et al., 1994), so müssen in der Testphase mindestens vier Quellenentscheidungen getroffen werden. Im *simultanen* Antwortformat resultiert daraus bereits ein 5AFC-Test (4 Quellenkombinationen plus „Neu“-Antwort). Im *sequentiellen* Format hingegen wird zunächst eine Alt-neu-Entscheidung verlangt, und dann werden die Quellendimensionen nacheinander jeweils im 2AFC-Format getestet.

In gekreuzten Quellengedächtnisaufgaben wird vor allem das sequentielle Antwortformat genutzt (von Hecker & Meiser, 2005; Meiser & Bröder, 2002; Meiser & Hewstone, 2004, 2006; Meiser et al., 2007). Light und Berger (1976) sowie Riefer, Chien und Reimer (2007) sind die einzigen Autoren, die einen 5AFC-Test verwendeten. Über die Gründe zur Wahl des Antwortformats werden in der Literatur keine Angaben gemacht. Vermutlich kann diese Präferenz auf die hohe Komplexität der 5AFC-Aufgabe zurückgeführt werden.

Systematisch verglich Meiser (2005) das simultane 4AFC-Antwortformat („four-alternative forced-choice test“; 4AFC-Test) mit dem sequentiellen Format, wobei in beiden Fällen die Alt-neu-Entscheidung vorangestellt wurde (daher auch 4AFC und nicht 5AFC). Die Daten wurden mit einem multinomialen Modell (Meiser & Bröder, 2002; siehe Kapitel 4.2.2) ausgewertet. In diesem experimentellen Vergleich fanden sich weder Effekte auf die Gedächtnis- noch auf die Rateparameter. Der Autor schlussfolgert daher: „Inasmuch as the model selection and the measures of source memory turned out to be independent of response format, experimenters can choose any of the response formats without jeopardizing the empirical results“ (Meiser, 2005, S.15). In nachfolgenden Studien von Meiser und Kollegen (von Hecker & Meiser, 2005; Meiser & Hewstone, 2006; Meiser et al., 2007) fand wieder das sequentielle Format Anwendung. In den Veröffentlichungen wird dies nicht weiter begründet, die Komplexität der Aufgabe scheint aber ein naheliegender, pragmatischer Aspekt zu sein.

Die Studie von Meiser (2005) ist die wichtigste Vergleichsstudie für die vorliegende Arbeit. Zum einen, da der simultane 4AFC-Test eine vergleichbare Komplexität zu dem assoziativen 5AFC-Test aufweist, zum anderen da nur das Antwortformat und keine inhaltlichen Aspekte manipuliert wurden. Meiser konnte keinerlei Auswirkungen des Antwortformats auf die Parameter seines multinomialen Modells finden. Ähnliche Ergebnisse sollten auch im Experiment 1a resultieren. Obwohl das Antwortformat keine Auswirkungen auf die Ergebnisse hatte, findet in der Mehrzahl der mehrdimensionalen Quellengedächtnisstudien das sequentielle Format Anwendung, wahrscheinlich weil es leichter verständlich ist. Diese Schlussfolgerung könnte auch für die vorliegende Arbeit gezogen werden.

Remember-Know-Aufgaben. Bei der Remember-Know-Methode (Gardiner, 1988; Rajaram, 1993; Tulving, 1985) gibt es ebenfalls die Möglichkeit der simultanen und sequentiellen Antwortfassung. Bei der simultanen Version werden die Antwortoptionen „Remember“, „Know“ sowie „Neu“ im 3AFC-Test parallel dargeboten. Bei der traditionell verwen-

deten sequentiellen Erfassung wird zunächst eine Alt-neu-Entscheidung erfragt, und nur für alte Items wird die Frage nach „Remember“ oder „Know“ gestellt.

Hicks und Marsh (1999) haben diese beiden Antwortoptionen verglichen. Sie konnten keine Unterschiede in der Sensitivität der Gedächtnisleistung (d') beobachten. Allerdings variierten die Antwortkriterien in Abhängigkeit des Testformats. Unter Verwendung des simultanen Antwortformats legten die Probanden ein eher liberales Kriterium für „Remember“-Antworten (Kriterium: RC) sowie für „Alt“-Antworten (Kriterium: C) an. Bei der sequentiellen Abfrage war ein eher konservatives Kriterium für „Alt“-Antworten zu verzeichnen. Eine Reanalyse mit anderen Auswertungsmethoden (Erdfelder, 2000; Erdfelder, Cüpper, Auer & Undorf, 2007) konnte den Einfluss der Testbedingung auf die Rateparameter bestätigen. Sowohl die Tendenz, „alt“ zu raten als auch die Tendenz, „Remember“ zu raten, war in der 3AFC-Bedingung stärker ausgeprägt. Es zeigte sich kein Einfluss auf die Erinnerungsleistung.

Hicks und Marsh (1999) führen verschiedene Erklärungsansätze an. Zum einen spekulieren sie, ob Probanden geneigt sind, ihre Antworten gleichmäßig auf die vorhandenen Antwortoptionen zu verteilen. Da in der simultanen Variante zwei „Alt“-Antworten (Remember sowie Know) einer „Neu“-Antwort gegenüberstehen, könnte dies erklären, warum eher die Tendenz besteht, ein Wort als gelernt zu bezeichnen. Allerdings kann dieser Ansatz nur die liberale Antworttendenz für „Alt“-Antworten erklären, nicht jedoch die Verschiebung des Kriteriums der „Remember“-Antworten. Um dies zu begründen, wird überlegt, ob die Einteilung des Vertrautheitskontinuums relativ zu den Antwortkategorien erfolgt. Weiter wird auch angesprochen, ob es sich bei der simultanen Entscheidung eventuell um eine schwierigere Entscheidung handelt und man daher allgemein eher liberale Antwortkriterien anwendet (Hicks & Marsh, 1999).

Die Erklärungsansätze werden von den Autoren nicht weiter empirisch untersucht. Ihre abschließende Empfehlung lautet, das sequentielle Aufgabenformat vorzuziehen. Dieses Format sei näher an der Konzeption der Remember-Know-Abstufung von Gardiner (1988), der ein solches Urteil nur für als alt beurteilte Items vorsieht. Weiterhin wird das sequentielle Verfahren von verschiedenen Autoren als weniger komplex in der Bearbeitung eingestuft (z.B. Eldridge, Sarfatti & Knowlton, 2002) und durch die Gleichverteilung von „Alt/Neu“ bzw. „Remember/Know“ sollten sich keine auffälligen Antworttendenzen ergeben. Auf der anderen Seite gilt hervorzuheben, dass die Effekte des Aufgabenformats sich nicht in der Gedächtnissensitivität niederschlugen, sondern maximal in einer Antwortverzerrung. Unter der Verwendung geeigneter Auswertungsmethoden sollte dieser Effekt also nicht die Gedächtnismaße verfälschen und muss lediglich bei der Interpretation der Antwortkriterien bzw. der Rateparameter berücksichtigt werden. Diese Schlussfolgerung konnte auch aus den Studien des Quellgedächtnisparadigmas abgeleitet werden.

Zusammenfassend ist im Hinblick auf Experiment 1a Folgendes festzuhalten: Die Forschungsergebnisse zur Remember-Know-Prozedur lassen vermuten, dass Versuchspersonen ihre Antwortkriterien an die gegebenen Antwortoptionen anpassen und es daher bei simultanen Formaten mit überwiegenderen „Alt“-Optionen zu einer liberalen Ratetendenz kommen könnte (Hicks & Marsh, 1999). Die Ergebnisse der mehrdimensionalen Quellengedächtnisforschung, deren Testparadigma der vorliegenden Studie sehr ähnlich ist, liefern keine Anhaltspunkte zu dieser Hypothese. Die experimentelle Untersuchung von Meiser (2005) zeigte keinen Unterschied zwischen einem simultanen 4AFC und einem sequentiellen Testformat. Bei mehrdimensionalen Aufgaben zeigte sich, dass ein 5AFC-Format praktisch möglich ist, häufiger aber das weniger komplexe sequentielle Format angewandt wird.

6.1.1 Methode

In Experiment 1a soll empirisch untersucht werden, welche Auswirkungen das Aufgabenformat auf die Parameterschätzungen des Modells hat. Zu diesem Zweck werden ein simultanes und ein sequentielles Aufgabenformat einander gegenübergestellt. In dem simultanen Test stehen fünf Antwortalternativen zur Auswahl: (1) beide Items sind neue Items, (2) nur das linke Item ist alt, (3) nur das rechte Item ist alt, (4) beide Items sind alt und wurden nicht in dieser Kombination gelernt oder (5) beide Items sind alt und wurden in dieser Kombination gelernt. Als sequentielles Aufgabenformat wird ein dreistufiges Aufgabenformat untersucht, in dem zunächst die binäre Alt-neu-Entscheidung für das rechte, dann für das linke Item getroffen werden muss. Wurden beide Items als „alt“ beurteilt, folgt die dritte binäre Entscheidung zur Kombination der Items.

Es ist zu erwarten, dass sich die Gedächtnisparameter nicht zwischen den Bedingungen unterscheiden. Dies würde die Paradigmenunabhängigkeit dieser zentralen Modellparameter zeigen (Erdfelder, 2000). Diese Nullhypothese gilt auch in Bezug auf die Rateparameter. Auswirkungen auf die Rateparameter würden aber kein kritisches Ergebnis darstellen (siehe Hicks & Marsh, 1999). Falls die Probanden wirklich geneigt sind, unter Unsicherheit ihre Antworten gleichmäßig auf die vorliegenden Antwortoptionen zu verteilen, so sollte dies bei der 5AFC-Bedingung eher zu einem konservativeren Rateparameter für Items führen als bei der sequentiellen Ja-nein-Bedingung.

Teilnehmer und Design

An diesem Experiment nahmen 43 Studierende der Universität Mannheim zwischen 18 und 30 Jahren teil ($M = 22.58$; $SD = 2.39$). Die Mehrzahl der Probanden studierte Psychologie (65%). Ihnen wurden Versuchspersonenstunden über die Dauer von 45 Minuten ausgestellt. Es wurden maximal vier Versuchspersonen gleichzeitig erhoben.

Als unabhängige Variable (UV) wurde das Aufgabenformat (5AFC- vs. Ja-nein-Format) innerhalb der Teilnehmer variiert. Die Probanden wurden zu Beginn des Experiments ran-

dominiert einer der beiden Testbedingungen zugeteilt und in dieser einen Bedingung erhoben.²⁵ Sie wurden nach Beendigung des Experiments gefragt, ob Sie Interesse hätten, an einem zweiten Experiment teilzunehmen, das mindestens eine Woche nach der ersten Testung stattfinden würde. An diesem zweiten Termin wurden die Personen in der jeweils anderen Testbedingung untersucht. Für die beiden Sitzungen wurde unterschiedliches Wortmaterial verwendet. Da die Teilnahme an dem zweiten Termin nicht verpflichtend war, sind nicht alle Teilnehmer wieder erschienen. Dieser Ausfall von Teilnehmern hatte zur Folge, dass keine perfekte Ausbalancierung von Erstteilnehmern an der Ja-nein-Bedingung und Erstteilnehmern an der 5AFC-Bedingung erreicht wurde. Es wurden nur die Daten der Teilnehmer ausgewertet, die an beiden Terminen teilnahmen und damit beide Testbedingungen durchliefen. Dies trifft für 43 Personen zu, davon wurden am ersten Termin 24 in der Ja-nein-Bedingung getestet, 19 in der 5AFC-Bedingung. Auswirkungen dieser ungleichen Verteilung von Erstteilnehmern sind nicht zu erwarten, denn die Lernbedingung erfolgte intentional, und ein Übungsdurchgang vor der ersten Lernphase klärte die Teilnehmer bereits über den gesamten Ablauf des Experiments auf.

Ablauf

Das gesamte Experiment von circa 45 Minuten wurde computergesteuert durchgeführt. Die Versuchssteuerung wurde mit dem Programm E-Prime erstellt (Psychological Software Tools, 2000), die Materialien an einem Intel-Pentium(4)-Rechner mit einer Monitorgröße von 19 Zoll dargeboten. Nach Unterzeichnung der Einverständniserklärung wurden die Teilnehmer instruiert, sich Wortpaare für einen späteren Test einzuprägen. Sie wurden darauf hingewiesen, dass es wichtig ist, sich sowohl die einzelnen Wörter zu merken, als auch die Kombination, in der die Wörter gezeigt werden, da auch diese getestet wird. Zunächst folgte dann ein Übungsdurchgang, in dem fünf Wortpaare für je 4 Sekunden präsentiert wurden. Die Wortpaare wurden in der Schriftart Courier New Größe 35 dargeboten. Das rechte Wort war jeweils linkszentriert, das linke Wort rechtszentriert, so dass der Abstand zwischen den Wörtern unabhängig von der Buchstabenanzahl konstant war. Abbildung 6.3 zeigt exemplarisch ein Wortpaar aus der Lernphase. Nach jeder Wortpaardarbietung folgte ein Interstimulusintervall von 500 Millisekunden. An die Präsentation der Übungssitems schloss sich direkt die Testphase für die Übungssitems an. Zu Beginn dieser Testphase wurde das Aufgabenformat erläutert.

²⁵Die Teilnehmer, die in der Ja-nein-Bedingung erhoben wurden, haben zugleich das Experiment 1b bearbeitet. In Experiment 1b wurde eine Manipulation der Darbietungszeit innerhalb der Probanden vorgenommen, daher fließen in die Auswertung von Experiment 1a nur die Daten der Kontrollbedingung ein. Dies sind 50% der Wortpaare. In die 5AFC-Bedingung flossen genauso viele Datenpunkte ein, da aufgrund einer weiteren Variablen, die zwischen den Teilnehmern manipuliert wurde, nur 50% der Probanden mit der Ja-nein-Bedingung vergleichbar sind und ausgewertet wurden. Diese Datenreduktion ist nicht mit der zufälligen Zuteilung zu der Ja-nein- und 5AFC-Bedingung konfundiert und gefährdet diese daher nicht.

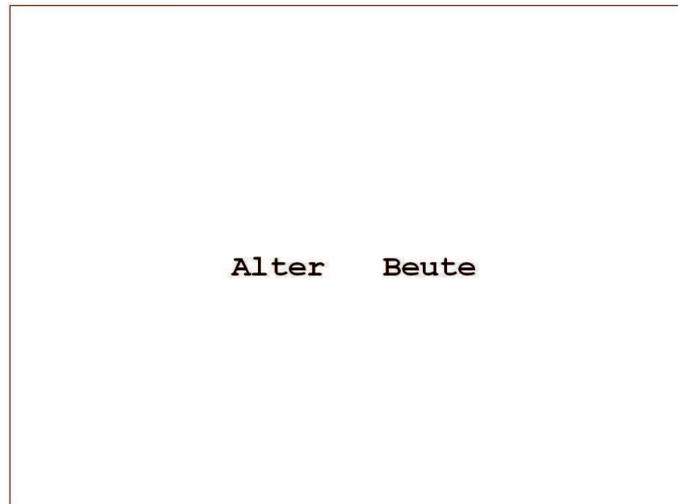


Abbildung 6.3: Abgebildet ist hier ein Ausschnitt aus der Lernphase. Jeweils zwei Wörter wurden zufällig zu einem Wortpaar kombiniert. Das linke Wort wurde rechtszentriert, das rechte Wort linkszentriert dargeboten, dadurch konnte der Abstand zwischen den Wörtern für alle Wortpaare konstant gehalten werden.

In der *sequentiellen Ja-Nein-Bedingung* wurde den Teilnehmer erklärt, dass sie im Folgenden wieder Wortpaare sehen würden und beurteilen sollen, ob die einzelnen Wörter und ggf. deren Kombinationen alt oder neu sind. Sie sollten dabei zunächst angeben, ob das linke Wort ein altes, d.h. gelerntes Wort ist oder ein neues Wort, das nicht gezeigt wurde. Dieselbe Einschätzung sollte danach für das rechte Wort getroffen werden. Für die Kombinationsfrage sollten die Teilnehmer „alt“ antworten, wenn beide Wörter in genau dieser Kombination gelernt wurden, eine „Neu“-Antwort wäre zu geben, wenn zwar beide Wörter gelernt wurden, aber nicht in dieser Zusammenstellung.

Nach diesem Hinweis auf die Antwortmöglichkeiten erschienen dann zunächst unter dem linken Wort die Felder „linkes Wort alt“ und „linkes Wort neu“. Die Beschriftungen waren farblich hinterlegt und über korrespondierende Tasten konnte die Versuchsperson ihre Eingabe vornehmen. Sobald eine Taste gedrückt wurde, verschwanden die Antwortoptionen unter dem linken Wort und es erschienen die Antwortmöglichkeiten unter dem rechten Wort. In Abbildung 6.4 A,B ist das Testformat für das linke und rechte Wort dargestellt. Falls die Versuchspersonen beide Wörter als alte, also gelernte Wörter beurteilten, wurden sie auch gefragt, ob die Wortkombination, in der die Wörter gezeigt wurden, alt ist, d.h. ob die Wörter in genau dieser Kombination gelernt wurden, oder ob die beiden Wörter zu einer neuen Kombination zusammengestellt wurden (siehe Abbildung 6.4 C). Bevor das nächste Testpaar erschien, wurde den Teilnehmern im Übungsdurchgang Rückmeldung über ihre Leistung gegeben. Um die Leistungsrückmeldung vergleichbar zur 5AFC-Antwortbedingung zu gestalten, wurden nur solche Antworten als korrekt beurteilt, bei denen alle gegebenen Antworten richtig waren. Falls also das linke Wort korrekt und

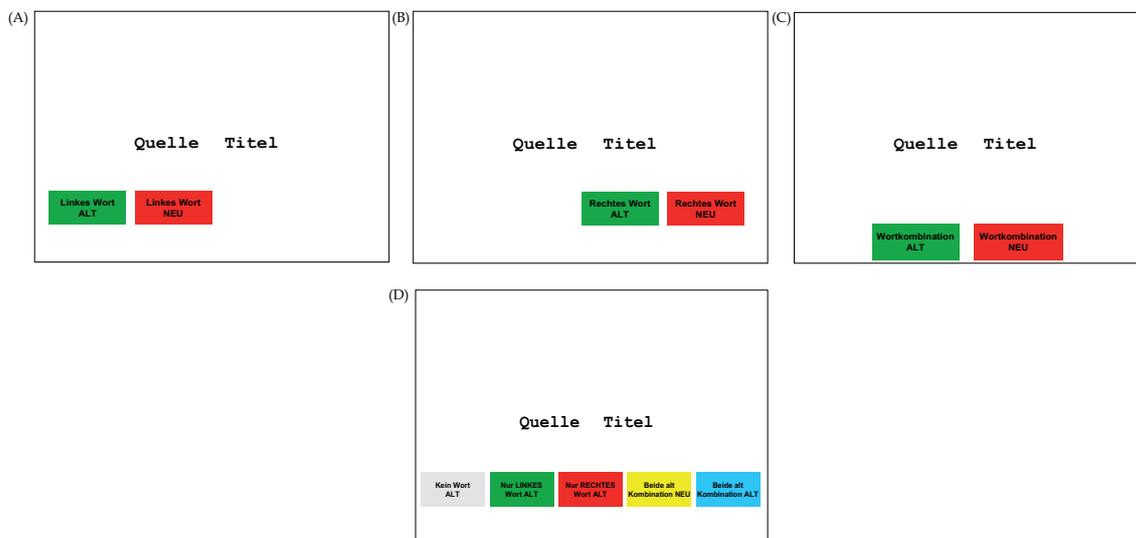


Abbildung 6.4: In dieser Abbildung wird das sequentielle Ja-nein-Format (A-C), dem simultanen 5AFC-Format (D) gegenübergestellt.

das rechte Wort falsch beurteilt wurde, bekam der Teilnehmer die Rückmeldung, dass die Antwort falsch war.

Die farblichen Markierungen für die „Alt“- und „Neu“-Tasten waren für rechtes und linkes Item sowie für die Kombinationsfrage identisch. Zwischen den Versuchspersonen wurde die Tastenbelegung (alt/neu) ausbalanciert, um eine möglichen Antwortverzerrung durch Händigkeitspräferenzen auszugleichen.

Auch in der *5AFC-Bedingung* wurde den Teilnehmern nach der Präsentation der Übungsitems erläutert, dass sie nun Wortpaare beurteilen sollen. Sie wurden darauf hingewiesen, dass entweder nur das linke Wort „alt“, d.h. gelernt sein kann, nur das rechte Wort „alt“ sein kann oder aber beide Wörter „alt“ sein können. Falls beide Wörter „alt“ sind, besteht entweder die Möglichkeit, dass sie auch in dieser Kombination gelernt wurden, oder aber, dass sie zu dieser Kombination neu zusammengestellt wurden. Ihnen wurden dann die zur Verfügung stehenden Antwortoptionen erläutert (siehe Abbildung 6.4 D). Den Teilnehmern wurden fünf Antwortmöglichkeiten vorgegeben. Diese waren beschriftet mit (1) „kein Wort alt“, (2) „nur linkes Wort alt“, (3) „nur rechtes Wort alt“, (4) „beide alt Kombination neu“ sowie (5) „beide alt Kombination alt“. Über farblich zu den Antwortoptionen korrespondierende Tasten gaben die Teilnehmer ihr Urteil ab. Die Antwortalternativen wurden auch hier zum Ausgleich von Händigkeitspräferenzen permutiert.²⁶ Während der Übungsdurchgänge erhielten die Versuchspersonen nach jeder Beurteilung Rückmeldung über ihre Leistung und Gelegenheit, dem Versuchsleiter weitere Fragen zu stellen.

²⁶Um eine sinnvolle Anordnung der Antwortoptionen zu gewährleisten, wurden zwei unterschiedliche Varianten ausbalanciert. Zum einen die in Abbildung 6.4 dargestellte, zum anderen eine gespiegelte Abfolge, bei der ganz links die Antwortoption „Beide alt, Kombination alt“ zu sehen war und ganz rechts die Option „kein Wort alt“.

Nach erfolgreichem Abschluss der Übungsphase folgte dann die eigentliche Lernphase, die für beide Gruppen identisch war. Fünfzig Wortpaare wurden für je 4 Sekunden präsentiert, 3 Füll-Wortpaare, die zur Aufhebung der ‚Primacy- und Recency‘-Positionseffekte dienten, waren jeweils am Anfang und Ende der Lernliste zusätzlich eingefügt. Die 50 Wortpaare wurden zufällig aus einer Liste von 280 Wörtern zusammengestellt. Nach einer Distraktoraufgabe von einer Minute folgte dann die Testphase. Hier sollten 70 Wortpaare beurteilt werden. Davon waren 20 Wortpaare in alter intakter Kombination, 20 Wortpaare neu kombiniert aus zwei alten Wörtern, 10 Target-Distraktor-Kombinationen, 10 Distraktor-Target-Kombinationen und 10 Wortpaare bestehend aus zwei Distraktoren. Die verschiedenen Wortkombinationen wurden jeweils zufällig zusammengestellt und in zufälliger Reihenfolge über die Testliste verteilt. Es wurden mehr alte intakte und neu kombinierte Paare getestet, da die zentralen Parameter c und r lediglich im ersten Baum des Modells vorkommen und sichergestellt werden sollte, dass ausreichend Datenpunkte zur reliablen Schätzung dieser Parameter vorhanden sind (für eine genauere Erläuterung dieses Vorgehens vergleiche auch Erdfelder, 2000).

Die Abfrage der Testwörter erfolgte genau wie in der jeweiligen Übungsphase beschrieben. Es wurde nun aber kein Feedback über die Beurteilungen gegeben. Nach Beendigung der Testphase musste eine Zwischenaufgabe bearbeitet werden, die circa 3 Minuten in Anspruch nahm. Dann erfolgte ein neuer Lern- und Testdurchgang unter Verwendung der verbleibenden 140 Wörter. Insgesamt wurden also 2 Lernlisten mit je 50 Wortpaaren (100 Einzelwörter) gelernt und zwei Testlisten mit je 70 Wortpaaren (140 Einzelwörter) getestet. Am Ende der Experiments erhielten die Probanden Rückmeldung über ihre Gesamtleistung in beiden Durchgängen.

Material

Das Kriterium zur Auswahl des Wortmaterials war die Vergleichbarkeit zu anderen Experimenten, die ebenfalls das Assoziationsdefizit untersuchten. In der Studie von Castel und Craik (2003) bestand das Wortmaterial aus zweisilbigen, konkreten, englischen Substantiven. Naveh-Benjamin (z.B. Naveh-Benjamin, 2000; Naveh-Benjamin, Brav & Levy, 2007) verwendete in seinen Experimenten entweder im Sprachgebrauch häufig vorkommende, zwei- bis dreisilbige hebräische Substantive (Balgur, 1968) oder hochfrequente, drei- bis vier buchstabige englische Substantive (Francis & Kucera, 1982). In keinem Experiment wurden dabei konkrete Angaben über die Häufigkeiten der Wörter gemacht.

In der vorliegenden Studie sollte daher das Wortmaterial ebenfalls aus Substantiven bestehen, die eine häufige Verwendung in der deutschen Sprache und eine ähnliche Silben-Buchstabenanzahl aufweisen wie die genannten Studien. Deutschsprachiges Wortmaterial kann der Celex-Datenbank (Max Planck Institute for Psycholinguistics, 1991) entnommen werden. Hier sind Häufigkeitsangaben aus geschriebenen und gesprochenen Quellen für 51728 Wortlemma aufgelistet. Zusätzlich zu Häufigkeitsangaben führen einige Normie-

Tabelle 6.1: Kennwerte der beiden parallelisierten Wortlisten

	Liste 1		Liste 2	
	Statistik	Standard- abweichung	Statistik	Standard- abweichung
Mittlere Häufigkeit	70.33	70.21	71.84	71.38
Minimale Häufigkeit	6.00		5.00	
Maximale Häufigkeit	498.00		451.00	
Mittlere Buchstabenanzahl	5.85	0.79	5.86	0.80
Minimale Buchstabenanzahl	4.00		4.00	
Maximale Buchstabenanzahl	7.00		7.00	
Mittlere Silbenanzahl	0.72	0.72	0.72	0.72
Schiefe	2.63		2.52	
Exzess	8.71		7.73	

Anmerkung.

Für die zwei parallelisierten Listen von je 280 Wörtern sind Mittelwert und Streuung der Verwendungshäufigkeit, der Buchstabenanzahl und der Silbenanzahl aufgeführt. Die Verwendungshäufigkeit in der deutschen Sprache wurde anhand des Maßes ‚MannMln‘ der Celex Datenbank berechnet und bezieht sich auf die Vorkommenshäufigkeit des Wortes in einem Korpus von einer Million Wörtern. Als Verteilungskennwerte sind außerdem Schiefe und Exzess angegeben.

rungsstudien (Hager & Hasselhorn, 1994) u.a. auch Angaben über Bildhaftigkeit, Konkretheit, Valenz, Potenz und Erregung. Da für die vorliegende Studie ein relativ großer Itempool von circa 600 Substantiven erstellt werden sollte, konnte nicht auf die Hager-Hasselhorn-Normen zurückgegriffen werden, da diese Normierungsstudien dafür keine ausreichende Itemmenge bereitstellen. Werte für die Konkretheit werden aus diesem Grund nicht berichtet.

Für die Auswahl des Wortmaterials wurde daher die Celex-Datenbank (Max Planck Institute for Psycholinguistics, 1991) zugrunde gelegt. Als Häufigkeitsmaß wird der Kennwert ‚MannMln‘ verwendet. Diesem Wert liegt die Vorkommenshäufigkeit eines Wortes im Wortkorpus von 6 Millionen Wörtern zugrunde. Um die Werte besser interpretieren, und mit Frequenzwerten anderer Studien vergleichen zu können, wurde das Maß dann auf die Häufigkeit innerhalb einer Million Wörter umgerechnet. Für die in der Datenbank aufgelisteten 51728 Wortlemma liegt eine mittlere Häufigkeit von 16.26 vor.

Um die Wortlisten für diese Studie zu erstellen, wurden zunächst nur Substantive mit mindestens vier und maximal sieben Buchstaben selektiert. Nach dem Ausschluss von Eigennamen, Namen von Ländern, Städten, Flüssen sowie der Bezeichnungen von Monaten und Wochentagen wurden 900 Wörter aus dem oberen Frequenzbereich ausgewählt. Diese Liste wurde von vier Personen gründlich nach weiteren Kriterien überprüft. Hierbei wurden zusammengesetzte Substantive (z.B. Haustür, Torwart), Fremdwörter (z.B. Omnibus, Minimum), Wörter die mehrere Bedeutungen aufweisen (z.B. Graben, graben), phonologisch nicht differenzierbare Wörter (z.B. Saite, Seite), genus-spezifische bzw. genus-untypische Nomen (z.B. Koch/Köchin, Kaiserin) und Zahlwörter entfernt. Für die verbleibenden Sub-

stantive wurde dann die Silbenanzahl ergänzt, da diese Information nicht in der Datenbank vorhanden ist. Diese Wortauswahl wurde in zwei Listen aufgeteilt, die hinsichtlich der Buchstabenanzahl, Silbenanzahl, Anzahl von Wörtern mit „ß“, Anzahl von Wörtern mit Umlauten, und dem wichtigsten Kriterium, der Frequenz, parallel sind. Nach diesem Vorgehen resultierten zwei Wortlisten, die es ermöglichen sollten, dieselben Versuchspersonen auch in einem relativ kurzen zeitlichen Abstand mit unterschiedlichem aber vergleichbarem Wortmaterial zu testen.

Beide Wortlisten umfassen je 280 Wörter, d.h. je 140 Wortpaare können daraus gebildet werden. Es sind jeweils 23 Wörter mit Umlauten und je 2 Wörter mit dem Buchstaben „ß“ enthalten. Als Verteilungskennwerte der Vorkommenshäufigkeit wurden neben Mittelwert und Streuung auch Schiefe und Exzess berechnet. Die Schiefe kennzeichnet mit 2.63 (Liste 1) und 2.52 (Liste 2) jeweils linkssteile Verteilungen, der Exzess ist positiv mit den Kennwerten 8.71 für die erste und 7.73 für die zweite Liste. Tabelle 6.1 gibt die Mittelwerte und Streuungen der wichtigsten Kennwerte an. Das gesamte Wortmaterial ist mit dem zentralen Häufigkeitskennwert in Anhang C angegeben. Für das erste Experiment (Experiment 1a und 1b) wurden beide parallelisierten Wortlisten 1 und 2 verwendet.

6.1.2 Ergebnisse

Die Daten dieses Experiments werden mit Hilfe des multinomialen ADH-Modells ausgewertet. In die multinomiale Modellanalyse gehen die absoluten Häufigkeiten aggregiert über alle Probanden ein. Pro experimentelle Gruppe wurden zunächst die Häufigkeiten der 25 Modellkategorien berechnet (5 Itemarten · 5 Antwortalternativen). Diese sollen modellunabhängig im Folgenden als deskriptives Maß berichtet werden. Da während der Testphase manche Wortkombinationen (z.B. alte intakte Wortpaare) häufiger getestet wurden als andere, können die absoluten Häufigkeiten nur schwer interpretiert werden. Daher werden relative Häufigkeiten dargestellt (siehe Tabelle 6.2). Die absoluten Häufigkeiten können Anhang F entnommen werden.

Aus den relativen Häufigkeiten (Tabelle 6.2) kann bereits abgelesen werden, dass die Teilnehmer einen hohen Anteil der Items richtig klassifizieren konnten (die richtigen Klassifikationen sind in der Diagonalen kursiv hervorgehoben). Für alte intakte Wortpaare in der 5AFC-Bedingung lag dieser Anteil bei .746, bei neu kombinierten Paaren war die korrekte Klassifikation mit .511 etwas niedriger und vergleichbar mit der Rate richtiger Antworten auf Zielitem-Distraktor Paare (.523).²⁷ Inwieweit diese Häufigkeiten korrekte Erinnerungsleistungen widerspiegeln oder auf Rateprozesse zurückgehen, soll anhand der

²⁷Die Daten der Ja-nein-Bedingung wurden analog zu der 5AFC-Bedingung ausgewertet, d.h. als korrekt wurden nur die Antworten gewertet, bei denen alle drei Entscheidungen richtig waren. Auf eine Target-Distraktor Kombination wurde also lediglich die Antwort „linkes Wort alt, rechtes Wort neu, Kombination neu“ als richtig angesehen. In Bezug auf das alte, linke Wort wäre aber auch die Antwort „beide Wörter alt, neue Kombination“ eine richtige Entscheidung. In der multinomialen Analyse wird dieser Aspekt natürlich berücksichtigt.

Tabelle 6.2: Relative Anwothhäufigkeiten der Ja-nein-Bedingung (Tabelle a) und der 5AFC-Bedingung (Tabelle b)

(a) Ja-nein-Bedingung

Wahre Itemkat.	Antwort der Versuchsperson				
	Altes WP	Neu komb. WP	Target Dis.	Dis. Target	Dis.Dis.
Altes WP	.718	.093	.065	.073	.051
Neu komb. WP	.105	.507	.172	.165	.050
Target Dis.	.040	.181	.526	.071	.181
Dis. Target	.045	.179	.090	.536	.150
Dis. Dis.	.024	.088	.225	.219	.444

(b) 5AFC-Bedingung

Wahre Itemkat.	Antwort der Versuchsperson				
	Altes WP	Neu komb. WP	Target Dis.	Dis. Target	Dis.Dis.
Altes WP	.746	.099	.060	.049	.046
Neu komb. WP	.121	.511	.160	.144	.064
Target Dis.	.045	.243	.523	.045	.145
Dis. Target	.045	.255	.088	.433	.180
Dis. Dis.	.033	.125	.188	.160	.495

Anmerkung.

Für beide Tabellen wurden relative Häufigkeiten berechnet, um eine bessere Vergleichbarkeit zwischen den Itemkategorien zu ermöglichen, die unterschiedlich häufig getestet wurden. Alte intakte Wortpaare sowie neu kombinierte Wortpaare kamen 20 mal im Test vor, die anderen drei Kombinationen jeweils 10 mal. Kursiv sind die korrekten Antworten hervorgehoben. Die Häufigkeiten sind über beide Testlisten aggregiert

multinomialen Analyse geklärt werden, da mit dieser Auswertungsmethode nicht nur die Unterscheidung von Gedächtnis- und Rateprozessen möglich ist, sondern des Weiteren die assoziativen Gedächtnisprozesse in Enkodierungs- und Abrufprozesse ausdifferenziert werden. Auf eine traditionelle Analyse der Differenz von Treffern und falschen Alarmen wird verzichtet, da diese weniger aussagekräftig ist als die multinomiale Analyse.

Multinomiale Modellauswertung

Der multinomialen Analyse liegen die absoluten Häufigkeiten (Anhang F) zugrunde. Die Parameterschätzungen wurden mit Hilfe der Programme AppleTree (Rothkegel, 1999), HmmTree (Stahl & Klauer, 2007) sowie multiTree vorgenommen (Moshagen, 2009). Die zur Auswertung erforderlichen Dateien können Anhang B sowie F entnommen werden.

Die G^2 -Statistik der Modellanpassung ist asymptotisch χ^2 -verteilt, daher wird als kritische Prüfstatistik ein χ^2 -Wert herangezogen. Da mit sehr großen Datenmengen die Wahrscheinlichkeit einer signifikanten Modellabweichung steigt, soll in den folgenden Analysen ein α -Level von $\alpha = .001$ angesetzt werden, um zu vermeiden, dass bereits minimale Modellabweichungen zu einer Ablehnung des Modells führen. Tatsächlich liegt in diesem Experiment eine große Datenmenge vor ($n = 5738$), die dazu führt, dass die Power, eine kleine Modellabweichung ($w = .01$) zu entdecken, selbst bei einem α -Niveau von $.001$ noch bei $1 - \beta > .99$ liegt und als sehr hoch beurteilt werden kann. Für die Parametertests wird allerdings das konventionelle α -Niveau von $.05$ verwendet, da diese in der Regel weniger teststark sind (für ein vergleichbares Vorgehen siehe Bayen, Erdfelder, Bearden & Lozito, 2006).²⁸

Für die folgenden Analysen liegen pro Gruppe 20 (d.h. insgesamt 40) freie Kategorie-wahrscheinlichkeiten vor. Insgesamt werden 22 Parameter geschätzt, daher resultieren für die Auswertung dieses Experiments 18 Freiheitsgrade. Die Modellanpassungsstatistik ist mit $G^2_{(18)} = 36.19$ unter dem kritischen Wert, $G^2_{krit,\alpha=.001} = 42.31$, kann also als befriedigend angesehen werden. Mit den vorliegenden Daten wurde der Rang der Jacobischen Matrix berechnet. Dieser entsprach mit einem Wert von 22 genau der Anzahl der geschätzten Parameter, d.h. eine lokale Identifizierbarkeit ist nicht ausgeschlossen. Daher wurde der EM-Algorithmus mehrmals mit zufälligen Startwerten aktiviert. Da die Parameterschätzungen bei identischer G^2 -Statistik nicht, bzw. nur in einem durch Rundungsfehler unkritischen Bereich, voneinander abwichen, kann von einer lokalen Identifizierbarkeit ausgegangen werden.

²⁸Eine A-Priori-Power-Analyse für die Parametertests kann in diesem Experiment schwer vorgenommen werden, da vor der Datenerhebung keine Anhaltspunkte darüber vorlagen, wie die Parameterschätzungen ausfallen würden. Die Power, einen Effekt z.B. im Abrufparameter r zu entdecken, ist aber abhängig von den anderen Modellparametern z.B. dem vorausgehenden Enkodierungsparameter c . Natürlich könnte man für unterschiedliche Kombinationen der Parameterwerte Daten simulieren und darüber eine Abschätzung der Power ermöglichen, in dieser Arbeit wurde darauf verzichtet, und anstatt einer A-Priori- wurde eine Post-Hoc-Power-Analyse durchgeführt.

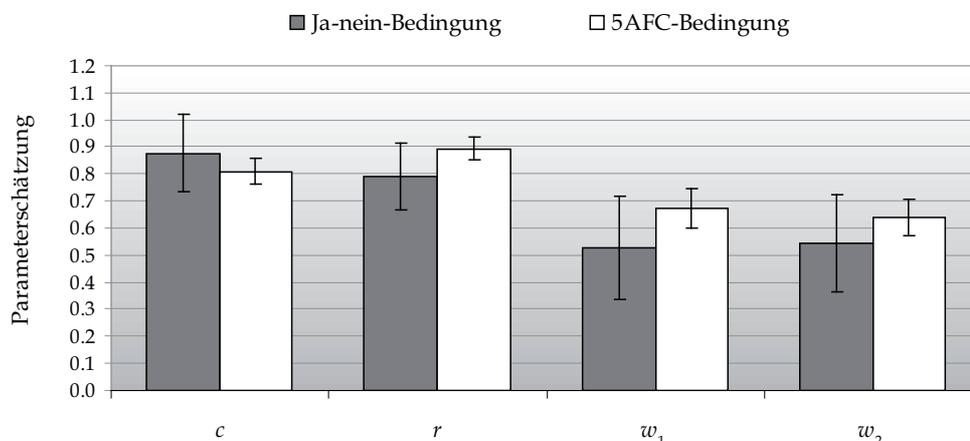


Abbildung 6.5: Diese Abbildung zeigt die Parameterschätzungen für die Gedächtnisparameter c , r , w_1 sowie w_2 des ADH-Modells. In Grau sind die Parameterschätzungen für die Ja-nein-Bedingung abgetragen, in Weiß die Schätzungen für die 5AFC-Bedingung. Als Fehlerbalken sind die 95%-Konfidenzintervalle dargestellt.

In Abbildung 6.5 sind die Schätzungen der *Gedächtnisparameter* abgetragen. Zunächst kann festgehalten werden, dass die verwendete Listenlänge und Lernzeit zu keinen Bodeneffekten führte, in den Parametern c und r sind tendenziell eher Deckeneffekte zu verzeichnen.

Im Vergleich zwischen den Gruppen sind deskriptiv leichte Unterschiede festzustellen. Diese wurden aber weder für den Parameter der Assoziationsenkodierung (c) signifikant ($\Delta G^2_{(1)} = 0.44$; kritischer Wert: $G^2_{krit, \alpha=0.05} = 3.84$) noch für den Abrufparameter ($\Delta G^2_{(1)} = 2.02$). Die Rekognitionsleistung für Items war in der Ja-nein-Bedingung etwas niedriger als in der 5AFC-Bedingung, signifikant unterschied sich die Rekognitionsleistung aber weder für das linke ($\Delta G^2_{(1)} = 2.46$) noch für das rechte Wort ($\Delta G^2_{(1)} = 0.53$).

Da in dieser Studie ein Nulleffekt postuliert wurde, muss überprüft werden, ob die Teststärke auch ausreichend hoch war, um überhaupt einen Effekt zwischen den Gruppen zu entdecken. Da die Power von den konkreten Parameterwerten abhängt, wurde eine Post-Hoc-Poweranalyse mit den geschätzten Parameterwerten als Ausgangswerten berechnet.

Die Teststärke, einen mindestens mittleren Parameterunterschied von .30 mit einem α -Level von .05 zu entdecken, ist für alle Gedächtnisparameter sehr hoch (r : $1-\beta = .99$; c : $1-\beta = .99$; w_1 : $1-\beta = .99$; w_2 : $1-\beta = .99$). Auch die Power, einen kleineren Parameterunterschied von .20 festzustellen, ist mit einer Power von mindestens .76 zufriedenstellend (r : $1-\beta = .94$; c : $1-\beta = .94$; w_1 : $1-\beta = .76$; w_2 : $1-\beta = .83$). Es kann daher geschlussfolgert werden, dass eine ausreichende Power vorlag, um Effekte auf die Erinnerungsparameter zu finden.

Im Hinblick auf die Gedächtnisparameter kann auch festgehalten werden, dass sich der Itemrekognitionsparameter für das linke Wort nicht maßgeblich von dem Rekognitionsparameter

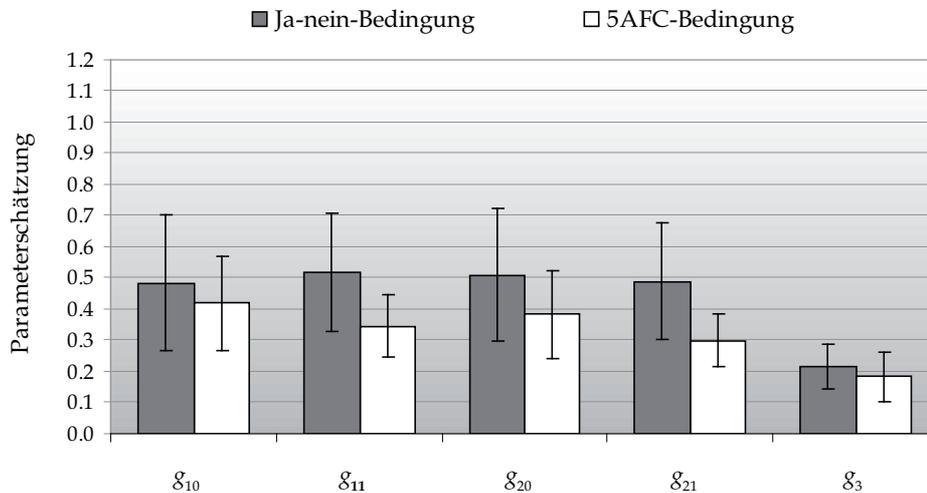


Abbildung 6.6: Hier sind die Schätzungen des ADH-Modells für die Rateparameter abgebildet. In Grau sind die Parameter der Ja-nein-Bedingung dargestellt, in Weiß die der 5AFC-Bedingung. Die Fehlerbalken repräsentieren die 95%-Konfidenzintervalle.

parameter für das rechte Wort unterschied. Weder in der Ja-nein-Bedingung wurde dieser Effekt signifikant ($\Delta G^2_{(1)} = 0.09$), noch in der 5AFC-Bedingung ($\Delta G^2_{(1)} = 1.52$). Eine Differenzierung dieser beiden Parameter scheint also in einer Standardbedingung nicht unbedingt notwendig, eine Reduktion auf das Submodell 10 ist in beiden Bedingungen möglich.

Die Schätzungen der *Rateparameter* sind in Abbildung 6.6 dargestellt. Hier sind weder Boden- noch Deckeneffekte zu verzeichnen. Im Vergleich zwischen den Gruppen sollten sich keine Unterschiede in den Rateparametern ergeben. Deskriptiv fallen die Parameter in der Ja-nein-Bedingung etwas höher aus als in der 5AFC-Bedingung. Dies ist aber nur eine Tendenz, und keiner der paarweisen Vergleiche erreichte das Signifikanzniveau (g_{10} : $\Delta G^2_{(1)} = 0.09$; g_{11} : $\Delta G^2_{(1)} = 1.31$; g_{20} : $\Delta G^2_{(1)} = 0.29$; g_{21} : $\Delta G^2_{(1)} = 2.82$; g_3 : $\Delta G^2_{(1)} = 0.58$), obwohl die Power einen großen Effekt zu entdecken mit $1 - \beta > .77$ zufriedenstellend war.

Innerhalb jeder Bedingung weisen die ausdifferenzierten Item-Rateparameter g_{10} , g_{11} , g_{20} , g_{21} keine großen Schwankungen auf. In der Ja-nein-Bedingung bewegen sich die Rateparameter in einem engen Bereich von 0.48 bis 0.51, in der 5AFC-Bedingung reichen die Werte von 0.29 bis 0.41. Auch hier scheint eine Ausdifferenzierung nicht unbedingt notwendig.

In der 5AFC-Bedingung können die Rateparameter auf zwei Rateparameter reduziert werden, die widerspiegeln, ob nur ein Wort geraten wurde oder beide Wörter geraten wurden ($\Delta G^2_{(2)} = 3.54$; $G^2_{krit,\alpha=0.05} = 5.99$; Modell 9a). Eine Einschränkung auf zwei Rateparameter, die die Position des geratenen Wortes repräsentieren (Modell 9b), ist in dieser Bedingung nicht möglich ($\Delta G^2_{(2)} = 10.71$). In der Ja-nein-Bedingung ist sowohl eine Ein-

Tabelle 6.3: Parameterschätzungen (Par. Schätzungen) des ADH-Modells für Experiment 1a sowie die dazugehörigen 95%-Konfidenzintervalle (95% KI)

Parameter	Ja-nein-Bedingung		5AFC-Bedingung	
	Par.Schätzung	95% KI	Par.Schätzung	95% KI
c	0.88	0.73 – 1.02	0.81	0.76 – 0.86
r	0.79	0.67 – 0.92	0.89	0.85 – 0.93
w_1	0.53	0.34 – 0.72	0.67	0.60 – 0.75
w_2	0.54	0.36 – 0.72	0.64	0.57 – 0.71
g_{10}	0.48	0.26 – 0.70	0.42	0.27 – 0.57
g_{11}	0.52	0.33 – 0.71	0.34	0.24 – 0.44
g_{20}	0.51	0.30 – 0.72	0.38	0.24 – 0.52
g_{21}	0.49	0.30 – 0.68	0.30	0.21 – 0.38
g_3	0.21	0.14 – 0.29	0.18	0.10 – 0.26
D	0.35	0.11 – 0.60	0.00	–0.32 – 0.32
a	0.15	–0.15 – 0.45	0.00	–0.31 – 0.31

Anmerkung.

Auf einem α -Niveau von .05 lagen keine signifikanten Effekte zwischen den Gruppen vor.

beschränkung auf zwei Prozessparameter ($\Delta G^2_{(2)} = 0.41$; Modell 9a) als auch auf zwei Positionsparameter umsetzbar ($\Delta G^2_{(2)} = 0.50$; Modell 9b). Hier ist sogar ein Rateparameter für Items ausreichend ($\Delta G^2_{(3)} = 0.51$; $G^2_{krit,\alpha=.05} = 7.81$; Modell 8c). Auch die zusätzliche Gleichsetzung der Itemparameter ($w_1 = w_2$), welche zu Modell 8a/b bzw. zu Modell 7 führt, kann vorgenommen werden (Modell 8a: $\Delta G^2_{(3)} = 0.55$; Modell 8b: $\Delta G^2_{(3)} = 0.54$; Modell 7: $\Delta G^2_{(4)} = 0.55$).

Die verbleibenden Parameter der Distraktorentdeckung (siehe Tabelle 6.3) unterscheiden sich zwischen den beiden Bedingungen nicht. Sowohl der Parameter der Distraktorentdeckung für Items konnte zwischen den Gruppen gleichgesetzt werden ohne die Modellanpassung signifikant zu verschlechtern ($\Delta G^2_{(1)} = 1.54$) als auch der Parameter der Distraktorentdeckung für Assoziationen ($\Delta G^2_{(1)} = 0.74$).

Allerdings weisen diese Parameter Bodeneffekte auf, und die breiten Konfidenzintervalle deuten auf ungenaue Schätzungen hin. Für den Parameter D sind die unreliablen Schätzungen wahrscheinlich eine Folge des niedrigen Anteils von Distraktoren im Test und sollten sich durch eine Erhöhung des Distraktorenanteils aufheben lassen. Auch für den Parameter a gibt es Manipulationen, unter denen sich reliablere Schätzungen ergeben sollten, darauf wird in Experiment 2 eingegangen.

Zusammenfassend kann festgehalten werden, dass in dieser ersten Modellanwendung eine gute Modellanpassung und lokale Identifizierbarkeit vorlag. Mit Ausnahme der Distraktorentdeckungsparameter konnten alle Parameter reliabel geschätzt werden, und es lagen keine gravierenden Boden- oder Deckeneffekte vor. Die Parameter konnten trotz ausrei-

chender Power ausnahmslos zwischen der Ja-nein-Bedingung und der 5AFC-Bedingung gleichgesetzt werden, ohne die Modellanpassungsstatistik signifikant zu verschlechtern.

6.1.3 Diskussion

In diesem Experiment wurden zwei alternative Aufgabenformate getestet. Im simultanen Antwortformat sollten die Probanden ihr Urteil in einem 5AFC-Test abgeben, das sequentielle bestand aus drei aufeinander folgenden Ja-nein-Abfragen.

Die Modellanalysen zeigten keine Unterschiede der Gedächtnisparameter in Abhängigkeit des Aufgabenformats. Das Ausbleiben dieser Effekte kann keine Folge unzureichender Teststärke sein; diese war zufriedenstellend. In den Rateparametern zeichnete sich eine leichte Tendenz zu konservativeren Ratetendenzen in der 5AFC-Bedingung ab (vgl. Hicks & Marsh, 1999), dies war aber nur eine Tendenz, die nicht das Signifikanzniveau erreichte. Auch die Parameter der Distraktorentdeckung konnten problemlos zwischen den Gruppen gleichgesetzt werden.

Die Ergebnisse bestätigen die Hypothese, dass die Art der Answerfassung keine Auswirkungen auf die Parameter des ADH-Modells hat, und konnten die Paradigmenunabhängigkeit erfolgreich zeigen. Die Resultate stehen in Einklang mit der Arbeit von Meiser (2005), der im Quellengedächtnisparadigma ebenfalls keine Effekte des Antwortformats feststellen konnte.

Im Theorieteil dieses Experiments wurde bereits diskutiert, dass bei vergleichbar komplexen Quellengedächtnisaufgaben vor allem das sequentielle Antwortformat Verwendung findet (von Hecker & Meiser, 2005; Meiser & Bröder, 2002; Meiser & Hewstone, 2004, 2006; Meiser et al., 2007). Obwohl die Gründe dieser Wahl nicht weiter ausgeführt wurden, scheint es plausibel, die einfachere Handhabung für die Versuchsperson als ausschlaggebend anzunehmen. Evidenz für diese Argumentation kann auch aus Experiment 1a gezogen werden: In der 5AFC-Bedingung wurden die Versuchsleiter häufig gefragt, ob die Probanden ihre gerade getroffene Entscheidungen im Nachhinein korrigieren könne, da die Antwortoption aus Versehen falsch ausgewählt wurde. Diese Fehler bei der Beurteilung traten in der Ja-nein-Bedingung bei keinem der Probanden auf. Es wurde zwar keine systematische Befragung der Teilnehmer durchgeführt, so dass an dieser Stelle keine Analysen berichtet werden können, die Auskünfte der Versuchsleiter weisen aber auf eine höhere Komplexität des 5AFC-Formats hin.

Wahrscheinlich könnte dieses Problem bei Studenten durch eine verlängerte Übungsphase minimiert werden. Das ADH-Modell soll aber auch in der gerontologischen Forschung genutzt werden. Da ältere Probanden meist weniger vertraut mit computergesteuerten Tests sind, scheint das 5AFC-Format für diese Probandengruppe ungeeignet zu sein. Daher wird in den folgenden Experimenten das sequentielle Ja-nein-Format verwendet.

Gegenüber den Vorteilen des Ja-nein-Formats ist im Einzelfall der Nachteil abzuwägen, dass keine sinnvolle Reaktionszeitanalyse vorgenommen werden kann. Falls also Hypothe-

sen über Reaktionszeiten vorliegen, scheint das 5AFC-Format in Verbindung mit einer ausgiebigen adaptiven Trainingsphase angemessener zu sein.

Im Hinblick auf die Modellevaluation kann zusammengefasst werden, dass das Modell eine befriedigende Modellanpassungsstatistik aufweist und die Parameterschätzungen - unter Ausnahme der Distraktorentdeckungsparameter - mit engen Konfidenzintervallen reliabel geschätzt werden konnten. Die Rateparameter konnten in der Ja-nein-Bedingung problemlos auf einen Itemrateparameter und einen Rateparameter für die Assoziation (g_3) reduziert werden (Modell 8c). In der 5AFC-Bedingung zeigte die Reduktion nach den vorliegenden Rateprozessen eine signifikant bessere Modellanpassung (Modell 9a) als die Reduktion aufgrund der Position (Modell 9b). Ob sich diese Ergebnisse replizieren lassen, wird anhand der folgenden Experimente untersucht. Des Weiteren scheint in einem Standardexperiment wie Experiment 1a eine Ausdifferenzierung der Itemrekognitionsparameter nicht unbedingt notwendig zu sein, es konnten keine signifikanten Unterschiede festgestellt werden (Modell 10). Trotzdem sollen im nächsten Experiment die beiden Itemrekognitionsparameter validiert werden, denn es sind durchaus Situationen vorstellbar, in denen sich eine Differenzierung der Itemparameter ergeben sollte. Liegen beispielsweise nicht zwei Wörter vor, sondern die Kombination von einem Wort und einem Bild, so sollten unterschiedliche Erinnerungsleistungen an die verschiedenen Itemklassen auch in unterschiedlichen Parameterschätzungen reflektiert werden. Im Sinne eines möglichst globalen Modelleinsatzes scheint es daher gerechtfertigt, eine ausdifferenzierte Modellversion zu validieren.

6.2 Experiment 1b

Ziel des Experiments 1b ist die Validierung der Itemparameter w_1 und w_2 , welche die Wahrscheinlichkeit einer erfolgreichen Erinnerung an ein einzelnes Wort repräsentieren. Im ADH-Modell stellen die Itemrekognitionsparameter hybride Parameter dar, d.h. es wird nicht zwischen Enkodierung und Abruf getrennt. Zur Validierung der Itemparameter gilt es, eine unabhängige Variable zu finden, die sich positiv oder negativ auf die Itemrekognitionsleistung auswirkt. Dabei sollte möglichst eine Manipulation gewählt werden, die experimentell etabliert ist und deren Auswirkung auf die Itemrekognitionsleistung klar vorhersehbar ist.

Hierfür kommt z.B. die Darbietungszeit in Betracht. Seit den Anfängen der experimentellen Gedächtnisforschung ist bekannt, dass die Lernzeit ein wesentliches Kriterium für den späteren Erinnerungserfolg darstellt. Dies gilt für unterschiedliches Itemmaterial z.B. für visuelle sowie verbale Informationen (Bugelski, Kidd & Segmen, 1968; Carroll & Burke, 1965; Potter & Levy, 1969). Der Effekt ist weiterhin nicht auf eine spezifische Art der Gedächtnistestung beschränkt, sondern konnte sowohl in der freien Reproduktion (Craik & Rabinowitz, 1985; Roberts, 1972; Waugh, 1967), der geförderten Reproduktion (Bugelski,

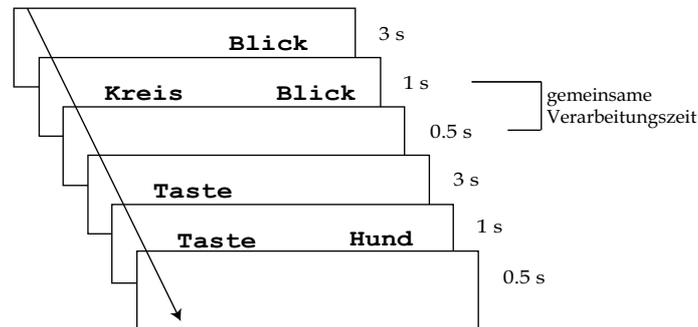


Abbildung 6.7: Hier wird die experimentelle Manipulation der ‚gemeinsamen Verarbeitungszeit‘ nach Jack (1986) dargestellt. Die gemeinsame Verarbeitungszeit definiert die Zeitspanne, die für die Enkodierung des zuletzt dargebotenen Wortes und damit auch für das Wortpaar zur Verfügung steht. Sie berechnet sich aus der Präsentationszeit des zweiten Wortes und aus der Dauer des Interstimulusintervalls.

1962; Carroll & Burke, 1965) als auch in Rekognitionsaufgaben (Hicks & Marsh, 1999; Potter & Levy, 1969) repliziert werden. Zahlreiche experimentelle Studien haben weitere Aspekte des Faktors Lernzeit untersucht. So wurde z.B. im Rahmen der ‚Total-Time-Hypothese‘ erforscht, ob nur die absolute Lernzeit für den Lernerfolg ausschlaggebend ist oder ob die Aufteilung dieser Lernzeit einen Einfluss auf die Erinnerungsleistung hat (Cooper & Pantle, 1967). Diese weiterführenden Forschungsfragen sollen hier allerdings nicht weiter ausgeführt werden, da für das vorliegende Experiment auf den Basiseffekt einer zunehmenden Gedächtnisleistung mit vermehrter Lernzeit Bezug genommen wird.

Waugh (1967) testete in einer Studie zur freien Reproduktion von Wörtern Darbietungszeiten von einer bis sechs Sekunden. Sie konnte dabei mit zunehmender Darbietungszeit eine klare Verbesserung der Reproduktionsleistung feststellen. Auch zwischen vier und sechs Sekunden war eine Leistungssteigerung möglich. Es wurde mit diesen Präsentationszeiten noch kein Leistungsplateau erreicht. Bugelski (1962) manipulierte in einem Experiment zum Paar-Assoziationslernen ebenfalls die Darbietungszeit, allerdings beginnend mit sechs Sekunden. Er verwendete die Zeiten 6, 8, 10, 12 und 19 Sekunden. Als Lernkriterium erhob er die Anzahl der Durchgänge bis zur fehlerfreien Wiedergabe. Auch er stellte einen signifikanten Effekt der Darbietungszeit fest, allerdings war der Effekt bei den kürzeren Darbietungszeiten von 6 bis 10 Sekunden stärker ausgeprägt als bei den Darbietungszeiten ab 10 Sekunden. Der Zusammenhang von Darbietungszeit und Lernerfolg ist also sicherlich kein linearer, sondern eher ein positiv beschleunigter, asymptotischer Zusammenhang (Calfée & Anderson, 1971; Carroll & Burke, 1965; Palfai, Halperin & Hoyer, 2003; Roberts, 1972). Für das vorliegende Experiment soll daher eine relativ kurze Darbietungszeit von einer Sekunde mit einer längeren Darbietungszeit von vier Sekunden kontrastiert werden (z.B. Hicks & Marsh, 1999; Palfai et al., 2003; Waugh, 1967).

Über die Manipulation der Darbietungszeit hinaus soll auch die gemeinsame Darbietungszeit des Wortpaares beeinflusst werden. Dies soll in Anlehnung an eine Studie von

Jack (1968) geschehen. In dieser Studie zum Paar-Assoziationslernen wurden Wortpaare entweder simultan dargeboten, d.h. beide Wörter wurden gleichzeitig und gleich lange gezeigt, oder aber die Präsentation erfolgte sequentiell (vgl. auch Van Mondfrans & Travers, 1965; Mueller & Travers, 1965). In der sequentiellen Bedingung wurde das linke Wort des Wortpaares zunächst für eine Sekunde angezeigt, danach folgte die Präsentation des zweiten, rechten Wortes. Experimentell manipuliert wurde in dieser Studie die Präsentationsdauer des zweiten Wortes bzw. die ‚gemeinsame Verarbeitungszeit‘ („Hook-up-Time“). Der Begriff ‚gemeinsame Verarbeitungszeit‘ definiert die Zeitspanne von Darbietungsbeginn des zweiten Wortes eines Wortpaares bis zur Darbietung des ersten Wortes des nächsten Wortpaares. Diese Spanne setzt sich also aus der Präsentationszeit für das zweite Wort eines Wortpaares und dem Interstimulusintervall zusammen. Die gemeinsame Verarbeitungszeit nach Jack beschreibt damit das Zeitintervall, das maximal zur Verarbeitung des zweiten Wortes bzw. zur Verarbeitung des Wortpaares genutzt werden kann (siehe Abbildung 6.7).

Die Ergebnisse von Jack (1968) konnten zeigen, dass bei vergleichbarer Gesamtlernzeit die simultane Darbietungsbedingung zu höheren Leistungen in der geförderten Reproduktion führte (siehe auch Mueller & Travers, 1965). Darüber hinaus stieg die Leistung in der sequentiellen Bedingung mit Zunahme der gemeinsamen Verarbeitungszeit an. Dieser Effekt war besonders drastisch innerhalb kurzer Darbietungszeiten von 1.25, 2 und 3 Sekunden.

Anlehnend an die berichteten Experimente sollte also neben der Lernzeit, die für ein einzelnes Item zur Verfügung steht, auch die Lernzeit, die für das Wortpaar aufgewendet werden kann, manipuliert werden.

6.2.1 Methode

Um sowohl die Darbietungszeit für die einzelnen Wörter als auch die Enkodierungszeit für das Wortpaar zu beeinflussen, wurden im vorliegenden Experiment drei experimentelle Bedingungen erzeugt (siehe Abbildung 6.8). Die Wortpaare wurden entweder *simultan* oder *sequentiell* dargeboten. In der simultanen Darbietung war das Wortpaar für vier Sekunden am Bildschirm zu sehen. In der sequentiellen Bedingung erschien zunächst ein Wort und wurde alleine für drei Sekunden dargeboten, dann folgte das zweite Wort für eine Sekunde. Es wurde ausbalanciert, ob das linke oder das rechte Wort als erstes Wort dargeboten wurde, konstant war, dass das erste Wort immer lange präsentiert wurde.

In der simultanen Bedingung konnte das Wortpaar also 4500 ms enkodiert werden (4000 ms Darbietungszeit plus 500 ms Interstimulusintervall), in der sequentiellen Bedingung standen für diesen Prozess nur 1500 ms zur Verfügung.

Die Erinnerungsleistung an die Einzelwörter sollte von der Präsentationszeit abhängen: Wörter, die in der sequentiellen Bedingung nur eine Sekunde gezeigt werden, sollten schlechter rekonstruiert werden als Wörter, die in der sequentiellen Bedingung drei Sekun-

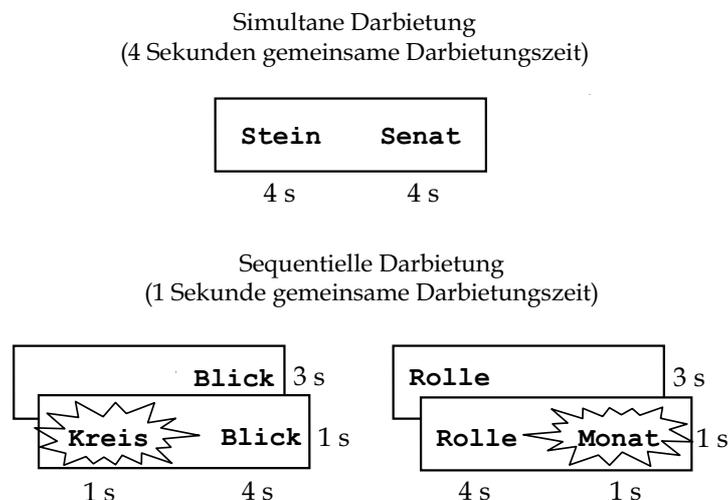


Abbildung 6.8: In dieser Abbildung ist die Manipulation der Darbietungszeit dargestellt. In der simultanen Bedingung wurden beide Wörter gleichzeitig und gleich lange dargeboten. In der sequentiellen Bedingung wurde zunächst ein Wort für drei Sekunden alleine dargeboten, danach wurde das zweite Wort für eine Sekunde gezeigt. Die gemeinsame Darbietungszeit des Wortpaares betrug also nur eine Sekunde.

den alleine dargeboten werden. Da nicht kontrolliert oder erfasst werden kann, wie die einzelnen Wörter eines simultan dargebotenen Wortpaares verarbeitet werden, wird für diese Darbietungsbedingung nur eine Vermutung, keine Hypothese geäußert. Es wird angenommen, dass die Rekognitionsleistung für simultan dargebotene Wörter zwischen der Leistung für kurz und lang dargebotene Wörter liegt. Dies ist plausibel unter der Annahme, dass die Versuchspersonen die zur Verfügung stehenden vier Sekunden ungefähr gleich auf die beiden Wörter aufteilen.

Die Dauer der gemeinsamen Darbietungszeit eines Wortpaares sollte sich auch auf den Enkodierungsparameter für Assoziationen auswirken. In der simultanen Bedingung sollten die Wortpaare mit einer höheren Wahrscheinlichkeit als Paar enkodiert werden als in der sequentiellen Bedingung. Welches Wort in der sequentiellen Bedingung zuerst dargeboten wird, sollte keinen Einfluss auf die assoziative Enkodierung haben.

Teilnehmer und Design

An diesem Experiment nahmen 60 Psychologiestudenten zwischen 18 und 30 Jahren teil ($M = 22.22$; $SD = 2.33$).²⁹ Ihnen wurden Versuchspersonenstunden über die Dauer der Teilnahme von 45 Minuten ausgestellt. Es wurden maximal vier Versuchspersonen gleichzeitig erhoben.

²⁹Von den 60 Probanden, die an Experiment 1b teilnahmen, haben 43 Probanden auch an Experiment 1a teilgenommen. Die Daten der simultanen Bedingung dieser Teilnehmer wurden unter einem anderen Gesichtspunkt in Experiment 1a analysiert.

Als unabhängige Variable wurde die Darbietungszeit in drei Stufen innerhalb der Versuchspersonen manipuliert. Die Hälfte der Wörter wurde dabei in der simultanen Bedingung, die andere Hälfte in der sequentiellen Bedingung präsentiert. Bei den Wortpaaren der sequentiellen Bedingung wurde genauso häufig das linke Wort als erstes Wort präsentiert wie das rechte.

Ablauf

Die Versuchssteuerung wurde mit dem Programm E-Prime erstellt (Psychological Software Tools, 2000). Die Erhebung wurde an Intel-Pentium(4)-Rechnern mit einer Monitorgröße von 19 Zoll durchgeführt. Genau wie in Experiment 1a wurden die Teilnehmer zunächst instruiert, dass sie sich im folgenden Experiment Wortpaare für einen späteren Test einprägen sollen und dass sowohl die einzelnen Wörter als auch deren Kombination von Bedeutung ist. Es folgte auch hier zunächst ein Übungsdurchgang, in dem fünf Wortpaare präsentiert und anschließend getestet wurden. Die Details der Wortdarbietung (bis auf die Darbietungsdauer) entsprachen der Darstellung in Experiment 1a. Die Teilnehmer gaben ihre Antworten über das sequentielle Ja-nein-Aufgabenformat ab. Auch in diesem Experiment wurde die Tastenbelegung ausbalanciert, um eine mögliche Antwortverzerrung durch Händigkeitspräferenzen auszugleichen.

Die 50 Wortpaare wurden zufällig aus einer Liste von 280 Wörtern zusammengestellt. Wieder wurden drei ‚Füller-Wortpaare‘ jeweils am Anfang und Ende der Lernliste eingefügt. Nach einer Distraktoraufgabe von einer Minute folgte die Testphase. Hier sollten 70 Wortpaare beurteilt werden. Davon waren, jeweils vollkommen zufällig zusammengestellt, 20 Wortpaare in alter intakter Kombination, 20 Wortpaare neu kombiniert aus zwei alten Wörtern, 10 Target-Distraktor-Kombinationen, 10 Distraktor-Target-Kombinationen und 10 Wortpaare bestehend aus zwei Distraktoren. Die verschiedenen Wortkombinationen wurden in zufälliger Reihenfolge über die Testliste verteilt dargeboten. Nach Beendigung der Testphase folgte eine Zwischenaufgabe, die circa drei Minuten in Anspruch nahm. Dann wurde ein neuer Lern- und Testdurchgang gestartet. Insgesamt wurden also zwei Lernlisten mit je 50 Wortpaaren (100 Einzelworte) gelernt und zwei Testlisten von 70 Wortpaaren (140 Einzelworte) getestet.

Material

Für das erste Experiment (1a und 1b) wurden die beiden parallelisierten Wortlisten 1 und 2 verwendet. Die zentralen Kennwerte wurden im vorherigen Kapitel in Tabelle 6.1 aufgelistet, zur genaueren Zusammenstellung des Wortmaterials sei auf Kapitel 6.1.1 sowie Anhang C verwiesen.

6.2.2 Ergebnisse

Im Auswertungsteil werden wieder zuerst die relativen Häufigkeiten berichtet, dann die Ergebnisse der multinomialen Modellanalyse.

In diesem Experiment wurde die Präsentationszeit der Wortpaare in drei Bedingungen manipuliert. Durch die zufällige Zusammenstellung der Wörter in der Testphase konnten verschiedene Kombinationen von Testpaaren resultieren: Die Wörter konnten zu Wortpaaren rekombiniert werden, in denen (1) beide Wörter vier Sekunden dargeboten wurden, (2) das rechte Wort vier Sekunden, das linke nur eine Sekunde gelernt wurde, (3) das rechte Wort eine Sekunde, das linke aber vier Sekunden präsentiert wurde oder (4) beide Wörter für nur eine Sekunde gezeigt wurden. Obwohl nur die ersten drei Bedingungen für die Auswertung von Interesse waren, wurden alle Bedingungen getestet, damit die Zusammenstellung der Wortpaare wirklich vollkommen zufällig erfolgen konnte. Aufgrund der experimentellen Manipulation fließen nur wenige Wortpaare in die Kategorie von zwei kurz präsentierten Wörtern ein. Diese Kategorie wurde daher in der Auswertung nicht berücksichtigt.

In der Tabelle 6.4 sind die relativen Anwohnhäufigkeiten für die drei experimentellen Bedingungen dargestellt. Bereits an den relativen Häufigkeiten ist zu erkennen, dass alte intakte Wortpaare in der simultanen Bedingung häufiger richtig klassifiziert werden (.717) als in den beiden sequentiellen Bedingungen (.675; .632). Auch in der Itemerkennung kann ein Effekt der längeren Darbietungszeit bereits vermutet werden, da Zielitem-Distraktor-Wortpaare, deren Zielitem länger präsentiert wurde, eine höhere Rate richtiger Klassifikationen erfahren (.598) als Zielitem-Distraktor-Kombinationen, deren Zielwort nur kurz gezeigt wurde (.498). Da in diesen Rohdaten allerdings nur die vollständig richtige Klassifikation eines Wortpaares berücksichtigt wurde und keine Korrektur der falschen Alarme erfolgte, soll eine inferenzstatistische Auswertung der multinomialen Analyse aufzeigen, inwieweit diese Unterschiede sich auch signifikant in den Parameterschätzungen niederschlagen.

Multinomiale Analyse

Die multinomiale Analyse basiert auf den absoluten Häufigkeiten, die im Anhang F nachgeschlagen werden können. In diesem Experiment lagen 31 Freiheitsgrade vor. Die Bäume für alte intakte, neu kombinierte, Target-Distraktor- und Distraktor-Target-Kombinationen kamen je dreimal in diesem Modell vor, für jede der experimentellen Bedingungen einmal. Der Distraktor-Distraktor-Baum war nur einmal in dem Modell enthalten, da die Manipulation der Darbietungszeit innerhalb der Versuchspersonen erfolgte und daher nur eine Klasse von Distraktoren vorlag.

Das Modell zeigte mit $G_{(31)}^2 = 31.13$ ($G_{krit, \alpha=.001}^2 = 61.09$) eine sehr gute Anpassung an die Daten, wieder lag lokale Identifizierbarkeit vor. Der Rang der Jacobischen Matrix ent-

Tabelle 6.4: *Relative Antworthäufigkeiten auf die drei unterschiedlichen Itemtypen von Experiment 1b*

(a) simultane Darbietungsbedingung

Wahre Itemkat.	Antwort der Versuchsperson				
	Altes WP	Neu komb. WP	Target Dis.	Dis. Target	Dis.Dis.
Altes WP	<i>.717</i>	.103	.069	.063	.048
Neu komb. WP	.111	<i>.495</i>	.176	.166	.051
Target Dis.	.032	.188	<i>.555</i>	.063	.162
Dis. Target	.037	.195	.085	<i>.542</i>	.142
Dis. Dis.	.024	.100	.228	.211	<i>.438</i>

(b) sequentielle Darbietungsbedingung (1 Sekunde, 4 Sekunden)

Wahre Itemkat.	Antwort der Versuchsperson				
	Altes WP	Neu komb. WP	Target Dis.	Dis. Target	Dis.Dis.
Altes WP	<i>.675</i>	.125	.051	.092	.057
Neu komb. WP	.146	<i>.446</i>	.146	.185	.077
Target Dis.	.047	.227	<i>.498</i>	.074	.154
Dis. Target	.067	.211	.067	<i>.565</i>	.090

(c) sequentielle Darbietungsbedingung (4 Sekunden, 1 Sekunde)

Wahre Itemkat.	Antwort der Versuchsperson				
	Altes WP	Neu komb. WP	Target Dis.	Dis. Target	Dis.Dis.
Altes WP	<i>.632</i>	.142	.107	.063	.056
Neu komb. WP	.162	<i>.432</i>	.236	.142	.027
Target Dis.	.060	.186	<i>.598</i>	.066	.090
Dis. Target	.066	.203	.090	<i>.472</i>	.169

Anmerkung.

Da die Manipulation der Darbietungszeit innerhalb der Teilnehmer erfolgte, liegt nur eine Distraktorklasse vor und in der modellbasierten Auswertung ist der Baum für die Kombination zweier Distraktoren nur einmal enthalten. Daher sind die Antworten auf Distraktor-Distraktor-Kombinationen nur in Tabelle (a) aufgeführt. Kursiv sind jeweils die korrekten Antworten hervorgehoben. Die Häufigkeiten sind über beide Testlisten aggregiert.

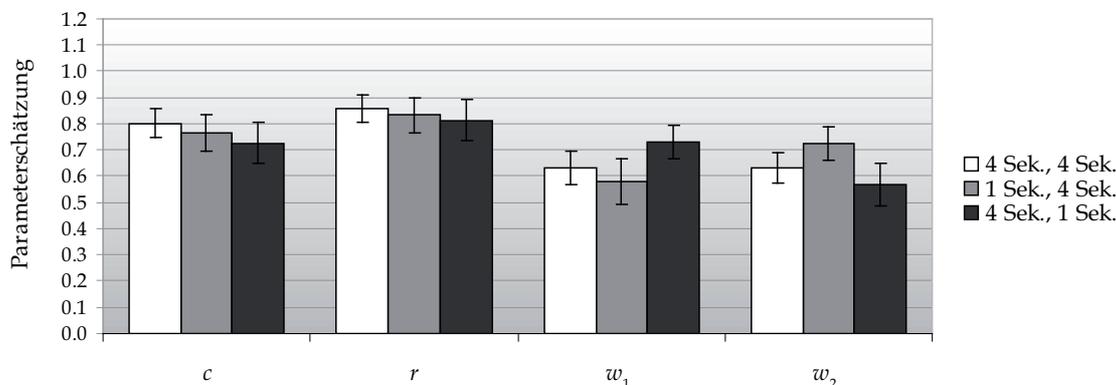


Abbildung 6.9: In dieser Abbildung sind die Parameterschätzungen der Gedächtnisparameter des ADH-Modells dargestellt. In Weiß ist die Bedingung der simultanen Darbietungszeit abgetragen; die Wörter wurden hier beide gleichzeitig und für vier Sekunden gezeigt. In Hellgrau sind die Schätzungen für die sequentielle Darbietungsgruppe dargestellt in der das rechte Wort nur für eine Sekunde gezeigt wurde und das linke für vier Sekunden. In Dunkelgrau sind die Schätzungen der zweiten experimentellen Gruppe mit einer langen Darbietungszeit für das rechte Wort und einer kurzen Darbietungszeit für das linke Wort zu sehen. Die Fehlerbalken geben jeweils die 95%-Konfidenzintervalle an.

sprach der Parameteranzahl und die mehrfach berechneten Parameterschätzungen waren bei gleicher Modellanpassung identisch. Wie aus Abbildung 6.9 entnommen werden kann, fallen die Itemerkognitionsparameter für lang dargebotene Wörter deskriptiv höher aus als die Itemerkognitionsparameter für kurz präsentierte Items ($(w_{1,lang} = 0.73) > (w_{1,kurz} = 0.57)$; $(w_{2,lang} = 0.72) > (w_{2,kurz} = 0.56)$). Nicht nur deskriptiv, sondern auch inferenzstatistisch zeigte sich in den Itemparametern ein signifikanter Effekt. Dieser ist sowohl im Vergleich der linken Items, die vier vs. eine Sekunde dargeboten wurden, hoch signifikant ($\Delta G_{(1)}^2 = 12.62$), als auch im Vergleich von rechts dargebotenen Items ($\Delta G_{(1)}^2 = 13.98$). Darüber hinaus unterscheiden sich die Items, die auf der rechten Seite des Wortpaares kurz dargeboten wurden, signifikant von Items, die auf der linken Seite lange gezeigt wurden ($\Delta G_{(1)}^2 = 11.56$). Umgekehrt ist der Effekt auch signifikant für Items, die auf der rechten Position lang und auf der linken kurz gezeigt wurden ($\Delta G_{(1)}^2 = 14.89$). Die Darbietungszeit zeigt also in allen Paarvergleichen einen höchst signifikanten Effekt.

Deskriptiv ist auch in den Parametern der Clusterenkodierung ein kleiner Effekt sichtbar. Der Clusterenkodierungsparameter der simultanen Bedingung unterscheidet sich aber nicht signifikant von der ersten sequentiellen Bedingung ($\Delta G_{(1)}^2 = 0.98$). Der Unterschied zur zweiten sequentiellen Bedingung wird sehr knapp signifikant ($\Delta G_{(1)}^2 = 3.85$).

Der Parameter für den Clusterabruf unterscheidet sich in keiner der beiden sequentiellen Bedingungen signifikant von der simultanen Bedingung ($\Delta G_{(1)}^2 = 0.44$; $\Delta G_{(1)}^2 = 1.17$),

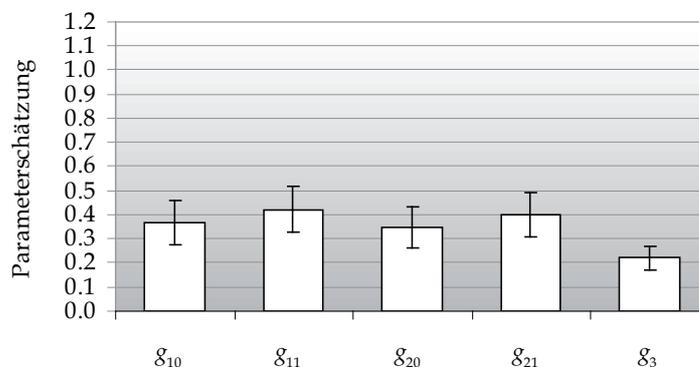


Abbildung 6.10: In dieser Grafik sind die Rateparameterschätzungen für Experiment 1b abgetragen. Da die experimentelle Manipulation innerhalb der Teilnehmer manipuliert wurde, liegt nur eine Klasse von Rateparametern vor. Die Fehlerbalken geben die 95%-Konfidenzintervalle an.

obwohl eine sehr hohe Power zur Entdeckung von Unterschieden von .20 vorliegt ($1 - \beta = .99$).

Die Rateparameter (vgl. Abbildung 6.10) sowie die Parameter der Distraktorentdeckung (vgl. Tabelle 6.5) zeigen keine Auffälligkeiten.

In den Rateparametern zeigte sich ein Muster, das die Ausdifferenzierung nach Rateprozessen nahelegt (Modell 9a), da g_{10} und g_{20} jeweils etwas niedriger ausfallen als g_{11} und g_{21} . Dies spiegelt sich auch in den Signifikanztests wider. Die Rateparameter können auf zwei Rateparameter eingeschränkt werden, die den Ablauf der Rateprozesse repräsentieren ($\Delta G_{(2)}^2 = 2.26$; Modell 9a), nicht jedoch auf zwei Parameter, die die Position der Wörter darstellen ($\Delta G_{(2)}^2 = 6.33$; Modell 9b). Da die experimentelle Manipulation keine Gleichsetzung der Itemparameter w_1 , und w_2 erlaubt, wurden weitere Submodelle nicht getestet.

6.2.3 Diskussion

In Experiment 1b sollten die Itemparameter für das linke und das rechte Wort validiert werden. Zu diesem Zweck wurden drei verschiedene Präsentationsmodi innerhalb von Versuchspersonen variiert. Die Wortpaare wurden (1) simultan dargeboten, d.h. beide Wörter eines Wortpaares waren gleichzeitig für vier Sekunden auf dem Bildschirm zu sehen, oder (2) sequentiell dargeboten, wobei zuerst das linke Wort für drei Sekunden alleine erschien und dann das rechte Wort hinzukam, oder (3) es erschien zuerst das rechte Wort für drei Sekunden alleine und dann folgte das linke Wort. Mit dieser Manipulation wurde zum einen die Enkodierungszeit für einzelne Wörter manipuliert, zum anderen stand auch unterschiedlich viel Zeit zur Enkodierung des Wortpaares zur Verfügung.

Tabelle 6.5: Parameterschätzungen (Par. Schätz.) und 95%-Konfidenzintervalle (95% KI) des ADH-Modells für Experiment 1b

Par.	4s – 4s		4s – 1s			1s – 4s	
	Par. Schätz.	95% KI	Par. Schätz.	95% KI		Par. Schätz.	95% KI
<i>c</i>	0.80	0.75 – 0.86	0.77	0.70 – 0.84	**	0.72	0.65 – 0.80
<i>r</i>	0.86	0.80 – 0.91	0.83	0.77 – 0.90		0.81	0.74 – 0.89
<i>w</i> ₁	0.63	0.57 – 0.69	0.58	0.49 – 0.67	*	0.73	0.67 – 0.80
<i>w</i> ₂	0.63	0.57 – 0.69	0.73	0.66 – 0.79	*	0.57	0.49 – 0.65
<i>g</i> ₁₀	0.37	0.28 – 0.46					
<i>g</i> ₁₁	0.42	0.33 – 0.52					
<i>g</i> ₂₀	0.35	0.26 – 0.43					
<i>g</i> ₂₁	0.40	0.31 – 0.49					
<i>g</i> ₃	0.22	0.17 – 0.27					
<i>D</i>	0.15	–0.04 – 0.33					
<i>a</i>	0.15	–0.03 – 0.34	0.00	–0.28 – 0.28		0.00	–0.27 – 0.27

*: Signifikante Effekte zwischen den beiden sequentiellen Bedingungen (4s – 1s vs. 1s – 4s) auf einem α -Niveau von .05.

** : Signifikante Effekte zwischen der simultanen Bedingung (4s – 4s) und der sequentiellen Bedingungen (1s – 4s) auf einem α -Niveau von .05.

Wie erwartet war die Erinnerungsleistung an kurz dargebotene Wörter schlechter als die Erinnerungsleistung an lang gezeigte Wörter. Dieser Effekt wirkte sich signifikant und selektiv auf die Itemparameter aus, womit dieses Experiment als erfolgreiches Validierungsexperiment für die beiden Itemparameter angesehen werden kann.

Die gemeinsame Darbietungszeit eines Wortpaares hatte zwar einen Effekt auf den Parameter der Clusterenkodierung, wenngleich dieser nicht besonders groß war. Gleichzeitig zeigte sich trotz einer sehr hohen Power kein Effekt auf den Parameter des Clusterabrufs. Die Ergebnisse von Experiment 1b sprechen daher für eine erfolgreiche Trennung von Enkodierungs- und Abrufprozessen durch das ADH-Modell. Die gemeinsame Darbietungszeit scheint aber keinen besonders starken Effekt auf die Enkodierung von Wortpaaren zu haben. Eventuell haben die Teilnehmer die intendierte gemeinsame Darbietungszeit von einer Sekunde verlängert, indem sie das Wortpaar sowohl im Interstimulusintervall als auch zu Beginn der Präsentation des folgenden Wortpaares noch wiederholt haben. Es könnte sein, dass eine solche Strategie den experimentell angestrebten Darbietungsunterschied abgeschwächt hat. Auf der anderen Seite ist auch denkbar, dass die Versuchspersonen eine sehr effiziente Strategie zur Enkodierung der Wortpaare verwendet haben und daher bereits eine kurze Darbietungszeit von einer Sekunde zur erfolgreichen Enkodierung ausreichend war. Eine weitere Validierung des Clusterenkodierungsparameters wird daher im nächsten Experiment angestrebt.

Das Modell zeigte in Experiment 1b eine sehr gute Anpassung an die Daten. Wie in Experiment 1a konnten auch in Experiment 1b die Rateparameter reduziert werden. Reduziert man die Parameter nach Rateprozessen anstatt nach Wortposition, so zeigte sich eine bessere Modellanpassung.

Kapitel 7

Experiment 2

Nachdem sich das Modell in den ersten beiden Experimenten (Experiment 1a und 1b) mit einer guten Datenanpassung bewährt hat und die Itemparameter erfolgreich validiert werden konnten, sollen in diesem Experiment die zentralen Parameter der Assoziationsenkodierung und des Assoziationsabrufs validiert werden.

Wie im theoretischen Teil bereits diskutiert wurde (Kapitel 4.1), stellt es eine Herausforderung dar, Enkodierungs- und Abrufprozesse über ein experimentelles Vorgehen zu trennen bzw. experimentelle Manipulationen zu finden, die sich gezielt nur auf Enkodierungsprozesse bzw. lediglich auf Abrufprozesse auswirken. Dies ist allerdings notwendig, um die entsprechenden Parameter des Modells zu validieren. Ausschlaggebend für dieses Ziel ist sogar weniger die genaue experimentelle Manipulation als vielmehr das experimentelle Design, in das eine geeignete Manipulation eingebettet ist, denn davon sind die Schlussfolgerungen, die aus den Daten gezogen werden können, abhängig.

Köhler (1947) argumentierte, dass die Auswirkungen von Enkodierungs- und Abrufmanipulationen in einem Drei-Gruppen-Design experimentell erfasst und getrennt werden können. Dazu muss eine erste Gruppe als Kontrollgruppe festgelegt werden. Im Vergleich zur Kontrollgruppe sollten in einer zweiten Gruppe lediglich die Enkodierungsbedingungen verändert werden. Unter diesen Voraussetzungen spiegeln die Veränderungen in dieser zweiten Gruppe den Effekt der Enkodierungsmanipulation wider. Werden in einer dritten Gruppe lediglich die Abrufprozesse manipuliert, alle anderen Prozesse konstant gehalten, so lassen sich nach Köhler (1947) in dieser Gruppe die reinen Effekte der Abrufmanipulation ablesen. Er schlussfolgerte, dass über ein solches Design die Auswirkungen von Enkodierungs- und Abrufmanipulationen ohne gegenseitige Konfundierung erfasst werden könnten. Wie Tulving (1983) allerdings einwendet, vernachlässigt Köhler (1947) das Auftreten von Wechselwirkungen zwischen den Enkodierungs- und Abrufmanipulationen. Es besteht die Möglichkeit, dass Enkodierungsmanipulationen sich durchaus auch auf Abrufprozesse auswirken können. Deshalb könnte die abhängige Variable der Enkodierungsgruppe sowohl von direkten Effekten der Enkodierungsmanipulation beeinflusst

		Abruf	
		X	Y
Enkodierung	A	AX	AY
	B	BX	BY

Abbildung 7.1: Diese Grafik skizziert den Aufbau des Enkodierungs-Abruf-Paradigmas nach Tulving (1983). Das Paradigma sieht ein Vier-Gruppen-Design vor in dem eine zweistufige Enkodierungsmanipulation (A,B) mit einer zweistufigen Abrufmanipulation (X,Y) gekreuzt wird.

sein, aber auch ein ‚Carry-over-Effekt‘ auf den Abruf kann nicht ausgeschlossen werden. Beispielsweise ist es denkbar, dass bestimmte Enkodierungsstrategien wie das Generieren von Verbindungswörtern sowohl während der Enkodierungsphase genutzt werden als auch in der Abrufsituation (Dunlosky et al., 2005). Um trotz dieser Einwände die Auswirkungen von Enkodierungs- und Abrufprozessen getrennt voneinander erfassen zu können, schlug Tulving (1983) das ‚Enkodierungs-Abruf-Paradigma‘ vor (‚encoding-retrieval paradigm‘). Dieses Paradigma sieht ein (mindestens) Vier-Gruppen-Design vor in dem zwei Enkodierungsmanipulationen mit zwei Abrufmanipulationen gekreuzt werden, und es hat den Vorteil, dass Wechselwirkungen zwischen den Manipulationen erfasst werden können (siehe Abbildung 7.1).

Dieses Paradigma soll nun am konkreten Beispiel von Experiment 2 erläutert werden: Eine Enkodierungsmanipulation (A), die eine schwache relationale Enkodierung zur Folge hat, soll mit einer Enkodierungsmanipulation (B) kontrastiert werden, die zu einer hohen relationalen Verarbeitung führt. In Gruppe A sollte daher die Wahrscheinlichkeit der Enkodierung von Assoziationen niedriger ausfallen als in Gruppe B. Dies sollte sich im c -Parameter der Assoziationsenkodierung zeigen. Parallel zu der Enkodierungsmanipulation gilt es, eine Abrufmanipulation zu finden, die in Gruppe X mit einer hohen Wahrscheinlichkeit zum Abruf von Assoziationen führt, in der Gruppe Y dahingegen die Abrufwahrscheinlichkeit senkt. Diese Manipulation sollte sich im Parameter des Abrufs von Assoziationen (r) widerspiegeln. Abbildung 7.2 zeigt, wie sich die Manipulationen auf die Parameter c und r auswirken sollten. Durch Wechselwirkungen könnte es der Fall sein, dass z.B. der Parameter r in der flachen Enkodierungsgruppe niedriger ausfällt als in der Bedingung tiefer Enkodierung, sich hier also ein Effekt der Enkodierungsmanipulation auf die Abrufparameter niederschlägt. Dieses Ergebnismuster wäre für die Validierung aber

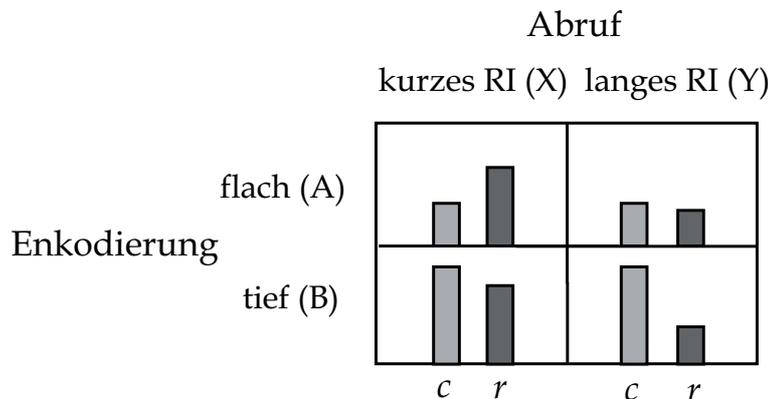


Abbildung 7.2: In Anlehnung an das Enkodierungs-Abruf-Paradigma von Tulving (1983) sind hier die Hypothesen von Experiment 2 darstellt. Innerhalb der vier Rechtecke sind jeweils die vorhergesagten Parameterschätzungen für den Enkodierungsparameter c und den Abrufparameter r abgetragen. Der Enkodierungsparameter sollte innerhalb einer Zeile konstant bleiben d.h. vom Retentionsintervall nicht beeinflusst werden; dafür aber zwischen den Zeilen variieren, d.h. den Effekt der Enkodierungsmanipulation widerspiegeln. Umgekehrtes gilt für den Abrufparameter: dieser sollte innerhalb der Zeilen variieren und damit den Effekt des Retentionsintervalls anzeigen, aber zwischen den Zeilen konstant bleiben und damit keine Beeinflussung durch die Enkodierungsmanipulation reflektieren.

nicht notwendigerweise problematisch. Wichtig wäre in diesem Fall, dass der Effekt der Abrufmanipulation zwischen Gruppe X und Y sich stärker auf den Abrufparameter auswirkt als der Effekt der Enkodierungsmanipulation. Um von einer erfolgreichen Validierung zu sprechen, müssten sich die *stärksten* Effekte im c -Parameter zwischen den Gruppen A und B zeigen, die *stärksten* Effekte im r -Parameter zwischen den Gruppen X und Y. Leichte Beeinflussungen, die nicht diesem Muster entsprechen, können dann als unkritisch bewertet werden. Das Enkodierungs-Abruf-Paradigma ist daher geeignet, die konvergenten und diskriminanten Effekte der Enkodierungs- und Abrufmanipulationen auf die entsprechenden Parameter zu erfassen, und wird in diesem zweiten Experiment angewandt.

Im Folgenden sollen nun adäquate Manipulationen für die Beeinflussung der assoziativen Enkodierung und des Assoziationsabrufs diskutiert werden, die ihre konkrete Anwendung im vorgeschlagenen Design finden.

Manipulationen zur Beeinflussung der relationalen Enkodierung. Das Ziel einer Enkodierungsmanipulation in diesem Experiment ist es, die Bildung von Assoziationen zu minimieren bzw. zu maximieren. Strategien zur Maximierung einer relationalen Verarbeitung wurden bereits im Kapitel 3.2.2 sowie im Kapitel 3.2.3 angesprochen. Als hilfreich für eine relationale Enkodierung von Wortpaaren wurden dabei folgende Manipulationen diskutiert: Verwendung von Komposita oder hoch assoziierten Wörtern; Definition eines neuen Kompositums, das aus beiden Wörtern besteht; (interaktive) Vorstellung der beiden Wörter sowie Einbindung der beiden Wörter in einen Satz.

Nicht alle diese Strategien können in diesem Experiment genutzt werden, da dasselbe abstrakte, unrelatierte Wortmaterial verwendet werden soll wie im ersten Experiment. Eine *geeignete* Strategie ist u.a. die Integration der beiden Wörter in einen Satz, denn diese kann problemlos auf abstrakte, unrelatierte Wörter angewandt werden. Weiterhin hat die Manipulation der Satzintegration den Vorteil, dass diese auf den klassischen Ansatz der Verarbeitungstiefe zurückgeht (‘levels of processing‘, Craik & Lockhart, 1972; Craik & Tulving, 1975) und daher zum einen ausgiebig untersucht wurde, zum anderen auch mit anderen Enkodierungsstrategien kontrastiert wurde, die eine relationale Verarbeitung eher minimieren. Es liegen also bereits Daten darüber vor, welche flachen Enkodierungsstrategien im Vergleich zu der tiefen Strategie der Satzenkodierung einen möglichst großen Effekt erzeugen. Zwar werden in den meisten Studien zur Verarbeitungstiefe keine Wortpaare sondern nur Einzelworte verwendet, die Strategien wurden aber vereinzelt auf Wortpaare übertragen (Goshen-Gottstein & Moscovitch, 1995). Im Folgenden wird ein kurzer Überblick über flache und tiefe Verarbeitungsstrategien gegeben, danach wird eine geeignete flache Strategie als Kontrast zur Satzenkodierung vorgeschlagen.

Aus den klassischen Studien zum Ansatz der Verarbeitungstiefe (Craik & Lockhart, 1972; Craik & Tulving, 1975; für einen Überblick siehe Craik, 2002) kann man entnehmen, dass eine flache Verarbeitung resultiert, wenn Wörter auf graphemische Merkmale hin beurteilt werden müssen (z.B. Groß-Kleinschreibung; Schriftart) oder wenn eine Analyse des Wortes nach oberflächlichen Merkmalen wie Vokal- Silben- oder Buchstabenanzahl erfolgt (Challis, Velichovsky & Craik, 1996; Craik & Tulving, 1975). Eine tiefe Verarbeitung kann durch semantische Aufgaben gefördert werden, z.B. indem die Passung eines Wortes in einen Satz beurteilt werden muss (Craik & Tulving, 1975; Glover, Rankin, Langner, Toderò & Dinnel, 1985; Hyde & Jenkins, 1973) oder indem Kategorieurteile abgegeben werden müssen (Craik & Tulving, 1975). Manipulationen wie z.B. Reimurteile über das zu verarbeitende Wort führen zu einer mittleren Verarbeitungstiefe (Morris, Bransford & Franks, 1977). Die genannten Manipulationen wurden 1975 von Craik und Tulving untersucht und danach in zahlreichen Studien repliziert (z.B. Gardiner, 1988; Hyde & Jenkins, 1973; Jacoby, Shimizu, Daniels & Rhodes, 2005; Mulligan & Hirshman, 1997; Russo, Cullis & Parkin, 1998). Auch entstanden weitere Manipulationen, die erfolgreich eine tiefe Verarbeitung induzierten: (1) indem Wörter als angenehm oder eher unangenehm eingestuft werden sollen (Jacoby et al., 2005; Toth, 1996), (2) indem Wörter als zutreffend oder unzutreffend auf die eigene Person beurteilt werden müssen (Challis et al., 1996) oder (3) indem Assoziationen zu dem Zielwort gebildet werden sollen (Java, 1992; Java, Gregg & Gardiner, 1997). All die genannten Manipulationen wurden in zahlreichen experimentellen Studien erfolgreich eingesetzt, um die Verarbeitungstiefe zu beeinflussen (z.B. Cleary, 2002; Gardiner, 1988; Habib & Nyberg, 1997; Hyde & Jenkins, 1973; Jacoby et al., 2005; Mulligan & Hirshman, 1997; Rajaram, 1993; Russo et al., 1998; Yonelinas, Dobbins, Szymanski, Dhaliwal & King, 1996).

Während in den meisten Studien nur Einzelwörter dargeboten wurden, sollen die Manipulationen in der vorliegenden Studie auf Wortpaare angewendet werden. Weiterhin sollte die tiefe Enkodierungsbedingung nicht nur eine tiefe Verarbeitung der Einzelwörter, sondern zugleich auch eine *relationale* Verarbeitung fördern. Wie bereits zu Beginn des Kapitels diskutiert wurde, kann dies durch die Integration der Wörter in Sätze erreicht werden. Diese Manipulation wurde von Goshen-Gottstein und Moscovitch (1995) genutzt, die ebenfalls Wortpaare verwendeten (siehe auch Prior & Bentin, 2003). Bei Goshen-Gottstein und Moscovitch (1995) wurden den Probanden Wortpaare für fünf Sekunden präsentiert und sie sollten in dieser Zeit einen Satz generieren, der beide Wörter enthält. Falls fünf Sekunden für diese Aufgabe nicht ausreichend waren, konnten sich die Probanden auch nach der Darbietung beliebig viel Zeit für die Generierung des Satzes lassen. Die Probanden konnten also in ihrem eigenen Tempo die Aufgabe durchführen. Die Generierung von Sätzen kann zwar zu einer besonders hohen Erinnerungsleistung an die Sätze inklusive der Wörter führen (Slamecka & Graf, 1978). Sie birgt aber auch die Gefahr, dass es nicht allen Probanden gelingt, einen Satz zu bilden, dass die generierten Sätze eine sehr unterschiedliche Komplexität aufweisen und dass es zu einer hohen Varianz in den Lernzeiten kommt. Daher werden im vorliegenden Experiment die Lernzeiten konstant gehalten, und die Probanden haben nicht die Aufgabe, einen Satz zu generieren, sondern sie bekommen einen Satz vorgegeben und sollen beurteilen, ob sie sich diesen Satz gut oder schlecht vorstellen können. Dadurch ist sichergestellt, dass alle Probanden die Wörter in einem sinnvollen, vergleichbar komplexen Kontext enkodieren und dass für alle Probanden eine konstante Lernzeit zur Verfügung steht. Weiterhin wird durch die Aufforderung zur Vorstellung des Satzes eine duale Kodierung angeregt; eine Manipulation, die ebenfalls zu einer tiefen Enkodierung und erhöhten Erinnerungsleistung führt (siehe Kapitel 3.2.2).

Als Kontrast zu dieser tiefen Enkodierungsbedingung können viele der flachen Enkodierungsbedingungen so modifiziert werden, dass sie auf Wortpaare anwendbar sind. So kann man z.B. die Probanden beurteilen lassen, ob die beiden Wörter in der gleichen Schriftart dargeboten werden, ob die gleiche Anzahl von Silben bzw. Vokalen in beiden Wörtern vorliegt oder ob die Wörter gleich viele Buchstaben aufweisen. Auf all diese Fragen kann - genau wie in den klassischen Manipulationen - ein Ja-nein-Urteil abgegeben werden. In Anlehnung an die Studie von Goshen-Gottstein und Moscovitch (1995) wurde für dieses Experiment der Vergleich der Vokalanzahl ausgewählt. Die Teilnehmer mussten also so schnell wie möglich beurteilen, ob die beiden Wörter gleich viele Vokale enthalten oder ob sie unterschiedlich viele Vokale aufweisen.

Dass der Vergleich dieser beiden Enkodierungsbedingungen (Vokalvergleich und Satzbildung) auf Wortpaare anwendbar ist und zu einem signifikanten Unterschied in der Rekognitionsleistung führt, konnte durch Goshen-Gottstein und Moscovitch (1995) nachgewiesen werden.

Nachdem die Enkodierungsmanipulation festgelegt wurde, bleiben noch zwei weitere Fragen zur konkreten Operationalisierung offen. Zum einen kann die Manipulation der Verarbeitungstiefe innerhalb oder zwischen den Teilnehmern umgesetzt werden; zum anderen könnte sowohl eine inzidentelle als auch eine intentionale Lernsituation genutzt werden. Da für das vorliegende Experiment ein möglichst starker Effekt erzielt werden sollte, orientiert sich die Wahl des Designs und der Lernsituation an diesem Kriterium.

Hinsichtlich des Designs zeigte sich (Thapar & Greene, 1994), dass eine geblockte Darbietung stärkere Effekte erzielt als eine randomisierte Darbietung. Allerdings weist auch die geblockte Darbietung innerhalb der Teilnehmer den Nachteil auf, dass eine Beeinflussung („Carry-over-Effekt“) der zweiten Lernbedingung durch die erste nicht ausgeschlossen werden kann. Um diesen Effekt zu kontrollieren und einen möglichst starken Effekt zu erreichen, wurde die Verarbeitungstiefe zwischen den Versuchspersonen manipuliert. In Bezug auf die Lernbedingungen, gibt es Studien die zeigen, dass eine intentionale Verarbeitung zu stärkeren Effekten führt als eine inzidentelle (Nyberg, Olofsson, Gardiner & Nilsson, 1997, siehe aber auch Hyde & Jenkins, 1973). Weiterhin reduziert eine intentionale Lernbedingung die Heterogenität zwischen den Probanden, falls manche Teilnehmer sich trotz inzidenteller Lernbedingungen die Wörter absichtlich merken.

Im Sinne eines maximalen experimentellen Effekts und einer Vermeidung von konfundierenden Störeffekten wurde die Verarbeitungstiefe zwischen den Teilnehmern manipuliert und die Probanden wurden darüber informiert, dass die Wörter in einem späteren Gedächtnistest abgefragt werden.

Manipulationen zur Beeinflussung des Abrufs von Assoziationen. Auch zur Beeinflussung des Abrufs von Informationen kann auf klassische Studien der Gedächtnispsychologie zurückgegriffen werden. Ein Faktor, der die Erinnerungswahrscheinlichkeit herabsetzt, ist das Retentionsintervall, also die Zeit zwischen Lern- und Testphase. Schon Ebbinghaus konnte mit seiner klassischen Vergessenskurve aufzeigen, dass ein Zusammenhang zwischen Retentionsintervall und Erinnerungsleistung besteht (Ebbinghaus, 1885). Um ein geeignetes Retentionsintervall für dieses Experiment festzulegen, sind zum einen Studien interessant, die sich mit der Vergessenskurve von Wortpaaren beschäftigen, zum anderen Studien, die das Retentionsintervall in Kombination mit einer Manipulation der Verarbeitungstiefe variiert haben.

Hockley (1991, 1992) untersuchte die Vergessenskurve der Item- sowie assoziativen Rekognition in Abhängigkeit von unterschiedlichen Retentionsintervallen. Für das vorliegende Experiment ist besonders die Studie von Hockley und Consoli (1999) interessant, da in ihrem ersten Experiment ein sofortiger Test mit einem Test nach 30 Minuten und nach 24 Stunden verglichen wurde. Über alle drei Messzeitpunkte hinweg war die Erinnerungsleistung an Wortpaare besser als die Erinnerungsleistung an einzelne Wörter. Zwischen den Messzeitpunkten nahm die Rekognitionsleistung jeweils ab, wobei der Ver-

lauf für Wortpaare und Einzelitems identisch war. Es trat also keine Interaktion zwischen Messzeitpunkt und Itemart auf. Obwohl die Teilnehmer eine relativ lange Wortliste von 120 Wortpaaren lernen mussten, sank ihre Erinnerungsleistung (gemessen über das Diskriminationsmaß A') für die Wortpaare von $A' = 0.90$ im sofortigen Test lediglich auf $A' = 0.87$ im Test nach 30 Minuten und auf $A' = 0.77$ nach 24 Stunden. Ein Retentionsintervall von 24 Stunden zeigt zwar einen deutlichen Effekt auf die Gedächtnisleistung, führt aber keineswegs zu Bodeneffekten. Die Probanden in der Studie von Hockley und Consoli (1999) wurden auf zwei hilfreiche Enkodierungsstrategien hingewiesen. Zum einen wurde ihnen vorgeschlagen, die Wörter in sinnvolle Sätze einzubinden, zum anderen wurde ihnen nahegelegt, sich die Worte in einem interaktiven Bild vorzustellen. Inwieweit die Strategien genutzt wurden, wurde nicht erfasst. Es ist also durchaus vorstellbar, dass bei einer konsequenten Anwendung der Strategien die Erinnerungsleistung noch besser ausgefallen wäre.

Die Anwendung einer tiefen vs. flachen Verarbeitung versuchten Morris et al. (1977) in ihrem Experiment sicherzustellen. In der tiefen semantischen Bedingung mussten die Probanden beurteilen, ob das Zielwort sinnvoll in einen vorgegebenen Satz passt; in der flachen Reimbedingung mussten die Teilnehmer angeben, ob sich das Zielwort und ein vorgegebenes Vergleichswort reimen. Neben der Verarbeitungstiefe manipulierten sie auch den Testzeitpunkt (direkt nach der Lernphase vs. 24 Stunden später). Sowohl im sofortigen Test als auch nach einem Tag war das Erinnerungsniveau in der Satzbedingung *deskriptiv* höher als in der Reimbedingung. Allerdings war dieser Unterschied nur in der sofortigen Bedingung signifikant, im Test nach 24 Stunden wurde kein signifikanter Unterschied mehr festgestellt. Ob dies auf eine zu schwache experimentelle Manipulation hinweist oder aber anzeigt, dass die Effekte der Verarbeitungstiefe sich über die Zeit abschwächen, wurde von den Autoren nicht weiter untersucht.

Aufschluss darüber gibt die Studie von Bentin, Moscovitch und Nirhod (1998). Unter einer etwas anderen Zielsetzung verglichen sie eine tiefe und flache Verarbeitungsbedingung im sofortigen Test und nach 24 Stunden. Im Vergleich zu Morris et al. (1977) verwendeten Bentin et al. (1998) eine stärkere experimentelle Manipulation der Verarbeitungstiefe. In ihrer flachen Enkodierungsbedingung mussten die Probanden so schnell wie möglich angeben, ob das Zielwort in roter Farbe oder in blauer Farbe gedruckt war. In der tiefen Bedingung mussten die Teilnehmer - wieder unter Zeitdruck - entscheiden, ob es sich bei einer Buchstabenkette um ein Wort oder Nonwort handelt. Diese lexikalische Entscheidungsaufgabe wird im deutsch- oder englischsprachigen Raum eher als flache Enkodierungsbedingung verwendet, in der hebräischen Sprache allerdings stellt eine solche Entscheidung höhere Anforderungen und wurde von den Autoren aufgrund der sprachlichen Besonderheiten als tiefe Verarbeitungsbedingung eingestuft. Ein signifikanter Unterschied in der Erinnerungsleistung flach vs. tief verarbeiteter Wörter findet sich sowohl im sofortigen Test als auch nach einem Tag. Das Retentionsintervall setzt in beiden Verar-

beitungsbedingungen die Erinnerungsleistung signifikant herab. Der Unterschied zwischen den beiden Enkodierungsbedingungen ist übereinstimmend mit Morris et al. (1977) im sofortigen Test größer als nach 24 Stunden.

Zusammenfassend lässt sich sagen, dass bei Hockley und Consoli (1999) auch nach 24 Stunden eine relativ gute Erinnerungsleistung an Wortpaare vorhanden war. Unter der Annahme, dass die angestrebte Satzmanipulation eine tiefe relationale Verarbeitung der Wörter als Assoziation fördert, scheint ein Retentionsintervall von 24 Stunden daher angebracht. Kürzere Retentionsintervalle könnten einen zu geringen Effekt hervorrufen oder aufgrund von Deckeneffekten scheitern.

Aus den Studien von Morris et al. (1977) sowie Bentin et al. (1998) kann man folgern, dass die Effekte der Verarbeitungstiefe sich sowohl in einer sofortigen Testung als auch in einem Test nach einem Tag zeigen. Zwar war der Effekt nach 24 Stunden schwächer, aber immer noch vorhanden. In der flachen Enkodierungsbedingung war die Leistung nach einem Tag bereits relativ schlecht, so dass längere Retentionsintervalle evtl. zu Bodeneffekten in der flachen Enkodierungsbedingung führen können.

Hinsichtlich des Designs sollte das Retentionsintervall innerhalb der Versuchspersonen manipuliert werden. Da eine intentionale Lernbedingung erfolgen sollte, wäre es inkonsistent mit den Instruktionen, keinen Gedächtnistest zu erheben. Außerdem besteht die Gefahr, dass die Probanden, die keinen sofortigen Gedächtnistest erhalten, vermuten, dass dieser am nächsten Tag stattfindet. Dies könnte dazu führen, dass einige Teilnehmer sich die Wörter notieren oder versuchen, diese zu wiederholen. Bei einer Manipulation innerhalb der Versuchspersonen werden 50% der Wörter sofort getestet, so dass die Probanden mit einer hohen Wahrscheinlichkeit „naiv“ an dem zweiten Gedächtnistest teilnehmen.

7.1 Methode

In Experiment 2 sollen die zentralen Parameter der Assoziationsenkodierung (c) und des Assoziationsabrufs (r) validiert werden. Da eine Beeinflussung der Enkodierungsleistung möglicherweise auch Auswirkungen auf den Abruf von Assoziationen haben kann, werden in diesem Experiment Assoziationsenkodierung und Assoziationsabruf gleichzeitig über das Enkodierungs-Abruf-Paradigma (Tulving, 1983) manipuliert. Dieses sieht vor, dass eine mindestens zweistufige Enkodierungsmanipulation mit einer ebenfalls mindestens zweistufigen Abrufmanipulation gekreuzt wird. Dadurch können nicht nur erwünschte Haupteffekte, sondern auch Interaktionen erfasst werden.

Konkret wird in diesem Experiment die Assoziationsenkodierung gefördert, indem die beiden Wörter in einem sinnvollen Satz dargeboten werden. Zusätzlich zu dem Satzkontext wird auch eine visuelle Enkodierung angeregt, denn die Probanden sollen angeben, ob sie sich den Satz gut oder schlecht vorstellen können. Für diese ‚Satzbedingung‘ sollte der assoziative Enkodierungsparameter signifikant höher ausfallen als in der ‚Vokalbedin-

gung'. In der Vokalbedingung sehen die Probanden zwei Wörter und müssen so schnell wie möglich entscheiden, ob beide Wörter dieselbe Anzahl von Vokalen enthalten. Diese oberflächliche Aufgabe fördert keine relationale Enkodierung des Wortpaares und sollte daher eine niedrige Schätzung im Enkodierungsparameter für Assoziationen zur Folge haben.

Die relationale Enkodierung sollte sich nicht nur auf den assoziativen Enkodierungsparameter auswirken, sondern auch auf den Parameter zur Distraktorentdeckung von Assoziationen, a . Wird zum Beispiel eine Target-Distraktor-Kombination getestet, so sollte es Probanden aus der Satzbedingung leichter fallen, den Satz, in den das alte Wort eingebettet war, zu erinnern und damit die Kombination als eine neue Wortkombination zu identifizieren ('recall to reject'). In der Satzbedingung sollte a daher signifikant höher ausfallen als in der Vokalbedingung.

Die Manipulation des assoziativen Abrufs wird über das Retentionsintervall operationalisiert. Die Hälfte der Wörter wird direkt (d.h. drei Minuten) nach der Lernphase getestet, die andere Hälfte der Wörter erst 24 Stunden später. Für das kurze Retentionsintervall sollte eine signifikant höhere Parameterschätzung des Abrufparameters resultieren als für das lange Retentionsintervall.

Sowohl das Retentionsintervall als auch die Verarbeitungstiefe werden die Hybridparameter der Itemerkennung (w_1 , w_2) beeinflussen. Die Rateparameter und der Parameter der Distraktorentdeckung D sollen sich zwischen den Gruppen gleichsetzen lassen.

Teilnehmer

An diesem Experiment nahmen 87 Studierende der Universität Mannheim teil. Das durchschnittliche Alter der Probanden lag bei 23.64 Jahren ($SD = 2.5$; Spannweite: 18-27 Jahre).

Material

In diesem Experiment wurde die Wortliste 2 verwendet. Um die Sätze zu generieren, wurden die Wörter zufällig zu 140 Wortpaaren zusammengestellt. Zu diesen Wortpaaren formulierten drei Personen unabhängig voneinander Sätze. Die Sätze sollten eine sinnvolle Aussage bilden und neben den beiden Zielwörtern keine weiteren Substantive enthalten. Aus den Satzalternativen für jedes Wortpaar wurde jeweils ein Satz ausgewählt; die Kriterien waren dabei Sinngehalt sowie interaktive Verbindung der beiden Wörter. Aus den 140 gebildeten Sätzen wurden 50 Sätze ausgeschlossen, da für diese Wortkombinationen keine sinnvollen Satzkonstruktionen möglich waren. Insgesamt bestand das Wortmaterial also aus 90 Sätzen. In Tabelle 7.1 sind beispielhaft einige Sätze dargestellt, die gesamte Satzliste kann Anhang D entnommen werden. Für den Vokalvergleich wurden dieselben randomisiert zusammengestellten Wortpaare verwendet.

Tabelle 7.1: Satzbeispiele aus Experiment 2

Linkes Wort	Rechtes Wort	Verbindender Satz
ABEND	PRESSE	Am ABEND hat die PRESSE bereits alles erfahren.
ABSICHT	FRIEDEN	Ihre ABSICHT ist es, FRIEDEN zu stiften.
BLECH	STAUB	Das BLECH ist dick mit STAUB bedeckt.
GRENZE	MÖNCH	An der GRENZE zögert der fliehende MÖNCH kurz.
LÄNGE	REGIE	Die endgültige LÄNGE bestimmt die REGIE.
POSTEN	LEICHE	Als er seinen POSTEN bezieht, entdeckt er die LEICHE.
SITZUNG	ZEUGNIS	In der SITZUNG wird ihr ZEUGNIS besprochen.
TRUPPE	NEBEL	Die TRUPPE schleicht vorsichtig durch den NEBEL.
ZIMMER	WOCHE	Sie mietet das ZIMMER für eine ganze WOCHE.

Aus den 90 Sätzen bzw. Wortpaaren wurden pro Proband 50 für die Lernphase zufällig ausgewählt, weitere 40 Einzelwörter wurden zufällig gezogen und als Distraktoren verwendet. Pro Versuchsperson wurden also von den 180 Wörtern nur 140 genutzt.

Vortest

In Experiment 1 wurden 50 Wortpaare für vier Sekunden präsentiert. Mit dieser Kombination wurde eine gute Erinnerungsleistung erreicht. Für Experiment 2 wurden zunächst Vortests durchgeführt um herauszufinden, ob vier Sekunden ausreichen, um die Sätze zu lesen und ein Urteil abzugeben. Für den Vokalabgleich sollte im Vortest sichergestellt werden, dass den Probanden nach ihrer Antwort nicht zu viel zusätzliche Zeit zur Verfügung steht, um elaborierte Lernstrategien anzuwenden.

Die Vortests ergaben, dass vier Sekunden für das Satzurteil etwas zu kurz waren, für den Vokalabgleich aber zu lang. Daher wurde die Präsentationszeit der Sätze auf fünf Sekunden festgelegt, die Präsentationszeit für die Wortpaare auf 3.5 Sekunden. Diese Zeiten wurden in einem erneuten Vortest nochmals validiert.

Ablauf

Zunächst unterzeichneten die Teilnehmer die Einverständniserklärung, in der sie über ihre Rechte aufgeklärt wurden. Um den Datensatz des ersten Tages und den Datensatz des zweiten Tages derselben Versuchsperson zuordnen zu können, wurde eine Teilnehmerliste geführt, in der der Name des Teilnehmers mit der zugehörigen Versuchspersonennummer aufgelistet war.³⁰ Diese Liste wurde nach Beendigung des Experiments vernichtet. Die Teilnehmer wurden in der Einverständniserklärung über dieses Verfahren informiert und hatten die Möglichkeit, die Teilnahme unter diesen Umständen zu verweigern.

³⁰Da in die multinomiale Analyse nur aggregierte Daten eingehen, ist es dafür nicht notwendig, die Datensätze einer Person zuordnen zu können. Allerdings sollten die Daten potentiell auch für andere Analysen wie z.B. eine Varianzanalyse verwendet werden können, und für diese Analysen ist es erforderlich, die beiden Datensätze einem Probanden zuordnen zu können.

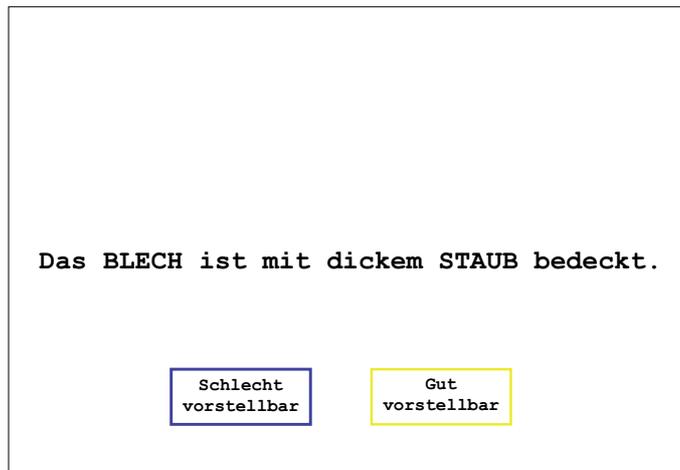


Abbildung 7.3: In dieser Abbildung ist dargestellt, wie die tiefe Enkodierungsbedingung umgesetzt wurde. Die gesamte Satzliste ist im Anhang D abgedruckt, einige Beispielsätze sind in Tabelle 7.1 wiedergegeben. Die Probanden mussten für jeden Satz beurteilen, ob sie sich den Inhalt gut oder schlecht vorstellen können. Über farblich korrespondierende Tasten konnte die Antwort eingegeben werden.

Nach der Angabe von demographischen Daten wie Alter und Studienfach wurden sie darüber in Kenntnis gesetzt, dass das folgende Experiment aus einer Lernphase besteht, in der sie Aufgaben zu Wortpaaren bearbeiten sollen, und aus einer Testphase, in der die Wortpaare erinnert werden müssen. Dann wurde ihnen abhängig von ihrer Bedingung die konkrete Aufgabe erklärt.

In der *Satzbedingung* wurde den Teilnehmern erläutert, dass ihnen Sätze dargeboten werden und sie beurteilen sollen, ob sie sich die Sätze bildlich gut vorstellen können. Weiterhin wurden die Teilnehmer darauf hingewiesen, dass sie sich die beiden großgeschriebenen Wörter jedes Satzes merken sollen, da ein Gedächtnistest für diese Wörter folgt. Die Beurteilung der Sätze wurde aber als wichtigere Aufgabe betont. Nach diesen Instruktionen folgte ein Übungsdurchgang in dem 6 Beispielsätze für je fünf Sekunden gezeigt wurden. Unter dem Satz waren die beiden Antwortoptionen „gut vorstellbar“ sowie „schlecht vorstellbar“ abgebildet. Über farblich korrespondierende Tasten konnten die Antworten eingegeben werden (siehe Abbildung 7.3). Die Urteile mussten innerhalb von fünf Sekunden erfolgen. Wurde diese Zeitspanne überschritten, so kam der Hinweis, beim nächsten Satz bitte schneller zu antworten. Wurde das Urteil nach z.B. zwei Sekunden abgegeben, so verschwanden die Antwortoptionen, der Satz wurde aber weiterhin dargeboten. Alle Teilnehmer konnten den Satz also fünf Sekunden lang enkodieren, unabhängig davon, wie schnell sie antworteten. In diesem Experiment wurden die Übungsitens nicht getestet, da der Fokus auf der Satzbeurteilung liegen sollte und nicht auf der Gedächtnisaufgabe.

In der *Vokalbedingung* wurde den Teilnehmern erklärt, dass sie im Folgenden zwei Wörter sehen und so schnell wie möglich beurteilen sollen, ob die beiden Wörter dieselbe



Abbildung 7.4: In dieser Grafik ist ein Beispiel für die flache Enkodierungsbedingung dargestellt. Die Probanden mussten für jedes Wortpaar entscheiden, ob beide Wörter dieselbe Anzahl an Vokalen enthalten oder nicht. Für diese Entscheidung standen ihnen 3.5 Sekunden zur Verfügung. Die Antwort konnte über farblich korrespondierende Tasten abgegeben werden.

Anzahl an Vokalen enthalten oder nicht. Zu diesem Zweck waren unter den Wörtern die beiden Antwortoptionen „gleiche Vokalanzahl“ sowie „ungleiche Vokalanzahl“ abgebildet (siehe Abbildung 7.4). Über farblich korrespondierende Tasten konnten die Urteile abgegeben werden. Die Wörter wurden für 3.5 Sekunden gezeigt, innerhalb dieser Zeitspanne musste die Vokalantwort gegeben werden. Falls die Probanden für ihre Eingabe länger benötigten, wurden sie aufgefordert, beim nächsten Mal schneller zu reagieren. Falls die Teilnehmer ihre Antwort z.B. nach 1.5 Sekunden abgegeben hatten, wurden die Antwortkästen ausgeblendet, die Wörter waren aber für insgesamt 3.5 Sekunden zu sehen. Die Teilnehmer wurden auch darauf hingewiesen, dass sie in einem Gedächtnistest nach den beiden Wörtern gefragt werden würden. Nach der Präsentation von 6 Übungssitem hatten die Teilnehmer Gelegenheit, Fragen zu stellen, und sie erhielten Rückmeldung über den Prozentsatz der richtig beurteilten Vokalvergleiche. Auch hier wurden die Übungssitem nicht getestet, um den Vokalvergleich als Hauptaufgabe hervorzuheben, nicht den Gedächtnistest.

Die Antworten wurden in beiden Gruppen mit dem sequentiellen Ja-nein-Format erfasst. Die Beschriftung der Tasten (gut/schlecht vorstellbar; gleiche/ungleiche Vokalanzahl sowie alt/neu) wurde ausbalanciert, um Händigkeitseffekte auszugleichen. Vor jeder Satz- bzw. Wortpaardarbietung wurde für 750 ms ein Fixationskreuz an der Stelle gezeigt, an der der Satz bzw. das Wortpaar erschien.

Auf den Übungsdurchgang folgte die Lernphase, in der 50 Wortpaare entweder in Satzform oder mit der Aufgabe, die Vokale zu beurteilen, präsentiert wurden. Wie in der Übungsphase wurden die Sätze für fünf Sekunden gezeigt. Der Vokalabgleich musste

innerhalb von 3.5 Sekunden erfolgen. Drei Füllersätze bzw. Füllerwortpaare gingen der eigentlichen Lernliste voraus. Im Anschluss an die Testphase folgte für drei Minuten eine Distraktoraufgabe. Danach wurden 50% der Wortpaare getestet. Die Wortpaare wurden in der Testphase randomisiert zu Kombinationen von (1) alten intakten Wortpaaren, (2) neu kombinierten Paaren alter Wörter, (3) Target-Distraktor-Kombinationen, (4) Distraktor-Target-Kombinationen sowie (5) Distraktor-Distraktor-Kombinationen zusammengestellt. Die Anzahl der jeweiligen Kombinationen wurde gleich auf die beiden Testphasen verteilt. Sowohl nach drei Minuten als auch nach 24 Stunden wurden also 10 alte intakte sowie 10 neu kombinierte Paare dargeboten. Target-Distraktor-Kombinationen, Distraktor-Target-Kombinationen sowie zwei Distraktoren wurden jeweils 5 mal getestet. Die Wortkombinationen wurden innerhalb einer Testbedingung randomisiert präsentiert.

Nach Beendigung des ersten Testdurchgangs wurden die Teilnehmer an ihren zweiten Termin am nächsten Tag erinnert. Es wurde ihnen gesagt, dass sie an diesem Termin erneut Aufgaben zu anderen Wortpaaren bearbeiten sollen. Die Teilnehmer wussten also nicht, dass am nächsten Tag die restlichen Wörter erinnert werden mussten. Am zweiten Tag wurden die verbleibenden Wortpaare getestet.

Design

Diesem Experiment liegt ein 2*2 gemischt faktorielles Design zugrunde. Der Faktor Enkodierungsbedingung (tief vs. flach) wurde zwischen den Teilnehmern variiert, das Retentionsintervall (3 Minuten vs. 24 Stunden) innerhalb der Teilnehmer.

7.2 Ergebnisse

Für Experiment 2 soll zunächst eine Überprüfung der Manipulation vorgenommen werden, um sicherzustellen, dass diese auch bei allen Probanden erfolgreich umgesetzt wurde. Dazu wurden die Antworten auf die Vokalfrage bzw. die Frage zur Vorstellbarkeit der Sätze analysiert.

Zunächst wurde der Anteil fehlender Antworten pro Versuchsperson berechnet. Hierbei fiel ein Proband auf, der in der tiefen Verarbeitungsbedingung zu keinem Satz eine Beurteilung abgegeben hatte. Dieser Proband wurde daher aus der Analyse ausgeschlossen. Zwei weitere Teilnehmer in der Vokalbedingung wiesen 68% bzw. 76% fehlende Antworten auf und hoben sich durch diese hohe Rate von über 50% von den restlichen Versuchsteilnehmern ab. Daher wurden auch die Daten dieser beiden Probanden entfernt, da unklar war, inwieweit für diese von einer flachen Enkodierungsbedingung ausgegangen werden kann.

In der Vokalbedingung wurde weiter analysiert, wie hoch der Anteil korrekt beurteilter Vokalvergleiche war. Auch hier mussten drei Ausreißer festgestellt werden, die sich mit weniger als 60% korrekter Antworten deutlich von den restlichen Probanden hervorhoben.

Auch diese wurden aus der Analyse ausgeschlossen. Nach Ausschluss der Probanden verblieben 87 Teilnehmer in der Stichprobe. Diese Anzahl wurde bei der Stichprobenbeschreibung berichtet. Diese Teilnehmer wiesen mit durchschnittlich 83% korrekten Antworten eine gute Leistung im Vokalvergleich auf. In der Satzbedingung wurden 58% der Sätze als gut vorstellbar beurteilt.

Nachdem sichergestellt wurde, dass alle in die Analyse eingehenden Versuchsteilnehmer die Aufgaben gewissenhaft bearbeitet hatten, wurden zunächst die relativen Häufigkeiten berechnet, dann die modellbasierte Auswertung vorgenommen. Um die folgende Auswertung übersichtlicher zu gestalten, wird diese in drei Teile aufgeteilt: (1) Im ersten Teil wird die tiefe Verarbeitungsbedingung mit der flachen Verarbeitungsbedingung verglichen, wobei jeweils nur die Testphase nach drei Minuten analysiert wird. Im Vergleich dieser Bedingungen sollten die Effekte unterschiedlich tiefer Enkodierung festgestellt werden.

(2) Im zweiten Teil wird für die tiefe Verarbeitungsbedingung die Leistung nach drei Minuten und nach 24 Stunden analysiert, um die Auswirkungen des Retentionsintervalls zu erfassen. (3) Um schließlich sowohl die Effekte konvergenter als auch diskriminanter Validität sicherzustellen und die Vorteile des gewählten Vier-Gruppen-Designs zu nutzen, werden in einem dritten Auswertungsteil die zentralen Parameter aller vier Bedingungen einander gegenübergestellt.

(1) Tiefe vs. flache Enkodierungsbedingung

Sowohl für die flache als auch für die tiefe Enkodierungsbedingung wurden die relativen Häufigkeiten ausgewertet (siehe Tabelle 7.2). Deskriptiv fällt bereits hier auf, dass die Raten korrekter Antworten in der flachen Verarbeitungsbedingung jeweils niedriger ausfallen als in der tiefen Verarbeitungsbedingung (Werte der flachen vs. tiefen Bedingung: $.335 < .661$; $.388 < .533$; $.340 < .548$; $.391 < .522$).

Ob die schlechtere Leistung der flachen Enkodierungsbedingung sich auch signifikant in einem niedrigeren Enkodierungsparameter zeigt, wurde in der multinomialen Analyse ausgewertet.

Multinomiale Analyse

Eine gute Modellanpassungsstatistik lag für die gemeinsame Auswertung aller vier Gruppen vor, der G^2 -Wert lag mit $G^2_{(36)} = 48.48$ unter dem kritischen Wert $G^2_{krit, \alpha=.001} = 67.98$. Die lokale Identifizierbarkeit war in diesem Experiment wieder gegeben, sowohl der Rang der Jacobischen Matrix als auch die identischen Parameterschätzungen bei mehrfacher Aktivierung des EM-Algorithmus mit zufälligen Startwerten konnten dies zeigen.

Aus Abbildung 7.5 wird ersichtlich, dass der Enkodierungsparameter in der flachen Verarbeitungsbedingung mit $c = 0.14$ deutlich geringer ausfällt als in der tiefen relationalen Verarbeitungsbedingung, $c = 0.66$. Die Gleichsetzung der beiden Parameter führt

Tabelle 7.2: *Relative Anwothhäufigkeiten auf flach enkodierte Wörter (a) sowie tief enkodierte Wortmaterial (b)*

(a) Vokalvergleich, Retentionsintervall 3 min

Wahre Itemkat.	Antwort der Versuchsperson				
	Altes WP	Neu komb. WP	Target Dis.	Dis. Target	Dis.Dis.
Altes WP	.335	.223	.156	.198	.088
Neu komb. WP	.074	.388	.230	.202	.105
Target Dis.	.051	.237	.340	.163	.209
Dis. Target	.065	.214	.102	.391	.228
Dis. Dis.	.023	.195	.181	.270	.330

(b) Satzenkodierung, Retentionsintervall 3 min

Wahre Itemkat.	Antwort der Versuchsperson				
	Altes WP	Neu komb. WP	Target Dis.	Dis. Target	Dis.Dis.
Altes WP	.661	.157	.078	.067	.037
Neu komb. WP	.057	.533	.161	.202	.048
Target Dis.	.039	.213	.548	.104	.096
Dis. Target	.035	.209	.087	.522	.148
Dis. Dis.	.043	.117	.239	.213	.387

Tabelle 7.3: *Parameterschätzungen (Par. Schätzungen) sowie 95%-Konfidenzintervalle (95% KI) für die beiden Gruppen der flachen und tiefen Enkodierungsbedingung (Vokalbedingung vs. Satzbedingung)*

Parameter	Vokalbedingung		*	Satzbedingung	
	Par.Schätzung	95% KI		Par.Schätzung	95% KI
c	0.67	0.58 – 0.76	*	0.14	–0.16 – 0.45
r	0.91	0.83 – 1.00		1.00	–0.59 – 2.59
w_1	0.64	0.55 – 0.73	*	0.26	0.15 – 0.37
w_2	0.66	0.57 – 0.74	*	0.34	0.21 – 0.48
g_{10}	0.31	0.18 – 0.44	*	0.61	0.48 – 0.75
g_{11}	0.40	0.23 – 0.58	*	0.58	0.52 – 0.64
g_{20}	0.32	0.18 – 0.46	*	0.89	0.70 – 1.09
g_{21}	0.39	0.22 – 0.56		0.46	0.36 – 0.55
g_3	0.25	0.13 – 0.36		0.40	0.28 – 0.52
D	0.02	–0.34 – 0.38		0.19	0.11 – 0.27
a	0.45	0.22 – 0.69	*	0.00	–0.40 – 0.40

*: Signifikante Effekte zwischen den Gruppen auf einem α -Niveau von .05.

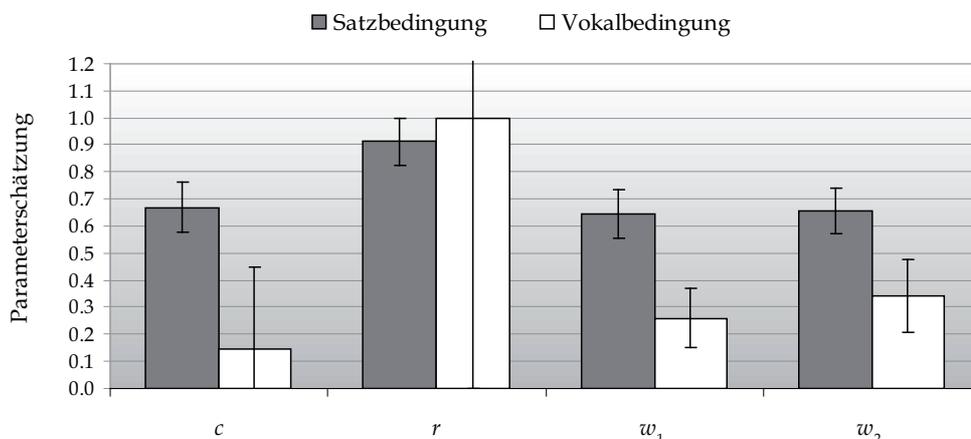


Abbildung 7.5: In dieser Abbildung sind die Parameterschätzungen für die Gedächtnisparameter abgetragen. In Grau sind die Schätzungen der tiefen Enkodierungsbedingung dargestellt, in Weiß die der flachen Enkodierungsbedingung. Alle dargestellten Parameterschätzungen beziehen sich auf den Testzeitpunkt nach drei Minuten. Die Fehlerbalken zeigen die 95%-Konfidenzintervalle an.

mit $\Delta G_{(1)}^2 = 13.54$ zu einer signifikanten Verschlechterung der Modellanpassungsstatistik (siehe auch Tabelle 7.3).

Die Hybridparameter der Itemerkennung können erwartungsgemäß weder für das linke Wort ($\Delta G_{(1)}^2 = 3.95$) noch für das rechte Wort ($\Delta G_{(1)}^2 = 4.38$) gleichgesetzt werden. Allerdings lassen sich die Parameter innerhalb jeder Gruppe gleichsetzen ($w_1 = w_2$). Das dadurch entstehende Submodell 10 kann sowohl in der Satzbedingung ($\Delta G_{(1)}^2 = 0.12$) als auch in der Vokalbedingung ($\Delta G_{(1)}^2 = 0.70$) angepasst werden.

Die Parameter des assoziativen Abrufs unterscheiden sich mit $r_{tief} = 0.91$ und $r_{flach} = 0.99$ weder deskriptiv noch inferenzstatistisch voneinander. Der Abrufparameter zeigt in der flachen Enkodierungsbedingung ein sehr breites Konfidenzintervall. Dies liegt daran, dass der c -Parameter in dieser Bedingung sehr niedrig ausfällt und daher wenig Datenpunkte in die Schätzung des r -Parameters einfließen. Um die diskriminante Validität des r -Parameters trotz der großen Konfidenzintervalle sicherzustellen, werden deshalb im letzten Teil der Analyse die c - und r -Parameter zwischen den vier experimentellen Gruppen verglichen.

Wie in Abbildung 7.6 zu sehen ist, fallen die Rateparameter in der Satzbedingung tendenziell niedriger aus als in der Vokalbedingung. Der paarweise Vergleich fällt für drei der vier Rateparameter auch signifikant aus (g_{10} : $\Delta G_{(1)}^2 = 5.26$; g_{11} : $\Delta G_{(1)}^2 = 6.54$; g_{20} : $\Delta G_{(1)}^2 = 3.90$; g_{21} : $\Delta G_{(1)}^2 = 0.52$). Innerhalb der Vokalbedingung fällt der Rateparameter g_{20} höher aus als die anderen Rateparameter, daher kann hier keine weitere Einschränkung auf spezifischere Submodelle vorgenommen werden. In der Satzbedingung dahingegen führt eine Reduktion zu Submodell 9a zu keiner signifikanten Verschlechterung der

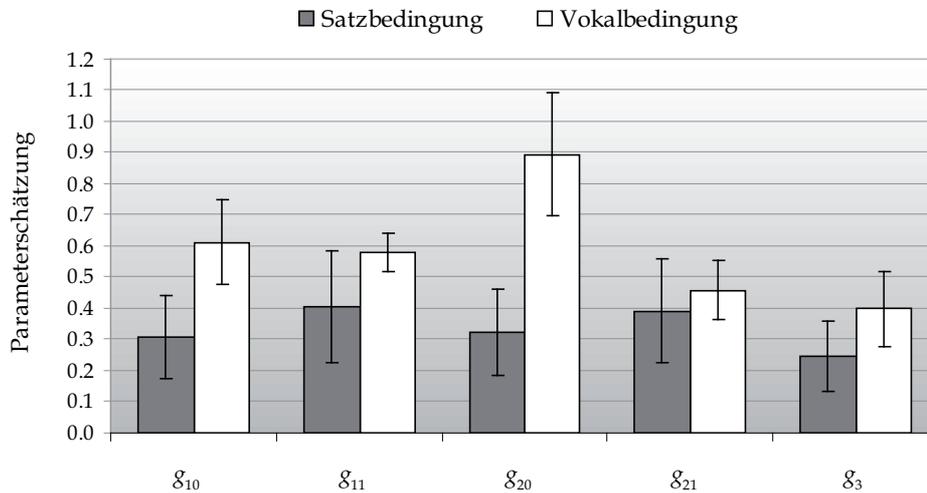


Abbildung 7.6: Hier sind die Schätzungen der Rateparameter des ADH-Modells für die beiden Teilbedingungen der tiefen (in Grau) und flachen (in Weiß) Enkodierung in Experiment 2 abgetragen. Alle Parameterschätzungen sind aus den Teildaten des sofortigen Tests gewonnen. Die Fehlerbalken zeigen die 95%-Konfidenzintervalle an.

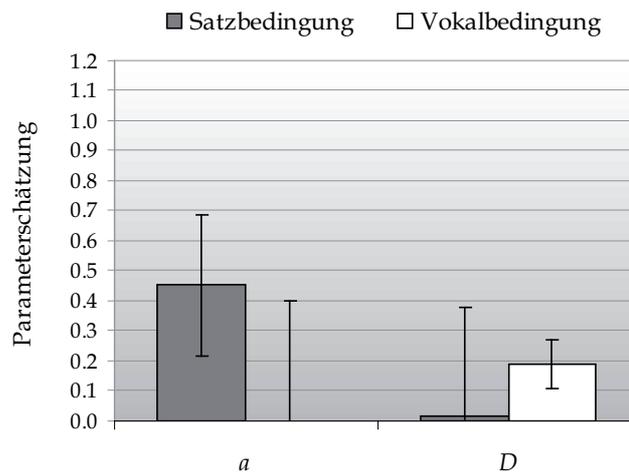


Abbildung 7.7: Diese Grafik zeigt die beiden Distraktorentdeckungsparameter D und a , einmal für die tiefe Enkodierungsbedingung (in Grau) und für die flache Enkodierungsbedingung (in Weiß). Alle Parameterschätzungen beziehen sich auf den sofortigen Testzeitpunkt drei Minuten nach der Lernphase. Die Fehlerbalken zeigen die 95%-Konfidenzintervalle an.

G^2 -Anpassungsstatistik ($\Delta G^2_{(2)} = 0.14$). Auch die Hinzunahme der Gleichsetzung von w_1 und w_2 zum Modell 8a kann ohne Probleme vorgenommen werden ($\Delta G^2_{(3)} = 0.32$). Eine Gleichsetzung der Rateparameter nach ihrer Position (Submodell 9b) führt dahingegen zu einer signifikanten Verschlechterung der Modellanpassung ($\Delta G^2_{(2)} = 6.12$).

Für den Parameter der Distraktorentdeckung D (siehe Abbildung 7.7) ist kein signifikanter Unterschied zu verzeichnen, $\Delta G^2_{(1)} = 1.32$. Die Distraktorentdeckung für Assoziationen fällt bei einer tiefen relationalen Enkodierung mit $a = .45$ signifikant höher aus als in der flachen Enkodierungsbedingung, $a = 0.01$ ($\Delta G^2_{(1)} = 7.21$).

Zwischenfazit. Über die Manipulation der relationalen Verarbeitungstiefe ist es erfolgreich gelungen, die assoziative Enkodierung sowie den Distraktorentdeckungsparameter a signifikant zu beeinflussen. Es ergaben sich keine Auswirkungen auf den assoziativen Abrufparameter. Allerdings lagen in der Satzbedingung konservativere Antwortkriterien vor als in der Bedingung des Vokalvergleichs. Für die Validierung des Enkodierungsparameters ist dieser signifikante Unterschied problematisch, da sich die Manipulation selektiv nur auf den Enkodierungsparameter und im Sinne der diskriminanten Validität nicht auf andere Parameter auswirken sollte. Post-Hoc kann allerdings plausibel erklärt werden, warum sich eine bessere Gedächtnisleistung auch auf die Raten tendenzen auswirkt. Dieser Erklärungsansatz wird in der Gesamtdiskussion von Experiment 2 ausgeführt.

(2) Vergleich des kurzen vs. langen Retentionsintervalls

Im Folgenden werden die beiden Satzbedingungen im Test nach 3 Minuten und im Test nach 24 Stunden verglichen. Schon die relativen Häufigkeiten zeigen, dass eine bessere Gedächtnisleistung nach 3 Minuten vorliegt (siehe Tabelle 7.4). Die relative Häufigkeit richtig klassifizierter alter intakter Wortpaare liegt in dieser Bedingung bei .661, nach 24 Stunden bei .324. Auch die neu kombinierten Wortpaare werden in der direkten Testphase häufiger richtig erinnert (.533) als im Test nach 24 Stunden (.302). Dasselbe Muster gilt für Zielitem-Distraktor-Kombinationen (.548 vs. .385) und Distraktor-Zielitem-Paare (.522 vs. .371).

Inwieweit dieser deskriptive Trend auch zu einem niedrigeren Abrufparameter für Assoziationen führt, wird in der multinomialen Analyse untersucht.

Multinomiale Analyse

Die Modellanpassungsstatistik war für die beiden Teilbedingungen mit $G^2_{(18)} = 16.12$ unter dem kritischen Wert, $G^2_{krit, \alpha=.001} = 42.31$ und zeigt damit eine sehr gute Beschreibung des Datenmusters durch das Modell an.

Die Länge des Retentionsintervalls zeigt einen Effekt auf den Abrufparameter für Assoziationen: In der unmittelbaren Testbedingung ist $r = 0.91$ (siehe Abbildung 7.8) und damit signifikant höher als in der Abrufbedingung nach einem Tag, $r = 0.38$ ($\Delta G^2_{(1)} = 4.10$).

Tabelle 7.4: *Relative Anwothhäufigkeiten für tief enkodierte Wörter nach 3 Minuten Retentionsintervall (a) sowie nach 24 Stunden Retentionsintervall (b)*

(a) Retentionsintervall 3 min (Satzbedingung)

Wahre Itemkat.	Antwort der Versuchsperson				
	Altes WP	Neu komb. WP	Target Dis.	Dis. Target	Dis.Dis.
Altes WP	.661	.157	.078	.067	.037
Neu komb. WP	.057	.533	.161	.202	.048
Target Dis.	.039	.213	.548	.104	.096
Dis. Target	.035	.209	.087	.522	.148
Dis. Dis.	.043	.117	.239	.213	.387

(b) Retentionsintervall 24 h (Satzbedingung)

Wahre Itemkat.	Antwort der Versuchsperson				
	Altes WP	Neu komb. WP	Target Dis.	Dis. Target	Dis.Dis.
Altes WP	.324	.278	.124	.163	.110
Neu komb. WP	.241	.302	.171	.183	.102
Target Dis.	.093	.205	.385	.078	.239
Dis. Target	.098	.200	.185	.371	.146
Dis. Dis.	.102	.141	.234	.268	.254

Tabelle 7.5: *Parameterschätzungen (Par. Schätzungen) und 95%-Konfidenzintervalle (95% KI) für die tief enkodierten Wortpaare nach 3 Minuten sowie 24 Stunden Retentionsintervall*

Parameter	Test nach 3 min			Test nach 24 h	
	Par.Schätzung	95% KI		Par.Schätzung	95% KI
c	0.67	0.58 – 0.76		0.73	0.27 – 1.18
r	0.91	0.83 – 1.00	*	0.39	0.14 – 0.63
w_1	0.64	0.55 – 0.73	*	0.27	0.11 – 0.42
w_2	0.66	0.57 – 0.74	*	0.00	–0.23 – 0.23
g_{10}	0.31	0.18 – 0.44		0.43	0.22 – 0.64
g_{11}	0.40	0.23 – 0.58		0.56	0.49 – 0.64
g_{20}	0.32	0.18 – 0.46	*	0.60	0.46 – 0.74
g_{21}	0.39	0.22 – 0.56	*	0.68	0.61 – 0.75
g_3	0.25	0.13 – 0.36		0.19	0.12 – 0.26
D	0.02	–0.34 – 0.38		0.29	0.20 – 0.38
a	0.45	0.22 – 0.69		0.35	–0.87 – 1.57

*: Signifikante Effekte zwischen den Gruppen auf einem α -Niveau von .05.

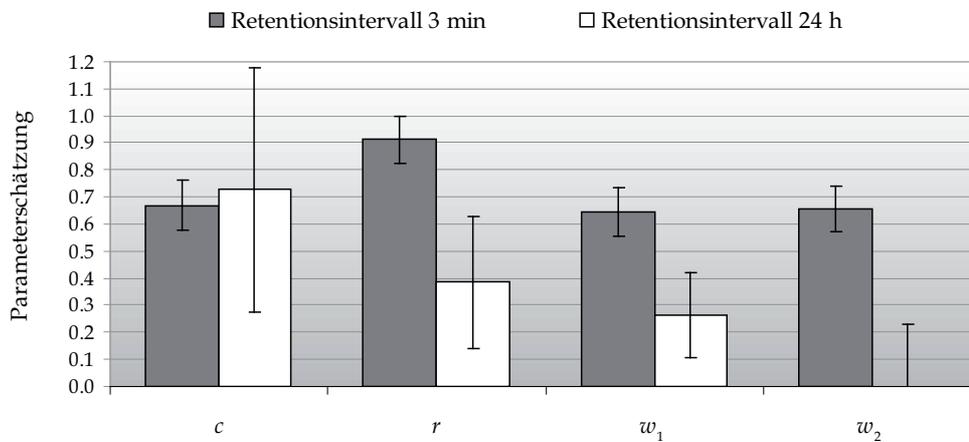


Abbildung 7.8: In dieser Grafik sind die Gedächtnisparameter des ADHD-Modells dargestellt. In Grau sind die Schätzungen für die sofortige Testbedingung abgetragen, in Weiß die Schätzungen für den Test nach 24 Stunden. Für beide Gruppen sind dies die Daten der tiefen Enkodierungsbedingung. Die Fehlerbalken zeigen die 95%-Konfidenzintervalle an.

Die Parameter der assoziativen Enkodierung weisen mit $c = 0.66$ und $c = 0.72$ erwartungsgemäß keinen signifikanten Unterschied auf ($\Delta G_{(1)}^2 = 0.06$). Die Power, für diesen Parametervergleich einen großen Effekt zu entdecken, war mit $1 - \beta = 0.80$ hoch. Die Hybridparameter der Itemerkennung (w_1, w_2) unterscheiden sich zwischen den Bedingungen sowohl für das linke Wort, $\Delta G_{(1)}^2 = 3.96$, als auch für das rechte Wort, $\Delta G_{(1)}^2 = 6.65$.

Übereinstimmend mit den Ergebnissen der Enkodierungsbedingung, sind die Rateparameter in der sofortigen Testbedingung mit Werten zwischen 0.30 bis 0.40 niedriger als die Werte in der verspäteten Testbedingung, 0.42 – 0.68 (siehe Abbildung 7.9). Signifikant wird dieser Effekt für den paarweisen Vergleich von zwei der Rateparameter (g_{11} : $\Delta G_{(1)}^2 = 4.68$; g_{21} : $\Delta G_{(1)}^2 = 4.09$), nicht jedoch für die verbleibenden Vergleichspaare (g_{10} : $\Delta G_{(1)}^2 = 0.97$; g_{20} : $\Delta G_{(1)}^2 = 3.10$). Die Rateparameter im Test nach 24 Stunden, können auf zwei Rateparameter reduziert werden, die die Position des zu ratenden Wortes widerspiegeln (Modell 9b: $\Delta G_{(2)}^2 = 3.81$), eine Einschränkung hinsichtlich des Rateprozesses ist in dieser Bedingung nicht möglich (Modell 9a: $\Delta G_{(2)}^2 = 9.45$). Das umgekehrte Bild zeigt sich in der sofortigen Testbedingung. Hier ist eine Reduktion auf Modell 9a möglich ($\Delta G_{(2)}^2 = 0.14$), nicht aber auf Modell 9b ($\Delta G_{(2)}^2 = 6.16$).

Die Distraktorentdeckungsparameter unterscheiden sich nicht signifikant zwischen den beiden Testbedingungen (D : $\Delta G_{(1)}^2 = 3.29$; a : $\Delta G_{(1)}^2 = 0.02$), siehe Tabelle 7.5.

Zwischenfazit. Als Zwischenfazit kann aus dem Vergleich der beiden Testbedingungen festgehalten werden, dass wie erwartet ein signifikanter Effekt auf den Abrufparameter für Assoziationen festgestellt wurde. Auch die Hybridparameter der Itemerkennung spiegeln eine schlechtere Itemgedächtnisleistung nach 24 Stunden wider. Im Sinne der diskrimi-

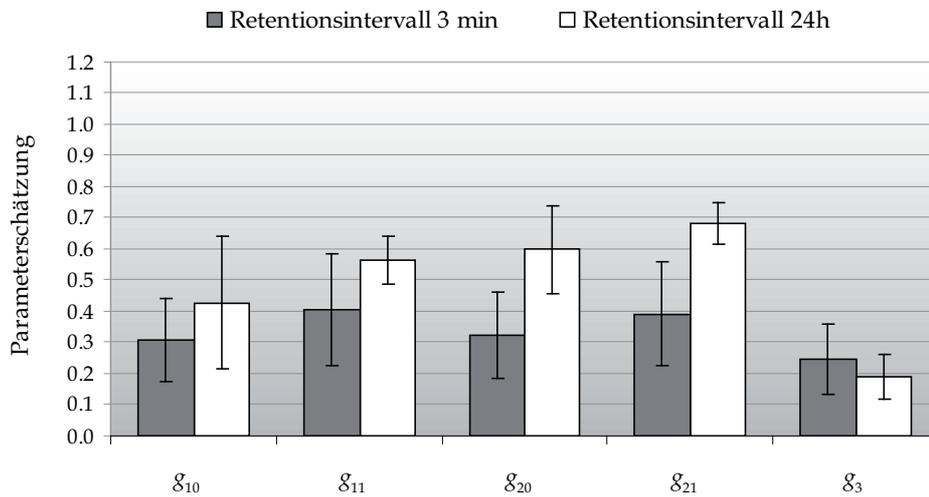


Abbildung 7.9: Diese Abbildung zeigt die Rateparameter für die sofortige Testung nach 3 Minuten sowie für die verzögerte Testung nach 24 Stunden. Alle Parameterschätzungen beziehen sich auf die in der Satzbedingung tief enkodierten Wortpaare. Die Fehlerbalken zeigen die 95%-Konfidenzintervalle an.

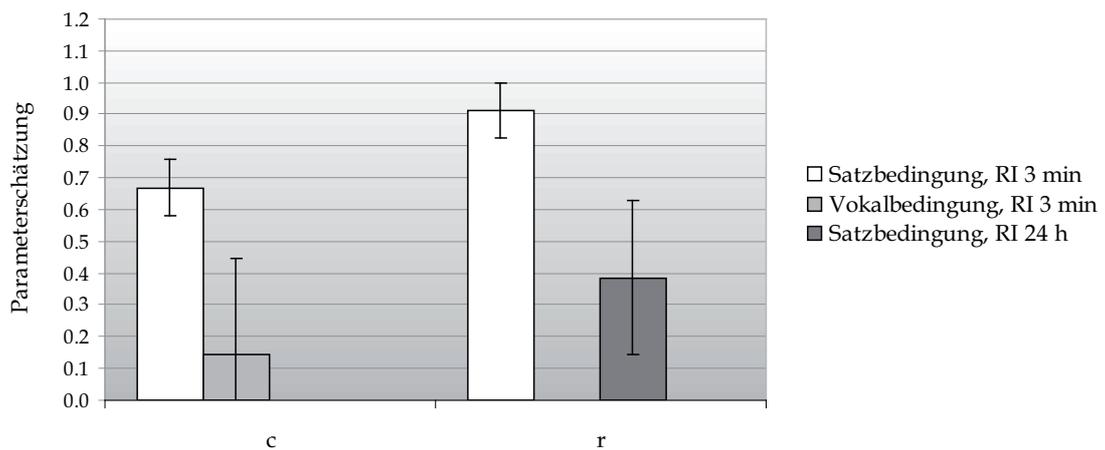


Abbildung 7.10: In dieser Abbildung sind die zentralen Gedächtnisparameter der assoziativen Enkodierung c und des assoziativen Abrufs r dargestellt. Die Fehlerbalken geben die 95%-Konfidenzintervalle an.

		Abruf							
		kurzes RI		langes RI		kurzes RI		langes RI	
Enkodierung	flach	0.144	n.s.	0.001	0.999	*	0.001	n.s.	n.s.
	tief	0.669	n.s.	0.726	0.912	*	0.385	n.s.	n.s.
		Enkodierungsparameter c				Abrufparameter r			

Abbildung 7.11: In dieser Abbildung sind die Ergebnisse von Experiment 2 in Anlehnung an das Enkodierungs-Abruf-Paradigma dargestellt. Nicht signifikante Gruppenunterschiede (n.s.) sowie auf einem α -Niveau von .05 signifikante Vergleiche (*) sind gekennzeichnet.

nanten Validität und eines selektiven Effekts konnte keine Veränderung im assoziativen Enkodierungsparameter beobachtet werden. Die Rateparameter weisen deskriptiv einen vergleichbaren Trend zu der vorherigen Analyse auf: In der Bedingung mit einer höheren Gedächtnisleistung sind die Antwortkriterien konservativer. Dieser Effekt ist in zwei Parametern signifikant. Eine Erklärung dieses Effekts sowie eine Ausführung der Problematik im Hinblick auf die Validierung des Modells wird in der Gesamtdiskussion gegeben.

(3) Gesamtanalyse

In den bisherigen Teilanalysen konnte bereits gezeigt werden, dass die Manipulation des Retentionsintervalls sich signifikant auf den Abrufparameter auswirkt, nicht aber auf den Enkodierungsparameter. Umgekehrt hatte die Manipulation der Enkodierungsbedingung zwar einen Effekt auf den Enkodierungsparameter nicht aber auf den Abrufparameter. Die bereits diskutierten Parameterschätzungen sind in Abbildung 7.10 dargestellt.

Im Folgenden sollen die beiden zentralen Parameter c und r zwischen allen vier Gruppen des Enkodierungs-Abruf-Designs verglichen werden. Dies soll aufzeigen, inwieweit nur die erwarteten Haupteffekte vorliegen.

In Abbildung 7.11 sind die zentralen Parameter der assoziativen Enkodierung und des assoziativen Abrufs in Abhängigkeit der vier experimentellen Gruppen dargestellt. Der linken Tabelle kann entnommen werden, dass die Manipulation der Enkodierungsbedingung sich auf den Parameter c auswirkt: Die Werte der oberen Zeile sind jeweils signifikant verschieden von den Werten der unteren Zeile ($\Delta G_{(1)}^2 = 13.54$; $\Delta G_{(1)}^2 = 7.19$). Die Auswirkungen der Enkodierungsmanipulation auf den Enkodierungsparameter zeigen sich also zu beiden Testzeitpunkten. Dies bestätigt die konvergente Validität der Manipulation.

In Abhängigkeit des Retentionsintervalls verändern sich die Enkodierungsparameter nicht, dies kann an den vergleichbaren Parameterschätzungen innerhalb jeder Zeile abgelesen werden. Diese Vergleiche erreichten keine Signifikanz, womit die diskriminante Validität bestätigt wurde ($\Delta G_{(1)}^2 = 1.02$; $\Delta G_{(1)}^2 = 0.06$).

In der rechten Tabelle sind analog zur linken Tabelle die Parameterschätzungen des Abrufparameters dargestellt. Hier zeigt sich das umgekehrte Muster: Der Parameter r variiert innerhalb der Zeilen, nicht aber zwischen den Zeilen. Der Abrufparameter unterscheidet sich also nicht zwischen den Enkodierungsbedingungen ($\Delta G_{(1)}^2 = 0.10$; $\Delta G_{(1)}^2 = 0.01$), wohl aber zwischen den beiden Abrufbedingungen. Dieser Vergleich wird allerdings nur in der Satzbedingung signifikant ($\Delta G_{(1)}^2 = 4.10$), nicht in der Vokalbedingung ($\Delta G_{(1)}^2 = 0.01$). In der Vokalbedingung lagen aufgrund des geringeren c -Parameters zu wenig Datenpunkte für eine reliable Schätzung des r -Parameters vor. Im Vergleich der verbleibenden drei experimentellen Gruppen zeigt sich die konvergente sowie diskriminante Validität des Abrufparameters.

7.3 Diskussion

Das Ziel von Experiment 2 war es, die beiden Parameter der assoziativen Enkodierung und des assoziativen Abrufs zu validieren. Die Enkodierungsmanipulation wurde operationalisiert über eine tiefe, relationale Verarbeitungsbedingung, in der die Probanden die Wortpaare eingebettet in sinnvollen Sätzen lernten, kontrastiert mit einer flachen Enkodierungsbedingung, in der die Probanden angeben mussten, ob beide Wörter dieselbe Anzahl an Vokalen aufweisen. Zusätzlich stand den Probanden in der Satzbedingung mit fünf Sekunden eine längere Lernzeit zur Verfügung als den Probanden in der Vokalbedingung, die den Vokalvergleich innerhalb von 3.5 Sekunden vornehmen mussten. Dies sollte den Effekt der Enkodierungsmanipulation zusätzlich verstärken.

Beide experimentellen Gruppen mussten zwei Testphasen durchlaufen. Die Hälfte der Wörter wurde drei Minuten nach der Lernphase getestet. Für die restlichen Wortpaare fand ein unangekündigter Test nach 24 Stunden statt. Diese Manipulation sollte sich auf den Erfolg des assoziativen Abrufs (Parameter r) niederschlagen. Es wurde erwartet, dass eine höhere Rate von Wortpaaren nach 3 Minuten erfolgreich abgerufen werden kann, als nach 24 Stunden. Sowohl die Enkodierungsmanipulation als auch die Abrufmanipulation sollte Effekte auf die Hybridparameter der Itemerkennung haben.

Die Auswertung zeigte, dass sich die Manipulation der Enkodierungsbedingung wie erwartet signifikant auf den Enkodierungsparameter sowie auf die Parameter der Itemerkennung auswirkt. Ebenfalls im Sinne der Hypothesen blieb ein Effekt auf den assoziativen Abrufparameter aus.

Die zwei unterschiedlichen Testintervalle hatten wie erwartet einen signifikanten Einfluss auf den Abrufparameter und auf die Itemerkennungsparameter. Der Enkodierungsparameter blieb von dieser Manipulation unbeeinflusst. In den Gedächtnisparametern konnte das erwartete Ergebnis also hypothesenkonform festgestellt werden. Dies zeigt die konvergente Validität des Experiments auf.

Des Weiteren hatten sowohl die Enkodierungsmanipulation als auch die Manipulation des Retentionsintervalls einen Effekt auf die Rateparameter. Dieser Effekt zeigte über beide Manipulationen hinweg ein konsistentes Muster: Die Ratetendenzen waren konservativer in den Bedingungen, in denen eine relativ gute Gedächtnisleistung vorlag, und liberaler in den Bedingungen mit geringerem Anteil korrekter Erinnerungen. Konkret lagen in der tiefen Enkodierungsbedingung konservativere Schätzungen vor als in der flachen Enkodierungsbedingung; die Rateparameter für tief enkodierte Worte fielen nach 3 Minuten konservativer aus als nach 24 Stunden.

Dieses Ergebnismuster wurde zwar nicht vorhergesagt, kann aber Post-Hoc sinnvoll interpretiert werden und stellt unter gewissen Voraussetzungen³¹ kein Problem für die Validierung der Parameter dar. Dies wird im Folgenden genauer ausgeführt.

Der Befund, dass eine Manipulation der Erinnerungsleistung, die auf die Gedächtnisparameter abzielt, sich auch in Rateprozessen niederschlägt, wurde im Kontext der multinomialen Modellierung bereits in anderen Studien beobachtet (z.B. Bayen et al., 1996; Meiser et al., 2007). Bayen et al. (1996) variierten in einem Quellengedächtnisexperiment die Ähnlichkeit von Zielitems zu Distraktoren, um damit den Itemrekognitionsparameter (D) ihres Quellengedächtnismodells zu validieren. Wie erwartet hatte diese Manipulation den gewünschten Effekt auf die Erinnerungsleistung. Zusätzlich zeigte sich aber ein Effekt im Itemrateparameter (b). Vergleichbar zu Experiment 2 der vorliegenden Arbeit fielen die Ratetendenzen liberaler aus, wenn Distraktoren und Targets eine hohe Ähnlichkeit aufwiesen und damit die Erinnerung erschwert war.

Auch in Ja-nein-Rekognitionsexperimenten (z.B. Hilford, Glanzer & Kim, 1997, Exp. 1,2; Hirshman, 1995), in ‚forced-choice‘-Experimenten (z.B. Bowles & Poon, 1982; Hilford et al., 1997, Exp. 3-5), in Experimenten mit Rating-Skalen (z.B. J. Brown, Lewis & Monk, 1977, Exp. 3) sowie in assoziativen Rekognitionsstudien (Greene, 1996; Hockley, 1994; Hockley & Niewiadomski, 2007; Xu & Malmberg, 2007) wurde eine Veränderung der Antwortkriterien aufgrund von Gedächtnismanipulationen beobachtet (für einen Überblick über den Effekt bei verschiedenen Testformaten siehe Glanzer & Adams, 1985). Unabhängig von der Auswertungsmethode findet sich in all diesen Experimenten eine erhöhte Trefferrate sowie eine herabgesetzte Rate falscher Alarme, wenn Manipulationen eine gute Gedächtnisleistung hervorrufen. Ein umgekehrtes Muster (niedrige Treffer und hohe Raten falscher Alarme) liegt unter erinnerungser schwerenden Bedingungen vor. Aufgrund des resultierenden Datenmusters ging dieses paradigmunenabhängige Phänomen als ‚Mirror-Effekt‘ (Glanzer & Adams, 1985; Glanzer & Bowles, 1976) in die Literatur ein.

Die Manipulationen, die einen solchen Effekt hervorrufen, sind sehr vielfältig. In einer Meta-Analyse von Glanzer und Adams (1985) sowie bei Hirshman (1995) werden berichtet: (1) niedrig- vs. hochfrequente Wörter; (2) konkrete (bzw. gut vorstellbare) vs. abstrakte

³¹Dieses Ergebnismuster kann dann als unproblematisch angesehen werden, wenn es in weiteren Experimenten gelingt, die Rateparameter unabhängig von den Gedächtnisparametern zu beeinflussen.

(bzw. schlecht vorstellbare) Wörter; (3) bedeutungshaltige Wörter vs. sinnlose Wortsilben; (4) Bilder vs. Wörter; (5) wiederholte vs. einfache Darbietung; (6) lange vs. kurze Darbietungsdauer; (7) tief vs. flach enkodierte Wörter. Diese Manipulationen wurden häufig verwendet, in weiteren Studien konnte der Mirror-Effekt auch mit anderen Mitteln erreicht werden (siehe Glanzer & Bowles, 1976).

Weiterhin ist der Effekt unabhängig vom experimentellen Design. Die meisten Studien variierten die Gedächtnisstärke zwar innerhalb der Versuchsteilnehmer, der Effekt wurde aber auch für Manipulationen zwischen den Teilnehmern nachgewiesen (Hilford et al., 1997; Hockley & Niewiadomski, 2007; Jenkins, Neale & Deno, 1967; Paivio & Csapo, 1969; Rowe & Rogers, 1975). Für die vorliegende Arbeit besonders relevant ist dabei die Tatsache, dass der Mirror-Effekt auch für Manipulationen zwischen Versuchspersonen im assoziativen Rekognitionsparadigma gezeigt werden konnte (Hockley & Niewiadomski, 2007), da dieses Paradigma und Design dem vorliegenden Experiment 2 sehr ähnlich sind.

Kaum weniger zahlreich als die empirischen Befunde zum Mirror-Effekt sind die Erklärungsansätze. Im Folgenden soll kurz auf drei dieser Ansätze (Batchelder & Batchelder, 2008; Greene, 1996; Hirshman, 1995) eingegangen werden (für einen Überblick über weitere Erklärungsansätze siehe z.B. Glanzer, Adams, Iverson & Kim, 1993; Hirshman, 1995; Greene, 1996).

Die Erklärungsansätze von Greene (1996) sowie Hirshman (1995) grenzen sich gegenüber anderen Erklärungsansätzen dadurch ab, dass sie den Fokus nicht auf Gedächtnisprozesse legen, sondern auf Antwortstrategien der Versuchspersonen. Greene geht davon aus, dass Probanden versuchen, ihre Antworten in etwa gleich auf die unterschiedlichen Stimuli zu verteilen. Haben Probanden z.B. Wörter tief und flach enkodiert, so besteht ihre Tendenz darin, „Alt“-Antworten genauso oft auf tief wie flach enkodierte Wörter zu geben, sowie gleich häufig „alt“ und „neu“ zu antworten. Da tief enkodierte Wörter besser erinnert werden, kann ein großer Anteil von „Alt“-Antworten aufgrund von Erinnerungsprozessen gegeben werden. Um das Ziel einer Gleichverteilung der Antworten zu erreichen, werden die Probanden in einem solchen Fall unter Unsicherheit eher dazu neigen, „neu“ zu raten. Greene demonstriert, dass dieser Ansatz sowohl auf Ja-nein-Rekognitionstests als auch auf 2AFC-Tests angewandt werden kann. Greene stützt seinen Erklärungsansatz in Experimenten, die nachweisen, dass durch die gezielte Einflussnahme auf Antwortstrategien der Mirror-Effekt eliminiert werden kann (Greene, 1996, Exp. 8,9).

Hirshman (1995) postuliert, dass Probanden ihr Antwortkriterium nach Abschätzung der Spannweite der Erinnerungsgüte für das vorliegende Itemset festlegen. Dieser Ansatz ist explizit auch für Manipulationen zwischen Teilnehmern anwendbar (Hirshman, 1995, S. 310). Hirshman verglich in mehreren Experimenten die Vorhersagen konkurrierender Ansätze mit empirischen Daten und konnte auf diese Weise Evidenz für seine Hypothese aufzeigen.

In eine ähnliche Richtung wie Greene und Hirshman argumentieren Batchelder und Batchelder (2008). Sie beschäftigen sich mit metakognitiv gesteuerten Einflüssen auf Rateprozesse im Quellengedächtnisparadigma³² und postulieren dabei zwei Heuristiken, die Probanden anwenden und die damit einen Einfluss auf die Rateprozesse haben.³³ Die erste Heuristik umschreiben sie als Heuristik der kausalen Ursache (,cause heuristic‘). Hierbei wird angenommen, dass Probanden im Zustand der Unsicherheit überlegen, welche Quelle bzw. Itemklasse diesen Zustand verursacht haben könnte. Im Kontext des ,reality monitorings‘ z.B. sollen Probanden angeben, welche Items von ihnen selbst generiert (oder erinnert) wurden und welche von einer anderen Quelle stammen (Voss, Vesonder, Post & Ney, 1987). Bei Quellenunsicherheit führt die Kausalitätsheuristik zu der Schlussfolgerung, dass dieses Item mit einer hohen Wahrscheinlichkeit aus der fremden Quelle stammen muss, da die Erinnerungsgüte an selbstgenerierte Items als besonders hoch eingeschätzt wird. Über diesen Kausalitätsmechanismus erklärt die Heuristik Veränderungen in den Rateparametern des Quellengedächtnisses (siehe z.B. Voss et al., 1987).

Mit derselben Logik können aber auch unterschiedliche Itemratetendenzen erklärt werden: Wurden in einem Experiment Items teilweise nach 3 Minuten, teilweise nach 24 Stunden getestet, so sollten in der Testbedingung nach 24 Stunden liberalere Kriterien angelegt werden. Die Probanden können das lange Retentionsintervall als Ursache für die schlechte Gedächtnisleistung ansehen und sind daher eher geneigt, in dieser Testsituation „alt“ zu raten.

Die zweite Heuristik wird als Basisratenheuristik bezeichnet (,base rate heuristic‘) und geht davon aus, dass Probanden die Basisraten (z.B. von alten und neuen Items im Test) berücksichtigen, wenn sie unter Unsicherheit raten. Auch dies trifft für Iteminformationen sowie Quelleninformationen zu: Kommen im Test mehr Aussagen vor, die von Quelle A präsentiert wurden als von Quelle B, so wird dies zu einer Ratetendenz für Quelle A führen. Liegen im Test mehr alte als neue Aussagen vor, wird eine liberale Itemratetendenz ausgebildet. Beide Heuristiken wurden von Batchelder und Batchelder (2008) durch bereits vorliegende Experimente untermauert.

Der Erklärungsansatz von Batchelder und Batchelder (2008) passt sehr gut zu den Experimenten von Meiser et al. (2007), die wie im vorliegenden Experiment ihre Daten über

³²Wie die Autoren in ihrem Artikel erläutern, ist ihr Erklärungsansatz keineswegs auf das Quellengedächtnisparadigma beschränkt. Er kann auch auf Rekognitionsstudien angewandt werden. Aufgrund der Ähnlichkeit des vorliegenden Paradigmas mit dem Quellengedächtnisparadigma scheint ein Transfer der Annahmen unproblematisch, insbesondere da die involvierten Gedächtnisprozesse sehr ähnlich sind. In beiden Paradigmen muss zunächst entschieden werden, ob einzelne Items gelernt wurden, dann, inwieweit diese von einer bestimmten Quelle bzw. im Kontext eines anderen Wortes präsentiert wurden.

³³Die Autoren schreiben zwar nicht explizit, ob die Probanden diese Heuristiken eher bewusst einsetzen oder ob es sich dabei um einen automatischen Prozess handelt. Ein reflektierter Einsatz der Heuristiken scheint aber aufgrund der Schilderung naheliegend zu sein. Wichtig ist deshalb an dieser Stelle hervorzuheben, dass es auch Erklärungsansätze für den Mirror-Effekt gibt, die nicht davon ausgehen, dass die Probanden ihr Antwortverhalten bewusst steuern. Diese Ansätze können den Mirror-Effekt auch in Experimenten erklären, in denen die Itemklassen für den Probanden nicht klar voneinander zu unterscheiden sind.

ein multinomiales Modell auswerteten. In drei Experimenten zeigten diese Autoren, dass sich Manipulationen wie die mehrfache Wiederholung von Items, die Präsentationszeit sowie metakognitive Informationen über unterschiedliche Quellen nicht nur in Gedächtnisparametern, sondern auch in Rateparametern niederschlagen. Dieser Befund ist gemeinsam mit dem Ergebnis von Bayen et al. (1996) für die vorliegende Studie relevant, da das vorliegende Modell mit einem ähnlichen Paradigma arbeitet, vergleichbare kognitive Prozesse beinhaltet und es sich bei allen drei Modellen um Zwei-Hochschwellenmodelle handelt. Insofern fügt sich das Ergebnismuster von Experiment 2 gut in die vorliegenden Befunde ein und findet durch die Erklärungsansätze zum Mirror-Effekt eine sinnvolle Interpretation.

Zusammenfassend wurde in dieser Diskussion erläutert, wie man die Ergebnisse aus Experiment 2 plausibel begründen kann und warum es sich dabei um weitere Evidenz für den Mirror-Effekt handeln könnte. Sowohl der Ansatz von Greene (1996) als auch der Ansatz von Batchelder und Batchelder (2008) können zur Erklärung von Experiment 2 herangezogen werden. Die Ergebnisse von Bayen et al. (1996) sowie Meiser et al. (2007) weisen parallele Ergebnisse unter Verwendung von sehr verwandten multinomialen Modellen auf.

Zwar wurde das Ergebnismuster von Experiment 2 Post-Hoc plausibel erklärt, für die erfolgreich Validierung der Parameter ist es allerdings unabdingbar zu zeigen, dass die Rateparameter unabhängig von den Gedächtnisparametern manipuliert werden können. Es ist also zwingend notwendig, in einem weiteren Experiment selektiv die Rateparameter zu verändern, und gleichzeitig die Gedächtnisparameter konstant zu halten. Nur wenn dies gelingt, kann von einer erfolgreichen Validierung gesprochen werden. Dieses Ziel wird daher in Experiment 3 und Experiment 4 verfolgt.

Kapitel 8

Experiment 3 und 4

Ziel der folgenden beiden Experimente ist die Validierung der Rateparameter für Items. In Experiment 2 hatten die Manipulationen des Abrufintervalls und der Enkodierungstiefe nicht nur einen Effekt auf die Gedächtnisparameter, sondern auch auf die Rateparameter. Wie diskutiert wurde, kann dieses Ergebnismuster über adaptive Kriterien der Versuchspersonen erklärt werden und steht im Einklang mit anderen Experimenten (siehe 7.3). Für die erfolgreiche Validierung des ADH-Modells ist es allerdings notwendig zu zeigen, dass es Manipulationen gibt, die sich selektiv nur auf die Rateparameter auswirken. Dies wird in den Experimenten 3 und 4 angestrebt.

In der Literatur werden - abhängig vom experimentellen Paradigma - sehr unterschiedliche Manipulationen aufgeführt, um Raten tendenzen zu beeinflussen:

(1) In Quellengedächtnisaufgaben können die Rateparameter z.B. über die Schema-Typikalität manipuliert werden (Bayen, Nakamura, Dupuis & Yang, 2000). Wird eine Quelle als Arzt, die andere als Anwalt vorgestellt, so werden unter Unsicherheit medizinische Aussagen eher dem Arzt zugeordnet als dem Anwalt (und umgekehrt für juristische Stellungnahmen). Die Probanden versuchen also bei Unwissenheit plausibel zu raten und wenden ihr metakognitives Wissen z.B. über die Quellen sinnvoll an (Batchelder & Batchelder, 2008).

(2) Auch in ‚Who said what?‘-Experimenten, einem Paradigma, das dem des Quellengedächtnisses sehr ähnlich ist, können die Rateparameter über Stereotype beeinflusst werden (Klauer, Wegener & Ehrenberg, 2002). Ebenso haben gezielte Informationen über Subgruppen einen Einfluss auf die Antworttendenzen (Klauer & Wegener, 1998, Exp. 4). Werden z.B. viele negative Aussagen von Männern getroffen und viele positive Aussagen von Frauen, so ordnen die Probanden unter Unsicherheit positive Aussagen eher einer Frau zu. Auch dies reflektiert den sinnvollen Einsatz metakognitiver Strategien.

(3) Durch den Anteil alter Wörter in der Testliste (z.B. 75% alte Wörter im Test oder nur 25% alte Wörter im Test) wurden u.a. im ‚Who said what?‘- und Remember-Know-Paradigma, im Kontext der Prozess-Dissoziationsprozedur und auch in einfachen Rekogni-

tionsstudien die Antworttendenzen verändert (Buchner et al., 1995; Hirshman & Henzler, 1998; Klauer & Wegener, 1998; Strack & Forster, 1995). Dieser Effekt beruht auf einem ähnlichen Prinzip wie die Subgruppeninformation im ‚Who said what?‘-Paradigma: Den Probanden wird entweder direkt über die Instruktionen oder indirekt über Leistungsfeedback der Anteil alter Wörter (bzw. alter Informationen) im Test vermittelt. Nutzen die Probanden dieses Wissen, so tendieren sie dazu, bei Unwissenheit die häufiger vorkommende Wortkategorie auszuwählen.

(4) Über ‚Auszahlungsmatrizen‘ (Payoff-Matrizen) gelang es, Rateparameter eines multinomialen Modells zur Prozess-Dissoziationsprozedur zu validieren (Buchner et al., 1995). Aber auch in Rekognitionsstudien zeigte diese Manipulation den erwarteten Effekt auf Rateprozesse (Curran, DeBuse & Leynes, 2007; A. F. Healy & Kubovy, 1978). Bei Auszahlungsmatrizen liegt der Gedanke zugrunde, die Probanden z.B. über Prämien zu einer möglichst guten Leistung zu motivieren. Im Test werden dann richtige (falsche) Antworten unterschiedlich stark belohnt (bestraft). Mit dem Ziel, eine möglichst hohe Leistung zu erreichen, wählen die Probanden unter Unsicherheit die Antwortalternative, für die es bei einer falschen Einschätzung eine geringe Bestrafung, bei einer richtigen Einschätzung eine hohe Belohnung gibt. Sie versuchen also unter Unsicherheit Belohnungen zu maximieren und Bestrafungen zu minimieren.

Diese Auflistung von Manipulationen ist definitiv nicht erschöpfend und könnte weiter ergänzt werden (siehe z.B. Rateparameter des multinomialen Modells zum Rückschaufehler, Erdfelder & Buchner, 1998 oder Rateparameter des Modells zur Wason-Aufgabe, Klauer, Stahl & Erdfelder, 2007). Für dieses Kapitel ist es aber relevant, solche Manipulationen gegeneinander abzuwägen, die auf den kombinierten Item- und assoziativen Rekognitionstest angewandt werden können. Hierfür sind vor allem die letzten beiden Manipulationsklassen geeignet. Sowohl die Veränderung der Basisraten als auch Auszahlungsmatrizen könnten direkt auf das vorliegende Paradigma übertragen werden und erfordern keine Veränderungen in zentralen Aspekten des Lern- oder Testkontexts. Des Weiteren wurden beide Manipulationen bereits erfolgreich zur Validierung von Rateprozessen in anderen multinomialen Modellen verwendet (z.B. Buchner et al., 1995; Klauer & Wegener, 1998). Daher wird im Folgenden ein kurzer Überblick über die bisherige Anwendung dieser beiden Manipulationen gegeben, bevor dann auf die konkrete Operationalisierung eingegangen wird.

Manipulation der Ratetendenzen über Basisraten. Der Idee, Ratetendenzen über das Verhältnis von alten Wörtern zu neuen Wörtern im Test zu manipulieren, geht die Annahme voraus, dass Probanden ihre Antwortkriterien an die jeweilige Testsituation anpassen. Im letzten Kapitel wurde dieser Gedanke bereits diskutiert und Evidenz dazu aufgeführt (siehe vor allem Batchelder & Batchelder, 2008). Bei Studien zur Manipulation

von Basisraten wird die Versuchsperson über unterschiedliche Wege auf die vorliegende Basisrate (direkt oder indirekt) hingewiesen. Den Probanden wird die Basisrate entweder in den Instruktionen explizit mitgeteilt, oder die Probanden erhalten eine Leistungsrückmeldung und können über dieses Feedback Wissen über die Zusammenstellung der Testliste erlangen. Auch eine Kombination der beiden Möglichkeiten ist denkbar.

Eine erfolgreiche Anwendung dieser Manipulation konnte z.B. von Buchner et al. (1995) demonstriert werden. Ihren Experimenten lag das Ziel zugrunde, die Rateparameter ihres multinomialen Modells zur Prozess-Dissoziations-Prozedur zu validieren. Im ersten Experiment veränderten sie die Basisraten, und den Probanden wurde nach jedem Item eine Leistungsrückmeldung gegeben. In den Instruktionen wurden die Probanden *nicht* auf das Verhältnis von alten zu neuen Wörtern hingewiesen. In einem weiteren Experiment erhielten die Probanden nur nach jedem zehnten Item eine Leistungsrückmeldung; die objektiv zugrunde liegende Basisrate wurde den Teilnehmern dafür in den Instruktionen mitgeteilt. In beiden Experimenten zeigte sich ein signifikanter Effekt der Manipulation auf die Rateparameter.

Auch Klauer und Wegener (1998) veränderten im ‚Who said what?‘-Paradigma die Basisraten und teilten das Verhältnis den Teilnehmern mit, auf ein Leistungsfeedback verzichteten sie. Die beiden experimentellen Gruppen hatten entweder eine Testliste mit 50% neuen Aussagen, oder eine Liste mit 70% neuen Aussagen zu beurteilen. Die Manipulation zeigte einen selektiven und signifikanten Effekt auf die Ratetendenzen.

Nicht nur im Kontext der Validierung von multinomialen Modellen, sondern zu ganz unterschiedlichen Fragestellungen wurde die experimentelle Veränderung der Basisraten verwendet und zeigte Auswirkungen auf die - zum Teil sehr unterschiedlich erfassten - Ratetendenzen (Estes & Maddox, 1995; Heit, Brockdorff & Lamberts, 2003; Hirshman & Henzler, 1998; Ratcliff, Sheu & Gronlund, 1992; Rotello, Macmillan, Hicks & Hautus, 2006; Strack & Forster, 1995). In den meisten Experimenten wurde versucht eine neutrale Antworttendenz über 50% alter Wortpaare zu erreichen; für liberale Antworttendenzen werden mindestens 70% alte Wortpaare getestet, für konservative Ratetendenzen maximal 30%. Diese Basisraten können daher als Orientierungsmaßstäbe für die vorliegende Studie verwendet werden.

Manipulation der Ratetendenzen über Auszahlungsmatrizen. Auszahlungsmatrizen werden in der Literatur seltener zur Beeinflussung von Antworttendenzen berichtet als Basisraten. Wahrscheinlich liegt dies daran, dass eine höhere Motivation auf Seiten der Teilnehmer erforderlich ist und die Aufgaben meist relativ komplex sind.

Der Grundgedanke dieser Experimente besteht darin, die Probanden für richtige Antworten unterschiedlich stark zu belohnen und sie für falsche Antworten differenziert zu bestrafen. Konkret könnten z.B. alte Items, die korrekt als „alt“ klassifiziert werden, mit 5 Pluspunkten stark belohnt werden, dafür wird eine „Neu“-Antwort auf ein altes Item mit

5 Minuspunkten auch stark bestraft. Dies sollte bei den Probanden unter Unsicherheit zu einer liberalen Antworttendenz führen: Für eine falsch geratene „Alt“-Antwort erhalten sie nur eine geringe Bestrafung von einem Minuspunkt, für eine falsch geratene „Neu“-Antwort werden ihnen aber 5 Punkte abgezogen.

Voraussetzung für die erfolgreiche Anwendung dieser Prozedur ist es, einen Anreiz zu schaffen, der die Probanden dazu motiviert, eine gute Leistung zu zeigen und Punkte zu sammeln. Außerdem müssen die Probanden sich genauer als in den Standard-Gedächtnisexperimenten mit den Instruktionen auseinandersetzen, da diese etwas komplexer sind. Wahrscheinlich haben diese beiden Aspekte dazu geführt, dass Basisratenmanipulationen häufiger verwendet wurden. Zahlreiche Studien bestätigen aber den Erfolg der Manipulation der Auszahlungsmatrizen (Buchner et al., 1995; Curran et al., 2007; A. F. Healy & Kubovy, 1978; Maddox & Bohil, 1998).

In der bereits erwähnten Studie von Buchner et al. (1995) wurden die Probanden mit einem vs. drei Punkten belohnt bzw. bestraft. Die Auszahlungsmatrix war so aufgestellt, dass sie zu liberalen vs. konservativen Antworttendenzen führen sollte. Zusätzlich erhielten die Probanden ein Leistungsfeedback nach jedem Urteil. Die Ergebnisse zeigten, dass sich die Rateparameter in die intendierte Richtung veränderten. In anderen Experimenten wurden zum Teil noch stärkere Unterschiede in der Auszahlung von einem vs. fünf Punkten verwendet (z.B. Curran et al., 2007).

Zusammenfassend zeigte sich sowohl in Validierungsstudien zu multinomialen Modellen als auch im Kontext von anderen Fragestellungen, dass Auszahlungsmatrizen sowie Basisraten erfolgreich angewandt werden können, um Ratetendenzen zu beeinflussen. Beide Manipulationen sollen daher auch zur Validierung des ADH-Modells genutzt werden: In Experiment 3 wird eine Veränderung der Basisraten vorgenommen, in Experiment 4 werden Auszahlungsmatrizen eingesetzt. In beiden Experimenten sollten selektive Effekte auf die Rateparameter resultieren. Welche Erkenntnisse Experiment 4 über Experiment 3 hinaus bringt, wird im Kapitel 8.2.1 erläutert.

8.1 Experiment 3

8.1.1 Methode

In Experiment 3 wird die Basisrate von alten Wörtern in der Testliste manipuliert, um einen Einfluss auf die Rateparameter auszulösen. Konkret sollen in der einer Gruppe 70% alte Wörter im Test vorkommen, in der anderen Gruppe 30% alte Wörter. Wie bereits geschildert, entsprechen diese Basisraten der gängigen experimentellen Praxis (z.B. Buchner et al., 1995; Hirshman & Henzler, 1998; Klauer & Wegener, 1998).

Anders als in einfachen Rekognitionstests muss im vorliegenden kombinierten Item- und assoziativen Rekognitionstest beachtet werden, dass trotz einer Veränderung der Basisra-

Tabelle 8.1: Häufigkeiten der Wortkombinationen in Experiment 3

Itemkategorie	Anzahl der Wortpaare	
	Gruppe 1	Gruppe 2
	Liberaler Ratetendenz	Konservative Ratetendenz
Lernliste	56	56
Testliste	80	80
Altes WP	22	8
Neu komb. WP	22	8
Target Dis.	12	8
Dis. Target	12	8
Dis. Dis.	12	48

ten das Verhältnis zwischen alten intakten Wortpaaren und neu kombinierten Wortpaaren gleich bleibt. Auch sollten genauso viele Target-Distraktor-Kombinationen wie Distraktor-Target-Wortpaare getestet werden.³⁴ Unter diesen Voraussetzungen sollten die vier Itemratetendenzparameter (g_{10} , g_{11} , g_{20} , g_{21}) innerhalb der beiden experimentellen Gruppen zu vergleichbaren Parameterschätzungen führen und sich nur zwischen den Gruppen unterscheiden.

Mit diesem Ziel wurden zunächst die Lernlisten im Vergleich zu den vorangegangenen beiden Experimenten um 6 Wortpaare verlängert. Dies ermöglichte eine gleichmäßige Aufteilung der Testwörter auf die unterschiedlichen Itemtypen. In der ersten Gruppe, in der eine *liberale Ratetendenz* erreicht werden sollte, wurden 22 alte intakte sowie neu kombinierte Wortpaare getestet und jeweils 12 Wortpaare aus den verbleibenden drei Itemkategorien. Insgesamt bestand diese Testliste aus 112 alten Wörtern (56 an der rechten Position, 56 an der linken Position) und 48 neuen Wörtern (ebenso gleichverteilt auf die beiden Positionen). Dies führte zu einer Testliste von 160 Items d.h. 80 Wortpaaren (siehe Tabelle 8.1).

Für die zweite Gruppe war die Länge der Lern- und Testliste mit der ersten Gruppe vergleichbar. Unterschiedlich lange Lern- oder Testlisten wurden vermieden, da dies einen Einfluss auf die Gedächtnisparameter haben könnte und damit die Interpretation der Ergebnisse erschweren würde. Daher wurden 56 Wortpaare gelernt und 80 getestet. Im Test waren diesmal 112 Wörter neu, nur 48 Wörter alt. Die häufigste Wortkombination in dieser Testliste waren Distraktor-Distraktor-Kombinationen mit insgesamt 48 Wortpaaren, alle anderen Wortkombinationen kamen jeweils 8 mal vor (siehe Tabelle 8.1). Die Probanden dieser Gruppe sollten eher eine *konservative Ratetendenz* zeigen.

Um einen möglichst starken Effekt zu erreichen, wurden nicht nur die Basisraten der Testlisten verändert, den Probanden wurde dies auch explizit mitgeteilt. Sie wurden in

³⁴Eine ungleiche Verhältnis von alten intakten Wortpaaren und neu kombinierten Wortpaaren würde einen zusätzlichen Effekt auf den assoziativen Rateparameter g_3 haben. Eine ungleiche Verteilung von Target-Distraktor- zu Distraktor-Target-Kombinationen müsste zu unterschiedlichen Rateparametern für das linke bzw. rechte Wort führen.

den Instruktionen darüber aufgeklärt, wie viel Prozent der Items in der Testphase alt bzw. neu sind. Weiterhin wurde ihnen Feedback über ihre Leistung gegeben. Auch dies sollte zu einem starken Effekt zwischen den Gruppen beitragen.

Teilnehmer und Design

An Experiment 3 nahmen 54 Studierende der Universität Mannheim teil, 81.5% davon aus dem Fachbereich Psychologie. Das mittlere Alter der Teilnehmer betrug 21.17 Jahre ($SD = 3.07$). Die Teilnehmer wurden randomisiert auf zwei Bedingungen aufgeteilt: Entweder erhielten sie eine Testliste mit 30% alten Wörtern oder eine Testliste mit 70% alten Wörtern.

Ablauf

Der Ablauf von Experiment 3 war bis zur Testphase identisch mit dem Ablauf von Experiment 1b. Die Probanden wurden über ihre Rechte als Versuchsperson aufgeklärt und erhielten die Instruktion, Wortpaare zu lernen, die später getestet werden. Nach einem Übungsdurchgang, der aus einer kurzen Lern- und Testphase bestand, startete dann die eigentliche Lernphase. Den Probanden beider Gruppen wurden 56 Wortpaare für vier Sekunden dargeboten. Die Präsentationszeit war mit Experiment 1a/b identisch, die Lernliste wurde um 6 Wortpaare verlängert. Nach einer Distraktoraufgabe von einer Minute begann die Testphase. In der ersten Gruppe mit *70% alten Wörtern in der Testphase* wurden den Probanden die folgenden Instruktionen gegeben:

„Bitte beachten Sie für die nun folgende Testphase: Die Wahrscheinlichkeit, dass ein Wort „alt“ ist, beträgt 70%. Dementsprechend beträgt die Wahrscheinlichkeit, dass ein gezeigtes Wort „neu“ ist, 30%.

Folglich wurden 70% der Wörter, die jetzt in der Testphase gezeigt werden, in der Lernphase gelernt. Die restlichen 30% der Wörter sind komplett neue Wörter.

Falls Sie sich also in einer Entscheidung unsicher sind, lohnt es sich für Sie, „alt“ zu raten anstatt „neu“ zu raten, denn es sind mehr alte Wörter in der Testliste als neue Wörter.

Nach vier Beurteilungen erhalten Sie Feedback darüber, wie viele Ihrer Antworten richtig waren. Falls Sie Fragen haben, richten Sie diese bitte nun an den anwesenden Versuchsleiter. Ansonsten starten Sie bitte die Testphase mit der Leertaste.“

Danach wurden die Wortpaare gezeigt, und die Probanden sollten über das sequentielle Ja-nein-Format ihr Urteil abgeben (siehe Abbildung 8.1 A). Nach jedem vierten Wort

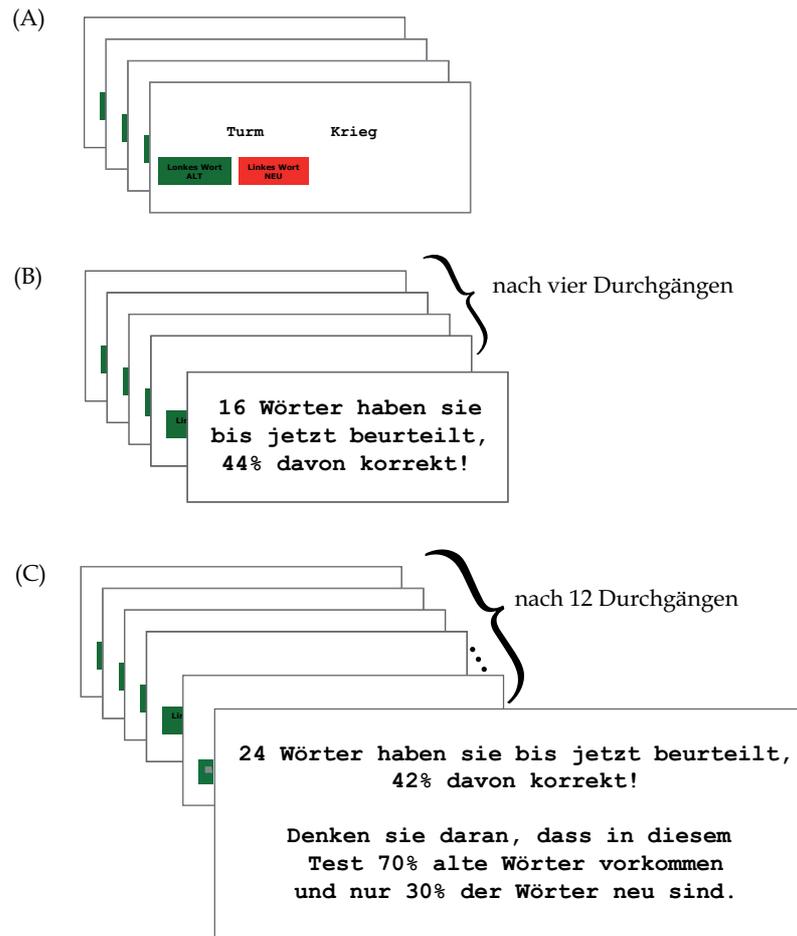


Abbildung 8.1: In Abbildung (A) ist das Ja-nein-Format zu sehen, über das die Teilnehmer ihre Antworten eingeben konnten. Abbildung (B) zeigt die Leistungsrückmeldung an, die nach vier Beurteilungen eingeblendet wurde. In Abbildung (C) ist diese Leistungsrückmeldung nach jedem 12. Durchgang dargestellt. Hier wurde zusätzlich daran erinnert, wie hoch der Anteil von alten Wörtern im Test ist.

erhielten die Probanden Feedback über ihre Leistung. Die Leistung bezog sich nur auf die Beurteilung der Einzelwörter, nicht auf die Wortkombination. Den Probanden wurde jeweils mitgeteilt, wie viele Wörter sie bisher beurteilt haben und welcher Prozentsatz davon richtig klassifiziert wurde (siehe 8.1 B). Diese Information wurde drei Sekunden lang angezeigt und konnte von den Probanden nicht vorzeitig beendet werden, da sichergestellt werden sollte, dass alle Probanden sich mit dem Feedback auseinandersetzen.

Nach der Beurteilung von jeweils 12 Wortpaaren wurde den Probanden zusätzlich zu der Leistungsrückmeldung mitgeteilt, welcher Prozentsatz alter Wörter in der Testliste vorlag (siehe 8.1 C). Diese Erinnerung sollte sicherstellen, dass die Probanden diese Information nutzen, um ihre Leistung durch geschicktes Raten zu verbessern.

Die zweite Gruppe mit *30% alten Wörtern in der Testphase* sah die zu ihrer Bedingung reziproken Instruktionen, danach war der Ablauf der Testphase identisch zu der ersten Gruppe.

Nach dem Testdurchgang erhielten alle Probanden ein Gesamtfeedback über die Leistung in der ersten Testphase. Auf eine dreiminütige Pause folgte dann der zweite Lern- sowie Testdurchgang (mit neuem Wortmaterial), der absolut parallel zu der ersten Lern-Testphase aufgebaut war.

Im Anschluss an die zweite Testphase wurden den Probanden zwei Fragen zu dem Experiment gestellt: (1) Sie sollten angeben, wie viel Prozent der Wörter in der Testphase *laut Instruktionen* alt, d.h. gelernt waren. (2) Die Teilnehmer wurden gefragt wie viel Prozent der Wörter in der Testphase *ihrer Einschätzung nach* alt, d.h. gelernt waren. Diese beiden Fragen dienten der Überprüfung der Manipulation.

Material

In diesem Experiment wurde eine längere Testliste von jeweils 80 Wörtern verwendet, daher war eine der parallelisierten Wortlisten nicht ausreichend. Aus diesem Grund wurden die Wörter randomisiert aus beiden Wortlisten 1 und 2 gezogen. Die Teilnehmer an Experiment 3 nahmen an keinem weiteren Experiment dieser Arbeit teil.

8.1.2 Ergebnisse

Bevor eine Datenanalyse vorgenommen wird, sollen zunächst die beiden Fragen zur Überprüfung der Manipulationen ausgewertet werden. Die Frage, wieviel Prozent der Wörter im Test laut Instruktionen alt waren, wurde von einem Probanden falsch beantwortet, dieser wurde daher aus der Analyse ausgeschlossen. Da die Information über den Anteil alter Wörter im Test sehr häufig wiederholt wurde, schien es plausibel anzunehmen, dass dieser Teilnehmer den Instruktionen und Rückmeldungen keine Aufmerksamkeit geschenkt hat.

Auf die Frage, wie viel Prozent alte Wörter *nach Einschätzung der Probanden* im Test vorkamen, wurde in der ersten Gruppe mit 70% alten Wörtern im Mittel ein Prozentsatz

Tabelle 8.2: Relativen Antworthäufigkeiten der Bedingung mit 70% alten Wörtern (a), sowie 30% alten Wörtern (b)

(a) 70% alte Wörter

Wahre Itemkat.	Antwort der Versuchsperson					
	Altes WP	Neu komb. WP	Target Dis.	Dis. Target	Dis.Dis.	
Altes WP	.710	.153	.061	.050	.026	
Neu komb. WP	.122	.618	.107	.121	.033	
Target Dis.	.054	.346	.458	.072	.069	
Dis. Target	.087	.353	.087	.393	.082	
Dis. Dis.	.061	.212	.250	.237	.240	

(b) 30% alte Wörter

Wahre Itemkat.	Antwort der Versuchsperson					
	Altes WP	Neu komb. WP	Target Dis.	Dis. Target	Dis.Dis.	
Altes WP	.569	.080	.105	.094	.152	
Neu komb. WP	.036	.326	.261	.212	.165	
Target Dis.	.020	.060	.533	.051	.335	
Dis. Target	.022	.074	.056	.536	.313	
Dis. Dis.	.009	.024	.118	.109	.741	

von 65.58% angegeben ($SD = 12.90$; Spannweite: 50% – 90%). In der zweiten Gruppe lag der mittlere Prozentsatz bei 31.82% ($SD = 13.75$; Spannweite: 6% – 60%). Da in den jeweiligen Gruppen die mittleren Werte nicht signifikant von 70% bzw. 30% abweichen, $t_{(25)} = -1.74$, $p > .05$ bzw. $t_{(27)} = 0.70$, $p > .05$, kann davon ausgegangen werden, dass die Probanden den Anteil alter Wörter wie intendiert wahrgenommen haben.

Auswertung der relativen Häufigkeiten

Zunächst werden die relativen Häufigkeiten berichtet, im Anschluss daran die multinomiale Analyse mit den relevanten Parametertests zu den Hypothesen dieses Experiments. In der Tabelle 8.2 sind die relativen Häufigkeiten der Antworten auf die unterschiedlichen Itemkategorien dargestellt.

Im Vergleich der beiden Gruppen zeigen sich Unterschiede in den Raten der korrekten Klassifikationen. Für die alten intakten Wortpaare z.B. ist die Rate korrekter Klassifikationen in der Bedingung mit 70% alten Wörtern höher, als in der Bedingung mit 30% alten Wörtern (.710 > .569). Gleiches gilt auch für die neu kombinierten Wortpaare (.618 > .326). Das umgekehrte Muster ist für Zielitem-Distraktor- und Distraktor-Zielitem-Kombinationen zu beobachten. Hier sind die Raten korrekter Klassifikationen in der Bedingung mit 30% alten Wörtern höher als in der anderen Bedingung (.533 > .458;

.536 > .393). Auch die Rate der korrekten Distraktorentdeckung für zwei Distraktoren ist in der 30%-Bedingung mit .741 höher als in der 70%-Bedingung (.240).

Diese Unterschiede sollten nicht auf unterschiedliche Gedächtnisleistungen zwischen den Gruppen zurückzuführen sein, sondern auf unterschiedlichen Ratetendenzen. Inwieweit dies zutrifft, wird durch die multinomiale Analyse mit getrennten Parametern für Ratetendenzen und Gedächtnisleistung deutlich. Auch für die relativen Häufigkeiten kann aber gezeigt werden, dass die Raten korrekter Klassifikationen durch Ratetendenzen beeinflusst sind: Für beide Gruppen wurden in einer weiterführenden Analyse korrigierte assoziative und Itemrekognitionsleistungen berechnet. Dazu wurden nicht nur die Hit-Raten berücksichtigt, sondern von den Hit-Raten jeweils die Raten falscher Alarme abgezogen. In dieser Analyse zeigt sich für die Gruppe mit 70% alten Wörtern eine assoziative Gedächtnisleistung von .588, in der anderen Gruppe fällt diese Leistung mit .533 nur leicht geringer aus. Auch die Gedächtnisleistungen der Item-Rekognition nähern sich mit .602 (70%-Bedingung) sowie .510 einander an. Dies deutet darauf hin, dass die großen Unterschiede in den relativen Häufigkeiten auf unterschiedliche Raten von falschen Alarmen zurückgehen. Genauer und nicht nur deskriptiv, sondern inferenzstatistisch soll dieser Aspekt in der multinomialen Auswertung analysiert werden.

Multinomiale Analyse

In der multinomialen Auswertung konnte mit $G^2_{(18)} = 31.38$ ($G^2_{krit, \alpha=.001} = 42.31$) eine befriedigende Modellanpassungsstatistik erreicht werden. Auch für diesen Datensatz lag eine lokale Identifizierbarkeit der Parameter vor, diese wurde analog zu den bisher berichteten Experimenten nachgewiesen. Wie Abbildung 8.2 entnommen werden kann, unterscheiden sich die Rateparameter in der 70%-Bedingung deutlich von den Rateparametern in der 30%-Bedingung. Dieser Unterschied ist in allen paarweisen Vergleichen signifikant (g_{10} : $\Delta G^2_{(1)} = 20.79$; g_{11} : $\Delta G^2_{(1)} = 9.82$; g_{20} : $\Delta G^2_{(1)} = 16.39$; g_{21} : $\Delta G^2_{(1)} = 10.93$). Darüber hinaus können die Parameter in der 70%-Bedingung alle auf den Wert 0.70 fixiert werden (alle $\Delta G^2_{(1)} < 0.93$), in der 30%-Bedingung ist eine Fixierung auf 0.30 möglich (alle $\Delta G^2_{(1)} < 2.10$). Es wurde also nicht nur eine Beeinflussung der Rateparameter in die gewünschte Richtung erreicht, die Rateparameter konnten sogar auf die intendierten Werte festgelegt werden.

Im Unterschied zu den Rateparametern zeigt sich in Abbildung 8.3 kein signifikanter Unterschied in den Gedächtnisparametern. Alle Parameter konnten über die beiden Bedingungen hinweg gleichgesetzt werden, ohne eine Verschlechterung der Modellanpassung hervorzurufen (r : $\Delta G^2_{(1)} = 1.63$; c : $\Delta G^2_{(1)} = 0.36$; w_1 : $\Delta G^2_{(1)} = 0.03$; w_2 : $\Delta G^2_{(1)} = 0.71$). Die Teststärke zur Entdeckung eines Effekts von 0.30 war für die Parameter r , w_1 , sowie w_2 mit $1 - \beta > 0.92$ sehr hoch, lediglich für den Enkodierungsparameter war die Power für einen mittleren Effekt nicht zufriedenstellend. Für diesen Parameter müsste man zugleich das α -Niveau auf .10 anheben, um eine höhere Teststärke sicherzustellen. An den Ergeb-

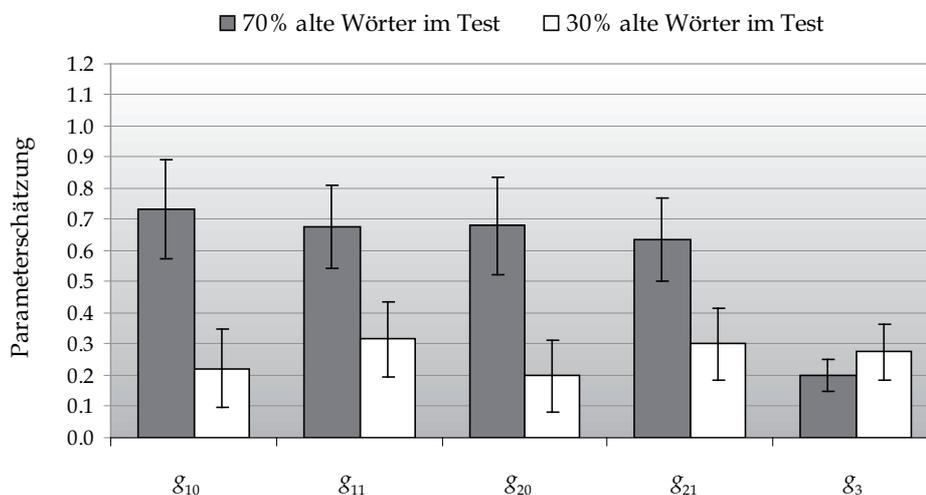


Abbildung 8.2: In dieser Abbildung sind die Rateparameter des ADH-Modells dargestellt. In Grau sind die Parameterschätzungen für die Bedingung mit 70% alten Wörtern im Test abgebildet, in Weiß die Parameterschätzungen für die Bedingung mit 30% alten Wörtern im Test. Die Fehlerbalken zeigen die 95%-Konfidenzintervalle an.

nissen hinsichtlich des c -Parameters ändert sich auch unter diesen Bedingungen nichts, so dass damit eine ausreichende Power für den Parameter c garantiert werden konnte.

In Tabelle 8.3 sind die Parameterschätzungen für die Distraktorentdeckungsparameter dargestellt. Deskriptiv fallen die Parameterschätzungen in der Bedingung mit einem höheren Distraktorenanteil größer aus. Dieser Unterschied ist allerdings weder für den Parameter a ($\Delta G^2_{(1)} = 3.25$), noch für den Parameter D ($\Delta G^2_{(1)} = 3.63$) signifikant.

Im Hinblick auf spezifische Submodelle führt in beiden experimentellen Bedingungen das Modell 10 zu einer guten Modellanpassung. In diesem Submodell werden die Itemerkennungparameter innerhalb der Bedingungen gleichgesetzt (70%-Bedingung: $\Delta G^2_{(1)} = 1.24$; 30%-Bedingung: $\Delta G^2_{(1)} = 0.07$). Auch die Restriktion auf zwei Rateparameter, welche die Rateprozesse reflektieren, kann in beiden Gruppen vorgenommen werden (Modell 9a: 70%-Bedingung: $\Delta G^2_{(2)} = 3.46$; 30%-Bedingung: $\Delta G^2_{(2)} = 1.50$). Die genannten Restriktionen können auch gleichzeitig vorgenommen werden ohne eine signifikante Verschlechterung der Modellanpassungsstatistik G^2 auszulösen (70%-Bedingung: $\Delta G^2_{(3)} = 3.52$; 30%-Bedingung: $\Delta G^2_{(3)} = 2.15$). Dies führt zu dem spezifischeren Modell 8a.

8.1.3 Diskussion

In Experiment 3 sollten die Rateparameter validiert werden. Als Manipulation wurde entweder eine Testliste verwendet, in der nur ein geringer Anteil von Wörtern (30%) alt war, oder es wurde eine Testliste genutzt, die hauptsächlich aus alten Wörtern bestand

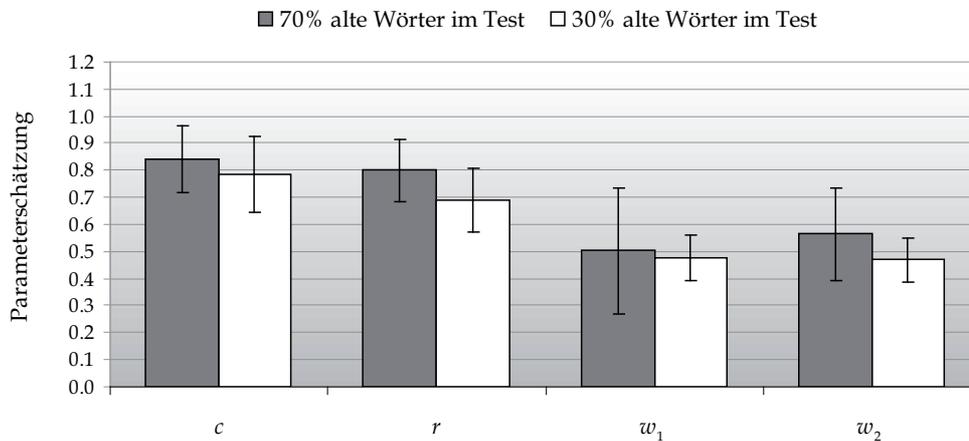


Abbildung 8.3: Die Gedächtnisparameter des multinomialen Modells sind in dieser Abbildung abgetragen. In Grau sind die Parameterschätzungen für die Bedingung mit 70% alten Wörtern zu sehen, in Weiß die Schätzungen für die Bedingung mit 30% alten Wörtern. Als Fehlerbalken sind die 95%-Konfidenzintervalle abgetragen.

Tabelle 8.3: Parameterschätzungen (Par. Schätzungen) in Abhängigkeit der beiden experimentellen Gruppen mit 70% bzw. 30% alten Wörtern im Test sowie die dazugehörigen 95%-Konfidenzintervalle (95% KI)

Parameter	70% alte Wörter			30% alte Wörter	
	Par.Schätzung	95% KI		Par.Schätzung	95% KI
c	0.84	0.72 – 0.96		0.79	0.65 – 0.92
r	0.80	0.69 – 0.91		0.69	0.57 – 0.81
w_1	0.50	0.27 – 0.74		0.48	0.39 – 0.56
w_2	0.56	0.39 – 0.73		0.47	0.39 – 0.55
g_{10}	0.73	0.58 – 0.89	*	0.22	0.10 – 0.35
g_{11}	0.67	0.54 – 0.81	*	0.32	0.20 – 0.44
g_{20}	0.68	0.52 – 0.83	*	0.20	0.08 – 0.32
g_{21}	0.64	0.50 – 0.77	*	0.30	0.18 – 0.42
g_3	0.20	0.15 – 0.25		0.27	0.18 – 0.36
D	0.25	0.10 – 0.40		0.45	0.26 – 0.65
a	0.17	–0.06 – 0.40		0.48	0.23 – 0.73

*: Signifikante Effekte zwischen den Gruppen auf einem α -Niveau von .05.

(70% gelernte Wörter). Die Probanden wurden jeweils vor der Testphase über den Anteil von alten Wörtern informiert. Während des Tests erhielten sie Feedback über ihre bisherige Leistung, und sie wurden in regelmäßigen Abständen an den Anteil alter Wörter im Test erinnert.

In der multinomialen Analyse zeigten sich hochsignifikante Effekte auf die Itemparameter: In der Bedingung mit 30% alten Wörtern wurden signifikant niedrigere d.h. konservativere Schätzungen erreicht als in der Bedingung mit 70% alten Wörtern. Auch konnten die Parameter jeweils auf die Werte von 0.30 bzw. 0.70 fixiert werden. Dies zeigt den Erfolg der experimentellen Manipulation und die konvergente Validität des Experiments auf.

Auch im Hinblick auf die diskriminante Validität konnte in Experiment 3 ein Erfolg erzielt werden. Die Gedächtnisparameter konnten zwischen den Bedingungen jeweils gleichgesetzt werden, dasselbe gilt für die Parameter der Distraktorentdeckung. Diese Ergebnisse sind nach den Resultaten von Experiment 2 besonders wichtig, denn in dem vorangehenden Experiment lag eine Konfundierung der Gedächtnisparameter und der Rateparameter vor. Es wurde diskutiert, inwieweit die mit einer hohen Gedächtnisleistung einhergehenden konservativen Ratetendenzen plausibel erklärt werden können. Um diese Post-Hoc-Erklärung zu stützen, war eine erfolgreiche Manipulation der Rateparameter ohne Einfluss auf die Gedächtnisparameter notwendig. Dies ist in Experiment 3 erfolgreich gelungen.

8.2 Experiment 4

In Experiment 3 gelang es über eine Manipulation der Basisraten die vier Itemrateparameter zu validieren. Das Ziel von Experiment 4 ist es zum einen, über eine alternative Manipulation die Ergebnisse des vorangegangenen Experiments 3 zu stützen und den Befund zu replizieren, dass die Rateparameter selektiv beeinflusst werden können. Zum anderen soll gleichzeitig eine differenzierte Validierung stattfinden: In Anlehnung an Experiment 1b soll gezeigt werden, dass die Rateparameter g_{10} und g_{11} jeweils Ratetendenzen für die linke Wortposition widerspiegeln, die Rateparameter g_{20} und g_{21} reziproke Antworttendenzen für das rechte Wort. Anders als in Experiment 3 sollen also nicht nur die Rateparameter global validiert werden, sondern die Aufspaltung nach der Wortposition soll gerechtfertigt werden.

8.2.1 Methode

Konkret wird in diesem Experiment eine Auszahlungsmatrix aufgestellt, die dazu führen soll, dass die Probanden für das rechte und linke Wort unterschiedliche Ratetendenzen entwickeln. Die Auszahlungsmatrix soll also eine liberale Antworttendenz für das rechte Wort und gleichzeitig eine konservative Antworttendenz für das linke Wort herbeiführen und umgekehrt. Jeder Proband muss für das rechte Wort eine andere Auszahlungsmatrix berücksichtigen als für das linke Wort (Manipulation innerhalb der Teilnehmer). Selekt-

tiv sollten sich also die Itemrateparameter sowohl innerhalb der Gruppen verändern als auch zwischen den Gruppen. Die Gedächtnisparameter und Parameter der Distraktorentdeckung sollten keine Differenzen zwischen den Gruppen aufweisen.

Teilnehmer

An Experiment 4 nahmen 44 Studierende der Universität Mannheim teil, 79.5% davon aus dem Fachbereich Psychologie. Das mittlere Alter der Probanden betrug 20.73 Jahre ($SD = 2.32$; Spannweite: 18-29 Jahre). Keiner der Probanden hat auch an Experiment 3 teilgenommen.

Design

Die Rateparameter wurden in diesem Experiment innerhalb der Probanden manipuliert. Zwischen den Probanden wurde ausbalanciert, welche Wortposition jeweils für alte Antworten stark bzw. schwach belohnt wurde. Konkret wurden die Probanden in Gruppe 1 für korrekte „Alt“-Antworten auf das linke Wort stark belohnt; „Alt“-Antworten auf das rechte Wort wurden nur schwach verstärkt. In Gruppe 2 wurde die Belohnungsmatrix umgekehrt: „Alt“-Antworten auf das linke Wort wurden schwach belohnt, „Alt“-Antworten auf das rechte Wort stark.

Ablauf

Bis zur Testphase war der Ablauf von Experiment 4 vergleichbar zu Experiment 3. Die Probanden wurden über ihre Rechte aufgeklärt; ihnen wurde erläutert, dass sie Wortpaare lernen müssen, die später getestet werden, und sie hatten Gelegenheit, sich in einer kurzen Übungsphase mit der Aufgabe vertraut zu machen. Analog zu Experiment 3 wurden in der Lernphase 56 Wortpaare für je 4 Sekunden präsentiert. Nach einem Retentionsintervall von einer Minute erhielten die Probanden die entsprechenden Informationen zur Testphase.

Die Teilnehmern beider Gruppen wurden informiert, dass sie in diesem Experiment Punkte sammeln können, die sie danach in Schokoriegel eintauschen können. Punkte erhalten sie jeweils, wenn sie ein Wort korrekt klassifizieren, d.h. auf ein neues Wort korrekt „neu“ oder auf ein altes Wort korrekt „alt“ antworten. Allerdings werden ihnen bei falschen Antworten auch Punkte abgezogen. Die genaue Belohnungsmatrix war für Gruppe 1 und 2 verschieden.

Den Teilnehmern *aus Gruppe 1* wurde dargelegt, dass sie

- (1) bei einer *korrekten* „Alt“-Antwort auf das linke Item 3 Pluspunkte erhalten;
- (2) bei einer *korrekten* „Neu“-Antwort auf das linke Item 1 Pluspunkt gewinnen;
- (3) bei einer *falschen* „Alt“-Antwort auf das linke Item ihnen 1 Punkt abgezogen wird;
- (4) bei einer *falschen* „Neu“-Antwort auf das linke Item sie 3 Punkte verlieren. Für das linke Item werden also korrekte „Alt“-Antworten stark belohnt und falsche „Alt“-Antworten nur

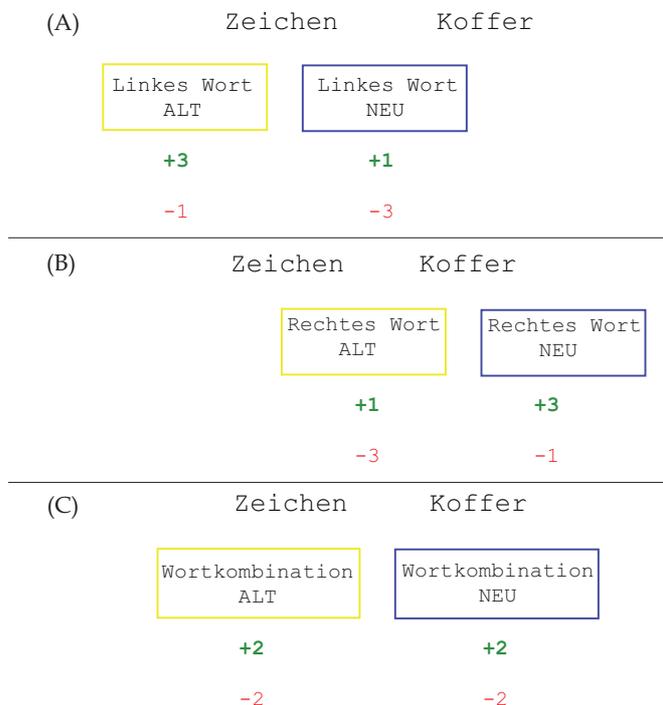


Abbildung 8.4: Diese Abbildung zeigt, wie die Testphase in Experiment 3 für die Teilnehmer aussah: Zunächst wurden die Teilnehmer gebeten, ihr Urteil über das linke Wort abzugeben (A), dann für das rechte (B) und ggf. für die Kombination (C). Die entsprechende Punktematrix für die Antworten auf das linke Wort wurde bei jedem Wortpaar eingeblendet. In der ersten Zeile waren die Punkte für korrekte Antworten in Grün (hier in Dunkelgrau) dargestellt. In der zweiten Zeile waren in Rot die Punkte für falsche Antworten abgetragen (hier in Hellgrau).

gering bestraft. Dies sollte dazu führen, dass für das *linke* Item die Tendenz besteht, unter Unsicherheit „alt“ zu raten.

Umgekehrt werden korrekte „Alt“-Antworten für das rechte Wort nur schwach belohnt, falsche „Alt“-Antworten werden aber streng bestraft. Für das *rechte* Wort sollten die Probanden also eher dazu tendieren, „neu“ zu raten. Die genaue Auszahlungsmatrix ist in Abbildung 8.4 dargestellt. Die Zusammenfassung der Instruktionen für die Probanden war mit der Matrix in Abbildung 8.5 vergleichbar.

Für die Teilnehmer aus *Gruppe 2* wurden die Punkteauszahlungen umgekehrt. Die Punkte waren so gestaffelt, dass sie unter Unsicherheit für das *linke Wort eine konservative Ratetendenz*, für das *rechte eine liberale Ratetendenz* entwickeln sollten.

Bei beiden Gruppen war die Punktevergabe für die Antworten auf die Wortkombination identisch und neutral. Richtige „Alt“-Antworten sowie richtige „Neu“-Antworten wurden mit 2 Pluspunkten belohnt, falsche „Alt“- sowie falsche „Neu“-Antworten wurden mit 2 Minuspunkten bestraft. Es sollte sich also keine Antworttendenz für das Raten der Assoziation ergeben.

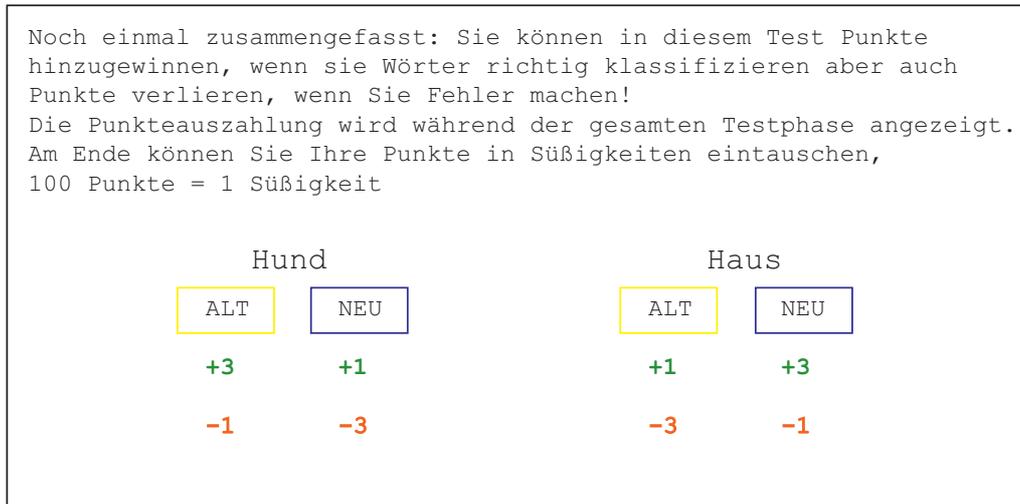


Abbildung 8.5: Diese Grafik zeigt die Instruktionen für die Versuchspersonen. Auf diesem Bildschirmausschnitt wurde zusammenfassend angegeben, wie die Punkteverteilung bei richtigen bzw. falschen Antworten auf das rechte bzw. linke Wort erfolgen würde. Die Pluspunkte waren in der ersten Zeile jeweils in Grün, die Minuspunkte in der zweiten Zeile in Rot dargestellt.

Alle Teilnehmer sahen in der Testphase unter jedem Wort die konkrete Punktevergabe abgebildet. Dies ist in Abbildung 8.5 veranschaulicht. Zudem wurde den Probanden nach jeweils 4 Beurteilungen Feedback über ihren aktuellen Punktestand gegeben.

Nach der ersten Testphase folgte eine dreiminütige Pause. Danach durchliefen die Teilnehmer unter denselben Bedingungen erneut einen Lern- und Testdurchgang. Nach Beendigung der zweiten Testphase konnten die Probanden ihre Punkte gegen Schokoriegel eintauschen, wobei sie für 100 Punkte einen Schokoriegel erhielten.

Material

Für dieses Experiment wurden dieselben Materialien wie für Experiment 3 verwendet. Die Wörter wurden zufällig aus den beiden Wortlisten 1 und 2 gezogen.

8.2.2 Ergebnisse

Wie in den vorangegangenen Experimenten werden zunächst die relativen Häufigkeiten der gegebenen Antworten berichtet. Diese sind in Tabelle 8.4 dargestellt. Bei beiden Gruppen zeigen die relativen Häufigkeiten mit Werten über .500 eine gute Gedächtnisleistung an.

Konform mit der Auszahlungsmatrix zeigen sich deskriptiv keine Unterschiede in den korrekten Antworten auf alte, intakte Kombinationen (.746 vs. .718) oder neue Kombinationen (.549 vs. .575). In der ersten Gruppe, die für das „Alt“-Raten des linken und „Neu“-Raten des rechten Wortes belohnt wurde, sind korrekte Klassifikationen auf Target-Distraktor-Items mit .635 höher als in der zweiten Gruppe (.521). Das Umgekehrte gilt für

Tabelle 8.4: *Relative Antworthäufigkeiten der ersten (a) und zweiten Gruppe (b) aus Experiment 4*

(a) Gruppe 1

Wahre Itemkat.	Antwort der Versuchsperson				
	Altes WP	Neu komb. WP	Target Dis.	Dis. Target	Dis.Dis.
Altes WP	.764	.100	.058	.049	.029
Neu komb. WP	.065	.549	.196	.136	.054
Target Dis.	.024	.136	.635	.041	.164
Dis. Target	.037	.183	.076	.541	.163
Dis. Dis.	.012	.064	.212	.117	.595

(b) Gruppe 2

Wahre Itemkat.	Antwort der Versuchsperson				
	Altes WP	Neu komb. WP	Target Dis.	Dis. Target	Dis.Dis.
Altes WP	.718	.089	.054	.099	.040
Neu komb. WP	.075	.575	.145	.177	.028
Target Dis.	.025	.225	.521	.086	.143
Dis. Target	.016	.171	.034	.650	.128
Dis. Dis.	.004	.063	.138	.286	.509

korrekten Klassifikationen auf Distraktor-Target-Kombinationen: Übereinstimmend mit der Auszahlungsmatrix fallen diese in der zweiten Gruppe höher aus als in der ersten (.650 > .541). Dieses Muster kann natürlich nicht nur über die falschen Alarme erklärt werden, sondern prinzipiell auch über Veränderungen in der Gedächtnisleistung. Welche Erklärung zutrifft, soll die multinomiale Analyse zeigen.

Multinomiale Analyse

Das Modell weist mit $G_{(18)}^2 = 21.589$ einen deutlich niedrigeren Wert auf als der kritische Wert ($G_{krit,\alpha=.001}^2 = 42.31$), daher kann die Modellanpassung als sehr gut beurteilt werden. Auch die lokale Identifizierbarkeit wurde für dieses Experiment über den Rang der Jacobischen Matrix und die mehrfache Berechnung der Parameterschätzungen überprüft und ist gegeben.

In Abbildung 8.6 sind die Rateparameter für die beiden Gruppen abgetragen. Da sich die Rateparameter auf die Wahrscheinlichkeit beziehen, für ein Wort „alt“ zu raten, ist über die Pluszeichen gekennzeichnet, ob die korrekte „Alt“-Antwort für das jeweilige Wort stark belohnt (+++) oder nur schwach belohnt wurde (+). In Gruppe 1 wurden korrekte „Alt“-Antworten für das linke Wort stark belohnt (g_{10} ; g_{11}) für das rechte Wort (g_{20} ; g_{21}) nur schwach. Erkennbar ist, dass die Ratetendenzen bei einer starken „Alt“-Belohnung mit

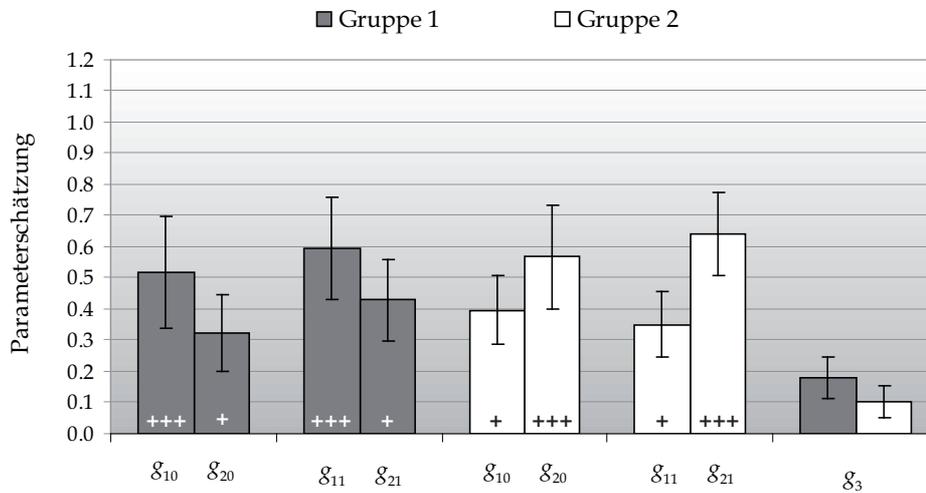


Abbildung 8.6: Diese Abbildung zeigt die Rateparameterschätzungen für Gruppe 1 (in Grau) und Gruppe 2 (in Weiß). Die ,+‘-Markierungen in den Säulen geben jeweils an, ob für dieses Wort eine „Alt“-Antwort stark (+++) oder nur schwach belohnt wurde (+). Als Fehlerbalken sind die 95%-Konfidenzintervalle abgetragen.

Werten von $g_{10;Gruppe1} = 0.51$, $g_{11;Gruppe1} = 0.59$, $g_{20;Gruppe2} = 0.56$, und $g_{21;Gruppe2} = 0.64$ jeweils höher ausfallen, als bei einer schwachen „Alt“-Belohnung, $g_{20;Gruppe1} = 0.39$, $g_{21;Gruppe1} = 0.35$, $g_{10;Gruppe2} = 0.32$, und $g_{11;Gruppe2} = 0.42$. Diese Differenz erreicht in allen vier paarweisen Vergleichen höchste Signifikanz (g_{10} : $\Delta G_{(1)}^2 = 15.124$; g_{11} : $\Delta G_{(1)}^2 = 15.19$; g_{20} : $\Delta G_{(1)}^2 = 10.85$; g_{21} : $\Delta G_{(1)}^2 = 43.82$).³⁵

Im Gegensatz zu den Gruppenunterschieden in den Rateparametern zeigen die Parameter der Gedächtnisleistung keine Beeinflussung durch die Auszahlungsmatrix auf (siehe Abbildung 8.7). Kritisch sind insbesondere die Parameter der Itemerkennung für das linke bzw. das rechte Wort, aber weder der Parameter w_1 zeigt einen signifikanten Unterschied zwischen den Gruppen ($\Delta G_{(1)}^2 = 1.96$), noch der Parameter w_2 ($\Delta G_{(1)}^2 = 0.19$). Gleiches gilt für den Enkodierungsparameter c ($\Delta G_{(1)}^2 = 1.24$) und den Abrufparameter r ($\Delta G_{(1)}^2 = 3.26$). Für die Parameter der Gedächtnisleistung lag mit einer Power von mindestens .95 eine sehr gute Teststärke vor, um mittlere Effekte zu entdecken (r : $1-\beta = .99$; c : $1-\beta = .95$; w_1 : $1-\beta = .99$; w_2 : $1-\beta = .99$).

Auch die Distraktorentdeckungsparameter a und D (siehe Tabelle 8.5), können zwischen den Gruppen gleichgesetzt werden, ohne eine signifikante Verschlechterung der Modellanpassungsstatistik hervorzurufen (a : $\Delta G_{(1)}^2 = 2.87$; D : $\Delta G_{(1)}^2 = 0.26$).

³⁵Eine Reduktion auf spezifische Submodelle ist in diesem Experiment nicht sinnvoll durchführbar, da die Manipulationen zu unterschiedlichen Rateparametern innerhalb einer Gruppe führen sollten.

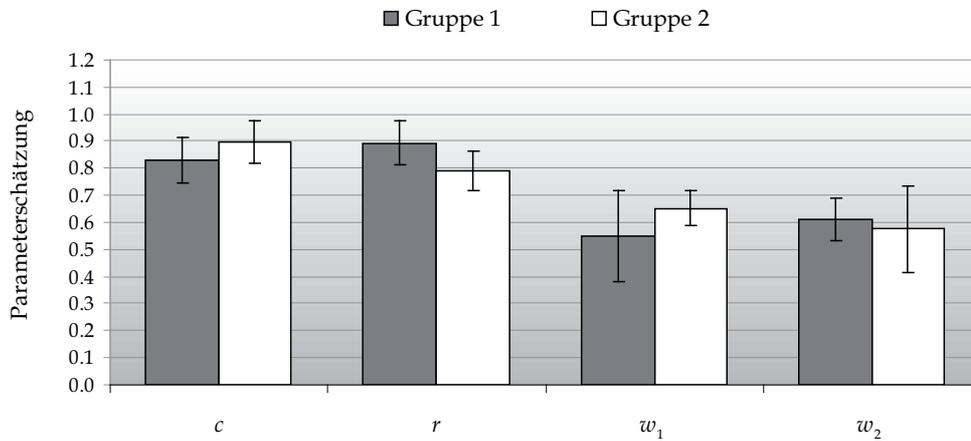


Abbildung 8.7: In dieser Abbildung sind die Parameterschätzungen der Gedächtnisparameter dargestellt. In Grau sind die Parameterschätzungen für Gruppe 1, in Weiß die korrespondierenden Schätzungen für Gruppe 2 abgebildet. Auch hier geben die Fehlerbalken die 95%-Konfidenzintervalle an.

Tabelle 8.5: Parameterschätzungen (Par. Schätzungen) des ADH-Modells sowie die 95%-Konfidenzintervalle (95% KI) in Abhängigkeit der beiden experimentellen Gruppen für die erste Lernliste

Parameter	Gruppe 1		Gruppe 2	
	Par.Schätzung	95% KI	Par.Schätzung	95% KI
c	0.83	0.75 – 0.92	0.90	0.82 – 0.97
r	0.89	0.81 – 0.98	0.79	0.72 – 0.86
w_1	0.55	0.38 – 0.72	0.65	0.59 – 0.72
w_2	0.61	0.53 – 0.69	0.58	0.42 – 0.73
g_{10}	0.52	0.34 – 0.70	0.40	0.29 – 0.51
g_{11}	0.59	0.43 – 0.76	0.35	0.25 – 0.46
g_{20}	0.32	0.20 – 0.45	0.57	0.40 – 0.73
g_{21}	0.43	0.30 – 0.56	0.64	0.51 – 0.77
g_3	0.18	0.11 – 0.25	0.10	0.05 – 0.15
D	0.48	0.32 – 0.63	0.43	0.29 – 0.56
a	0.28	0.00 – 0.56	0.00	-0.43 – 0.43

*: Signifikante Effekte zwischen den Gruppen auf einem α -Niveau von .05.

8.2.3 Diskussion

In Experiment 4 wurden die Rateparameter innerhalb der Versuchsperson manipuliert. Die Probanden erhielten für korrekte Antworten abhängig von der Itemposition unterschiedlich viele Pluspunkte, für falsche Antworten wurden differenziert Minuspunkte abgezogen. Die Auszahlungsmatrix wurde so aufgestellt, dass die Probanden für das linke Wort eine liberale Ratetendenz entwickeln sollten und für das rechte Wort eher eine konservative Ratetendenz (Gruppe 1). Für die zweite Gruppe wurde das umgekehrte Muster angestrebt. Die Motivation der Probanden, möglichst viele Punkte zu gewinnen, wurde dadurch geweckt, dass die Punkte in Süßigkeiten eingetauscht werden konnten.

Die Rateparameter zeigten ein hypothesenkonformes Muster, das jeweils zu hochsignifikanten Unterschieden zwischen den beiden experimentellen Gruppen führte. Damit konnte die Differenzierung der Rateparameter für das linke und rechte Wort validiert werden. Experiment 4 führte in diesem Punkt zu einem Erkenntnisgewinn, der über die Replikation von Experiment 3 hinausgeht.

Weitere diskriminante Effekte auf die Gedächtnisparameter oder die Parameter der Distraktorentdeckung blieben aus. Wie schon in Experiment 3 konnten die Rateparameter also selektiv beeinflusst werden. Dieses Ergebnismuster ist nach den Resultaten des zweiten Experiments von besonderer Wichtigkeit. In Experiment 2 wirkte sich eine Manipulation der Gedächtnisleistung auch auf die Rateparameter aus. Dieser Effekt konnte zwar theoretisch erklärt werden, um sicherzustellen, dass die Rate- und Gedächtnisparameter nicht konfundiert sind, war es aber notwendig zu zeigen, dass es möglich ist, die Rateparameter zu manipulieren, ohne Auswirkungen auf die Gedächtnisparameter zu finden.

In Experiment 3 und 4 gelang die erfolgreiche Validierung der Rateparameter. In beiden Experimenten wirkten sich die Manipulationen lediglich signifikant auf die Ratetendenzen und auf keine weiteren Parameter aus. Über zwei unterschiedliche Methoden konnte damit die Validität der Rateparameter überzeugend dargelegt werden.

Kapitel 9

Zusammenfassung und Diskussion der Validierungsexperimente

Nachdem in den letzten Kapiteln die vier Validierungsexperimente detailliert geschildert wurden, soll in diesem Kapitel ein kurzer, zusammenfassender Überblick über deren Ergebnisse im Hinblick auf die Ziele der empirischen Überprüfung des ADH-Modells gegeben werden. Dabei sollen die in Kapitel 5 genannten Kriterien (1. Identifizierbarkeit, 2. gute Modellanpassung sowie 3. Parametervalidierung) aufgegriffen werden.

Grundlegend ist zunächst die Identifizierbarkeit des Modells (1. Kriterium). In allen Experimenten wurde die lokale Identifizierbarkeit überprüft, und sie konnte in allen Datensätzen sichergestellt werden. Der Rang der Jacobischen Matrix entsprach jeweils der Anzahl der geschätzten Parameter und schließt damit die lokale Identifizierbarkeit nicht aus. Das mehrfache Aktivieren des EM-Algorithmus mit zufälligen Startwerten führte in allen Experimenten zu dem gleichen³⁶ Schätzvektor Ω bei identischen Modellanpassungsstatistiken. Dies zeigt, dass die lokale Identifizierbarkeit für alle Experimente gegeben war. Darüber hinaus wurde für zufällig gewählte Daten auch der Simulationsnachweis der Identifizierbarkeit (Erdfelder, 2000; Rouder & Batchelder, 1998) über multiTree (Moshagen, 2009) erbracht. Die durchschnittliche Abweichung bei einem Toleranzkriterium von 0.001 lag lediglich bei 0.0004. Auch bei einem identifizierten Modell sind minimale Abweichungen zu erwarten, da das Konvergenzkriterium größer Null sein muss und Rundungsfehler einer perfekten Übereinstimmung entgegenwirken. Es ist daher angesichts des Simulationsergebnisses wahrscheinlich, dass die lokale Identifizierbarkeit der vorliegenden Datensätze keine Ausnahmen darstellen und eine globale Identifizierbarkeit plausibel ist. In jedem Fall kann das Modell für die in dieser Arbeit diskutierten Fälle als identifizierbar angesehen werden.

³⁶Minimale Abweichungen waren zwar gegeben, können aber auf Rundungsfehler zurückgeführt werden.

Voraussetzung für die inhaltliche Interpretation der Parameter und weiterführende Parametertests ist neben der Identifizierbarkeit eine gute Anpassung des Modells an die Daten (2. Kriterium). Bisher wurde als Statistik der Modellanpassung der G^2 -Wert berichtet. Da mit zunehmender Anzahl von Datenpunkten die Wahrscheinlichkeit einer Modellablehnung aufgrund von geringen Modellverletzungen steigt, lagen in den Validierungsexperimenten jeweils sehr hohe Teststärken zur Entdeckung von kleinen Modellabweichungen ($w = .01$) vor. So war die Teststärke in Experiment 3 und 4 jeweils größer als .999. Daher wurde das α -Niveau für alle Experimente auf .001 festgelegt. Wie aus Tabelle 9.1 zu entnehmen ist, wurde mit diesem Kriterium in allen vier Experimenten eine gute Modellanpassung erreicht. Für das konventionelle α -Niveau von .05 war wurde in drei Experimenten eine Modellanpassung erreicht, in Experiment 3 allerdings nicht. Dies könnte im günstigsten Fall nur eine Folge der unterschiedlich großen Datenmengen sein, die jeweils in die Analyse einfließen.

Um diese Hypothese zu überprüfen, kann eine alternative Statistik für die Modellanpassung berechnet werden, welche die zugrundeliegende Datenmenge berücksichtigt. Dieses Kriterium wird u.a. von der BIC -Statistik (Bayesian Information Criterion, Schwarz, 1978) erfüllt. Die BIC -Statistik weist gegenüber der G^2 -Statistik den Vorteil auf, dass die Anzahl der Datenpunkte in deren Berechnung einfließt. Damit wird der Nachteil der hohen Sensitivität der G^2 -Statistik bei großen Datenmengen relativiert. Legt man die ΔBIC -Statistik³⁷ zugrunde, so wird das Modell unter der Nullhypothese dann beibehalten, wenn die Statistik negative Werte annimmt. Wie aus Tabelle 9.1 abgelesen werden kann, ist dies für alle vier Experimente der Fall. Die BIC -Werte reichen von -131.77 bis -264.40 . Diese Analysen zeigen, dass das Modell eine gute Modellanpassung aufweist und die schlechtere G^2 -Statistik nicht auf problematische Modellannahmen zurückzuführen ist.

Diese Schlussfolgerung kann auch über einen alternativen Weg demonstriert werden. Im Sinne eines Kompromisses zwischen α - und β -Fehler wird häufig eine Anpassung des α -Niveaus vorgenommen, um eine bestimmte Mindestteststärke ($1 - \beta$) zu garantieren (Faul, Erdfelder, Lang & Buchner, 2007). Dieses Verfahren kann auch im vorliegenden Fall angewandt werden. Ziel ist es dabei, das α -Niveau in allen Experimenten so anzupassen, dass eine vergleichbare Power für den χ^2 -Test der Modellanpassung vorliegt. Die Power, bereits kleine Modellabweichungen ($w = .01$) zu entdecken, wurde für alle Experimente auf .99 festgelegt. Um mit den zugrundeliegenden Datenmengen diese Teststärke zu erreichen, ist in Experiment 1 und 2 ein α -Niveau von .0009 bzw. .009 ausreichend. Da in Experiment 3 eine größere Datenmenge einfließt, kann hier eine Power von .99 mit einem α -Niveau von $5.84 \cdot 10^{-7}$ erreicht werden, bei Experiment 4 genügt ein α -Niveau von $7.78 \cdot 10^{-5}$. Dies zeigt, dass die schlechtere Modellanpassung in Experiment 3 angesichts der großen Datenmenge nicht als kritisch für das Modell angesehen werden muss. In Anbetracht der hohen

³⁷Der ΔBIC -Wert bezieht sich auf den Vergleich zu dem jeweiligen saturierten Modell ($\Delta BIC = PD^\lambda - \ln(N)df$) und wird verwendet, da er einfacher zu interpretieren ist als der absolute BIC -Wert.

Power kann also in allen vier Experimenten von einer zufriedenstellenden Modellanpassung ausgegangen werden.

Bevor zusammenfassend auf die Parameterschätzungen eingegangen wird, soll diskutiert werden, inwieweit diese als robust gegenüber potentiellen Verletzungen der Verteilungsannahmen eingestuft werden können. Die bisher berichteten Ergebnisse beruhen wesentlich auf den Verteilungsannahmen, die bei einer multinomialen Modellanalyse erfüllt sein müssen. Hierzu gehört z.B. dass allen Beobachtungen der Stichprobe die gleichen Kategoriewahrscheinlichkeiten zugrunde liegen und dass die Beobachtungen weder innerhalb noch zwischen den Gruppen eines Modells stochastisch voneinander abhängig sind. Es wird also angenommen, dass für alle Probanden die gleichen Parameterwerte resultieren (Parameterhomogenität). Allerdings wurde dies bisher nicht getestet. Tatsächlich kann eine Verletzung dieser Voraussetzung in psychologischen Gedächtnisstudien als nicht unwahrscheinlich angesehen werden, da trotz einer relativ homogenen Stichprobe interindividuelle Unterschiede in den kognitiven Fähigkeiten zu erwarten sind (für weitere Beispiele siehe Klauer, 2006). Fraglich ist, inwieweit die Parameterschätzungen des ADHD-Modells robust gegenüber solchen Differenzen sind.

Aus individuellen Unterschieden in den Parametern (Parameterheterogenität) könnte (a) eine schlechte Modellanpassung, (b) fehlerhafte Schätzungen der Konfidenzintervalle und (c) eine verzerrte Schätzung der Modellparameter folgen (Erdfelder, 2000; Klauer, 2006). Vorliegende Parameterheterogenität könnte den Validierungserfolg der berichteten Experimente also dann gefährden, wenn sich die Effekte der Manipulationen lediglich als Konstrukte von verzerrten Parameterschätzungen auf Aggregatsebene herausstellen würden. Es gilt daher zu überprüfen, ob Parameterheterogenität vorliegt und ob die Ergebnisse der Validierungsexperimente auch auf Ebene homogener Subgruppen beibehalten werden können.

Um das Problem der Parameterheterogenität zu identifizieren und einer Lösung zuzuführen, kann eine latente Klassenanalyse vorgenommen werden (Klauer, 2006). Einer solchen Analyse werden nicht die aggregierten Daten über alle Versuchspersonen zugrunde gelegt, sondern die individuellen Datenpunkte. Die Software *HmmTree* (Stahl & Klauer, 2007) ermöglicht auf einfachem Wege die Durchführung einer latenten Klassenanalyse. Zunächst können für die Ein-Klassen-Lösung Statistiken berechnet werden, die eine Beurteilung der Heterogenität erlauben (Klauer, 2006). Im Falle von Heterogenität kann dann eine Mehr-Klassen-Lösung berechnet werden, aus der Parameterschätzungen für die einzelnen Klassen resultieren.

Für die vorliegenden Experimente wurde mittels *HmmTree* eine Heterogenitätsanalyse durchgeführt (Erdfelder, 2000; Klauer, 2006). Das Hauptziel war dabei sicherzustellen, dass sich die Manipulationen der vier Experimente nicht nur auf Aggregatsebene in den Parametern niederschlagen, sondern auch auf Ebene von homogeneren Klassen.

Wie zu erwarten war, wurde in den Experimenten³⁸ Heterogenität festgestellt. Diese reduzierte sich bereits für eine Zwei-Klassenlösung deutlich.³⁹ In beiden Subklassen konnten die erwarteten Effekte im Sinne der konvergenten Validität repliziert werden. Die absoluten Werte der Parameter unterschieden sich zwischen den Klassen, der relative Unterschied war aber unabhängig von den absoluten Werten immer in der erwarteten Richtung vorhanden. Daher kann geschlussfolgert werden, dass die Effekte der Validierungsexperimente nicht künstlich durch die Aggregation mehrerer Subpopulationen entstanden, sondern dass sich auch innerhalb jeder Klassenlösung die erwarteten konvergenten Effekte zeigen. Es muss also eingeräumt werden, dass Parameterheterogenität vorliegt. Diese ist aber für die Schlussfolgerungen aus den Validierungsexperimenten und die Interpretation der Parameter nicht als kritisch anzusehen. Auch die Parameter auf Aggregatsebene können daher sinnvoll interpretiert werden.

Damit können die Parameterschätzungen der vier Validierungsexperimente nun im Hinblick auf das 3. Kriterium, die Parametervalidierung, beurteilt werden. In Tabelle 9.1 ist für jedes Experiment aufgelistet, auf welche Parameter die experimentelle Manipulation ausgerichtet war (Zielparameter; erste Zeile) und worin die experimentelle Manipulation bestand (zweite Zeile). In den letzten beiden Zeilen ist abgetragen, auf welche Parameter die Manipulation im Sinne der konvergenten Validität Einfluss nehmen sollte und welche Parameter im Sinne der diskriminanten Validität unbeeinflusst bleiben sollten.

Zusammenfassend konnten alle postulierten Hypothesen im Sinne der konvergenten Validität bestätigt werden. Bis auf eine Ausnahme konnten auch alle Nullhypothesen der diskriminanten Validität trotz ausreichender Teststärke beibehalten werden. Die Ausnahme betrifft Experiment 2. Hier schlugen sich die Effekte einer tiefen vs. flachen Enkodierungsbedingung sowie die Effekte des Testzeitpunkts nicht nur in den entsprechenden Gedächtnisparametern sondern auch in Rateparametern nieder. Allerdings konnte dieses Ergebnismuster Post-Hoc im Sinne des Mirror-Effekts sinnvoll interpretiert werden (siehe Kapitel 7) und stellt im Zusammenhang mit den Ergebnissen aus Experiment 3 und 4 (in denen es gelang, die Rateparameter unabhängig von den Gedächtnisparametern zu beeinflussen) kein Problem für die Validierung dar. In der Tabelle ist diese Abweichung von den Hypothesen gekennzeichnet (*). Alle anderen Hypothesen konnten bestätigt werden.

³⁸Für eine Heterogenitätsanalyse mit der Software HmmTree ist es Voraussetzung, dass von jedem Probanden gleich viele Datenpunkte vorliegen. Daher konnten nur für die Experimente 2 bis 4 Heterogenitätsanalysen vorgenommen werden. In Experiment 1 wurde die Manipulation der unterschiedlich langen Darbietungszeiten innerhalb der Probanden operationalisiert; durch die randomisierte Zusammenstellung der Testpaare resultierten für jeden Probanden leicht unterschiedliche Kategoriehäufigkeiten. Da aber sowohl die Items als auch die Probanden aller Experimente aus derselben Grundpopulation stammten, ist nicht zu erwarten, dass für Experiment 1 andere Ergebnisse resultieren als für die übrigen Experimente.

³⁹Da in der Zwei-Klassen-Lösung meist etwa gleich große Klassen resultierten, wohingegen bei der Drei-Klassen-Lösung in eine Klasse nur sehr wenige Probanden einfließen, wurde an der Lösung mit reliablen Parameterschätzungen in allen Klassen festgehalten, obwohl sie die Heterogenität nicht perfekt beseitigen konnte.

Tabelle 9.1: Zusammenfassung der vier Validierungsexperimente

	Validierungsexperiment				
	Exp. 1b	Exp. 2		Exp. 3	Exp. 4
Zielparameter	w_1, w_2	c	r	$g_{10}, g_{11},$ g_{20}, g_{21}	$g_{1j},$ g_{2j}
Exp. Manipulation	Präs.zeit kurz/lang	Enkodierung tief/flach	Ret.Intervall 3 min/24 h	Basisrate Testwörter	Aus.matrix lib./kons.
G^2	31.13	48.48		31.38	21.58
df	31	36		18	18
p	.45	.07		.02	.25
α^*	.0009	.009		$5.84 \cdot 10^{-7}$	$7.78 \cdot 10^{-5}$
BIC	-242.72	-264.40		-131.77	-137.89
Lok. Identifiz.	gegeben	gegeben	gegeben	gegeben	gegeben
Erw. Effekte					
(a) Konv. Val.	$w_1, w_2,$ c	$c, a,$ w_1, w_2	$r,$ w_1, w_2	$g_{10}, g_{11},$ g_{20}, g_{21}	$g_{1j},$ g_{2j}
(b) Diskr. Val.	$r, g_{ij},$ $a, D, g\mathfrak{3}$	$r, g_{ij}^*,$ $D, g\mathfrak{3}$	$c, g_{ij}^*,$ $a, D, g\mathfrak{3}$	c, r, w_i $a, D, g\mathfrak{3}$	c, r, w_i $a, D, g\mathfrak{3}$

* Auf diesem α -Niveau ist jeweils eine Teststärke von .99 zur Entdeckung von kleinen Modellabweichungen ($w = .01$), gegeben.

In den vorliegenden vier Experimenten konnten alle Gedächtnisparameter (c , r , w_1 , w_2), der Parameter der Distraktorentdeckung für Assoziationen (a) sowie die Itemrateparameter (g_{10} , g_{11} , g_{20} , g_{21}) validiert werden. Die einzigen Parameter, deren Validierung noch aussteht ist der Rateparameter für Assoziationen (g_3) und der Distraktorentdeckungsparameter (D). Der Rateparameter g_3 könnte validiert werden, indem man das Verhältnis von alten intakten Wortpaaren zu neu kombinierten Wortpaaren verändert (vergleiche Kapitel 8.1.1 für eine analoge Manipulation auf Itemebene). Auch eine Auszahlungsmatrix die nicht wie in Experiment 4 auf die Rateparameter für Items, sondern auf den Rateparameter für die Kombination ausgelegt ist, sollte eine Validierung ermöglichen. Der Distraktorentdeckungsparameter D könnte über die Ähnlichkeit von Distraktoren und gelernten Wörtern beeinflusst werden (siehe Bayen et al., 1996).

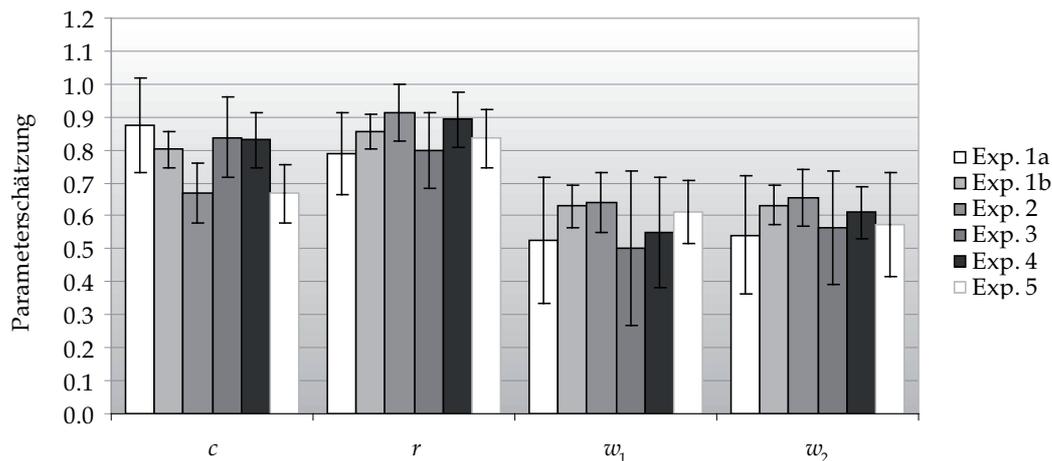


Abbildung 9.1: In dieser Abbildung sind die Gedächtnisparameter der Kontrollgruppen im Vergleich zwischen den vier Experimenten abgebildet. Die Fehlerbalken geben die 95%-Konfidenzintervalle an.

Hinsichtlich der spezifischeren Submodelle ergab sich kein eindeutiges Ergebnis. Eine einheitliche Reduktion auf eines der vorgestellten Submodelle war über alle experimentellen Bedingungen hinweg nicht möglich. Es zeichnete sich aber bis auf wenige Ausnahmen (Exp. 2) das Bild ab, dass eine Reduktion der Rateparameter aufgrund der vorausgehenden oder folgenden Rateprozesse (Prozessparameter, Modell 9a) sinnvoller erscheint als eine Reduktion aufgrund der Position des Wortes, für das geraten wird. Anhand der vorliegenden Experimente soll an dieser Stelle das Fazit gezogen werden, dass es gerechtfertigt

ist an der validierten 11-Parameter-Lösung festzuhalten. Diese Modellvariante ist identifizierbar, die Parameter konnten validiert werden und wiesen sinnvolle Schätzungen auf. Es besteht daher keine dringende Notwendigkeit, auf eines der Submodelle auszuweichen.

Abschließend sind in Abbildung 9.1 die zentralen Gedächtnisparameter des Modells über die Kontrollgruppen der vier Validierungsexperimente hinweg abgebildet. Vorausgreifend wurden auch die Parameterschätzungen für Experiment 5 in diesen Vergleich aufgenommen. Dabei zeigt sich, dass vergleichbare, reliable Parameterschätzungen über die Experimente hinweg erreicht werden konnten.

Zusammenfassend kann also festgehalten werden, dass das ADH-Modell in der Serie von Validierungsexperimenten durchweg eine zufriedenstellende Modellanpassung aufwies und die zentralen Parameter des Modells erfolgreich validiert werden konnten, auch im Hinblick auf das strenge Kriterium der diskriminanten und konvergenten Validität. Rückblickend auf die in Kapitel 4.2 formulierten Anforderungen an das Modell kann festgehalten werden, dass mit dem ADH-Modell folgende Ziele erreicht wurden:

- (1) Sowohl die Gedächtnisleistung für einzelne Items (w_1, w_2) als auch die Erinnerung an die Assoziation zweier Items (c, r) kann erfasst werden.
- (2) Die Erinnerungsleistung an die Assoziation zweier Items wird auf die Subprozesse der assoziativen Enkodierung (c) und des assoziativen Abrufs (r) zurückgeführt.
- (3) Das Modell erlaubt die Analyse von Rekognitionsdaten und
- (4) an das Itemmaterial werden keine besonderen Anforderungen gestellt. Bisher wurden zufällig kombinierte hochfrequente Wörter verwendet, das Modell erlaubt aber prinzipiell auch die Analyse von z.B. visuellem Material oder auch assoziierten visuellen oder verbalen Stimuli.

Da das ADH-Modell sowohl die inhaltlichen als auch die methodischen Kriterien erfüllt, kann es nun zur Untersuchung von substanzwissenschaftlichen Fragestellungen verwendet werden. Insbesondere ist es mit diesem Modell möglich, die Lokalisationsfrage des assoziativen Defizits älterer Menschen zu beantworten. Auf dieses Ziel wird im nächsten Kapitel eingegangen.

Teil III

Gerontologische Modell Anwendung

Kapitel 10

Experiment 5

Im vorangegangenen Teil der Arbeit wurde das ADH-Modell erfolgreich validiert: In vier Experimenten war die lokale Identifizierbarkeit gegeben, die Modellanpassung war zufriedenstellend und die Parameter zeigten die aufgrund der experimentellen Manipulationen erwarteten Effekte. Daher kann das Modell nun zur Datenanalyse in einer kognitiven Altersstudie genutzt werden.

Das Ziel dieses Altersvergleichs ist es herauszufinden, inwieweit das assoziative Defizit auf Enkodierungs- oder Abrufprozesse zurückgeht. Bisher wurde das assoziative Defizit über den Vergleich eines Itemerkennungstests und eines assoziativen Erkennungstests erfasst. In der Analyse der korrigierten Erkennungsleistungen (Treffer minus falsche Alarme) zeigte sich dabei eine signifikante Wechselwirkung zwischen der Testart und der Altersgruppe. Diese Interaktion beruhte auf einer Leistungsveränderung der älteren Menschen. Während die Senioren im Itemtest eine mit den Jüngeren vergleichbare Leistung aufwiesen, waren sie in der assoziativen Aufgabe signifikant schlechter: Sie zeigten ein spezifisches Gedächtnisdefizit für assoziiertes Material.

Dieses Defizit sollte sich auch mit dem neuen kombinierten Item- und assoziativen Erkennungstest zeigen lassen, da der Alterseffekt nicht von einem spezifischen Testparadigma abhängen sollte. Anders als in den Validierungsexperimenten wird daher auch die varianzanalytische Auswertung der korrigierten Erkennungsleistung vorgenommen, um das klassische Ergebnismuster zu replizieren. In einem zweiten Schritt wird über die multinomiale Auswertung eine genauere Lokalisation des Altersunterschieds ermöglicht.

10.1 Methode

Um in einer Altersstudie den maximalen Anteil des Gruppenunterschieds auf das biologische Alter zurückführen zu können und den Einfluss von konfundierenden Faktoren wie unter anderem Gesundheits- und Bildungsunterschieden minimal zu halten, sollten die Altersgruppen in Hinsicht auf diese Faktoren parallelisiert sein. Dies wurde auch in Ka-

pitel 2 genauer begründet und diskutiert. In Anlehnung an die Standards der kognitiven Altersforschung wurden daher in diesem Experiment mehrere Befragungen bzw. Tests durchgeführt, um relevante Kontrollvariablen zu erheben: Die objektive Gesundheit wurde über einen Gesundheitsfragebogen erfasst, der vor allem altersbedingte Krankheiten sowie Krankheiten, die das Erinnerungsvermögen beeinflussen, erfragt. Zum Ausschluss aus der Datenerhebung führten dabei Erkrankungen wie Schlaganfall, Herzinfarkt, Diabetes, Depressionen, ärztlich nicht kontrollierter Bluthochdruck, Hirnschädigungen oder längere Bewusstlosigkeit. Diese Kriterien galten auch für die jüngeren Teilnehmer. Da in der vorliegenden Studie verbale Materialien verwendet wurden, sollte des Weiteren sichergestellt werden, dass beide Altersgruppen ein vergleichbares verbales Verständnis bzw. eine vergleichbare verbale Intelligenz aufweisen. Es wurde daher eine Wortschatztest (SASKA, Riegel, 1967) durchgeführt. Um die Beschuldungsdauer, sowie die in höheren Bildungsinstitutionen verbrachte Zeit zwischen den Probandengruppen vergleichen zu können, wurde der Schulabschluss sowie die Ausbildungs- bzw. Studienzzeit erfragt. Wie erwähnt entspricht die Erfassung dieser Variablen den Standards der kognitiven Altersforschung.

Auch bei der Rekrutierung und Datenerhebung wurde einerseits darauf geachtet, zwischen den Altersgruppen identische Bedingungen sicherzustellen, andererseits die Bedingungen aber so zu gestalten, dass die älteren Probanden nicht benachteiligt werden. So wurde unter anderem bei der Rekrutierung die Aktivierung bestimmter Altersstereotype vermieden (siehe Rahhal et al., 2001, Kapitel 2). Eine ausreichend lange Übungsphase und der Verzicht auf die Computermaus sollten die geringere Vertrautheit älterer Menschen mit Computereperimenten ausgleichen. Die Schriftgröße und die Position der Materialien auf dem Bildschirm wurden so angepasst, dass sie auch für ältere Menschen gut wahrnehmbar waren. Mit diesen Maßnahmen sollte sichergestellt werden, dass ein möglichst großer Anteil der Altersvarianz auf die durch das biologische Alter verursachten Gruppenunterschiede zurückgeführt werden kann.

In den bisher durchgeführten Altersexperimenten zur Assoziationsdefizit-Hypothese zeigte sich, dass der Alterseffekt unter anderem von bestimmten Eigenschaften des Wortmaterials moderiert wird. Naveh-Benjamin (2000, Experiment 4) sowie Naveh-Benjamin, Hussain et al. (2003) manipulierten den semantischen Bezug des Wortmaterials. In beiden Experimenten wurde das typische Assoziationsdefizit nur repliziert, wenn die Wörter keinen semantischen Bezug zueinander aufwiesen (siehe Kapitel 3.2.2). Aus diesem Grund sollte sichergestellt werden, dass die in dieser Altersstudie verwendeten Wortpaare nicht semantisch miteinander assoziiert sind. Die Wörter wurden daher nicht zufällig zu Wortpaaren zusammengestellt, sondern es wurde ein mehrstufiger Vortest durchgeführt, in dem überprüft wurde, ob in den verwendeten Wortkombinationen Wörter vorkamen, die einen offensichtlichen Bezug zueinander erkennen lassen. Genauer ist das Verfahren im Abschnitt „Material“ beschrieben.

Ein weiterer Faktor, der den Alterseffekt moderiert, sind die Instruktionen in der Lern- und Testphase. Um einen möglichst großen Alterseffekt zu erfassen, war die Lernphase

intentional. Durch einen Übungsdurchgang war auch die Art des Tests bekannt. Weder in der Lern- noch in der Testphase wurden irgendwelche Hinweise auf Strategien gegeben. Im Unterschied zu einigen Altersexperimenten von Naveh-Benjamin sollte die Lernsituation in diesem Experiment für beide Altersgruppen absolut vergleichbar sein. Daher wurde die Präsentationszeit nicht zwischen den Altersgruppen variiert, sondern eine konstante Präsentationszeit verwendet. Insgesamt wurde also zum einen das Experiment so gestaltet, dass das assoziative Defizit möglichst durch keine Randbedingung verringert wird, zum anderen wurde versucht, vergleichbare Altersstichproben zu erheben, um einen möglichst klaren Rückschluss auf den Altersfaktor zu erlauben.

Teilnehmer und Design

An diesem Experiment nahmen 40 Studenten der Universität Mannheim teil. Zwei Teilnehmer wurden ausgeschlossen, da sie im Gesundheitsfragebogen angaben, unter Depressionen zu leiden. Ein Teilnehmer wurde ausgeschlossen, da er eine länger andauernde Bewusstlosigkeit aufgrund einer starken Gehirnerschütterung berichtete. Bei einem weiteren Teilnehmer ging aus technischen Gründen der Datensatz verloren. Daher konnten die Daten von 36 Studenten ausgewertet werden (Durchschnittsalter 25.14 Jahre). Die Studenten wurden über Versuchspersonenstunden für ihre Teilnahme entlohnt.

Die älteren Probanden wurden größtenteils im Zusammenhang mit einer Evaluation des Seniorenstudiums der Universität Mannheim rekrutiert. Ein Teil dieser Evaluation bestand darin, die geistige Aktivität der Seniorstudenten über verschiedene kognitive Tests zu erfassen, sowie ein Evaluationsgespräch zu führen. Im Rahmen dieser Laborstudie wurden die Probanden im Vorhinein gefragt, ob sie Interesse hätten, an einem weiteren Experiment teilzunehmen. Die Evaluationsstudie und das vorliegende Altersexperiment wurden dann zusammen durchgeführt. Einige zusätzliche Probanden wurden aus einer Versuchspersonendatenbank des Lehrstuhls für Psychologie III rekrutiert. Diese Teilnehmer waren zu einem früheren Zeitpunkt über Anzeigen oder Handzettel geworben worden.

Die auf diesem Weg rekrutierte Stichprobe bestand aus 42 älteren Menschen. Von den älteren Probanden mussten 8 aus der Analyse ausgeschlossen werden, da sie eines oder mehrere der folgenden gesundheitlichen Probleme berichteten: Diabetes, Schlaganfall, Hirnschädigungen oder Depressionen. Zwei Probanden wurden ausgeschlossen, da sie angaben, Probleme beim Lesen und Verstehen der Materialien gehabt zu haben. Die 32 in der Stichprobe verbleibenden älteren Menschen hatten ein Durchschnittsalter von 66.53 Jahren, der jüngste Proband war 48 Jahre, der älteste 83 Jahre, die Standardabweichung lag bei 6.81 Jahren. Die Stichprobe kann damit als eher „jüngere“ ältere Stichprobe eingeordnet werden. Von den Teilnehmern waren 84.4% Seniorenstudenten der Universität Mannheim. Die älteren Probanden erhielten für die Teilnahme eine Aufwandsentschädigung von 10 Euro.

Im Hinblick auf die Kontrollvariablen wurde durch den Ausschluss von Probanden mit objektiven Gesundheitsproblemen bereits eine erste Parallelisierung der Altersgruppen vorgenommen. Mit durchschnittlich 17.78 Bildungsjahren ($SD = 6.25$) wies die ältere Probandengruppe einen vergleichbaren Wert zu der jüngeren Stichprobe auf (18.36 Bildungsjahre, $SD = 4,50$), hier wurde kein signifikanter Unterschied festgestellt, $t(66) = 0.44, \eta_p^2 = .003, p = .66$. Im verbalen Test schnitten die älteren Probanden mit einem Mittelwert von 17.41 ($SD = 1.79$) deskriptiv besser ab als die jüngeren Probanden ($M = 14.56, SD = 2.32$). Dieser Unterschied wurde auch im inferenzstatistischen Vergleich signifikant, $t(66) = -5.61, \eta_p^2 = .32, p = .16$.

Damit kann festgehalten werden, dass die beiden Stichproben im Hinblick auf ihre Bildungsjahre und ihre objektive Gesundheit als parallelisierte Gruppen angesehen werden können. Vergleichbar zu anderen Altersstudien zeigten die älteren Menschen eine bessere Fähigkeit im Umgang mit verbalen Materialien. Dieser signifikante Effekt zugunsten der Senioren ist für die Untersuchung keineswegs problematisch, da er eher zu einer konservativen Unterschätzung des Alterseffekts führt.

In diesem Quasi-Experiment variierte lediglich das Alter zwischen den Probanden. Es wurden keine weiteren experimentellen Manipulationen durchgeführt.

Ablauf

Alle Probanden wurden zunächst über den Ablauf und das Ziel der Studie aufgeklärt. Wie in den Validierungsexperimenten wurde auch dieses Experiment an Intel-Pentium(4)-Rechnern und einem 19-Zoll-Monitor durchgeführt. Den Probanden wurden dieselben Instruktionen erteilt wie in den Validierungsstudien. Sie wurden darauf hingewiesen, dass ihnen in dieser Aufgabe Wortpaare gezeigt werden, die sie sich einprägen sollen, da sie später sowohl nach den einzelnen Wörtern als auch nach der Kombination der Wörter gefragt werden. Auf diese Instruktionen folgte ein Übungsdurchgang, in dem sechs Wortpaare präsentiert und anschließend in den verschiedenen Testkombinationen abgefragt wurden. Anders als in den bisherigen Studien hatten die Teilnehmer die Möglichkeit, den Übungsdurchgang mehrmals zu wiederholen. Wie in den bereits durchgeführten Studien stand während der gesamten Studie der Versuchsleiter für Rückfragen zur Verfügung.

Die Wörter wurden in derselben Schriftgröße präsentiert wie in den Validierungsexperimenten. Eine Anpassung auf eine altersgerechte Darstellung der experimentellen Materialien war nicht erforderlich, da die Schriftgröße in den Validierungsexperimenten mit Blick auf die anschließende Altersstudie bereits sehr groß gewählt wurde. Lediglich die Instruktionen wurden nun etwas größer dargestellt als in den Validierungsexperimenten.

Auf den Übungsdurchgang folgte die Präsentation der ersten Lernliste, nach einem Retentionsintervall von 3 Minuten wurde diese Liste getestet. In der Lernliste wurden nach 3 Füll-Wortpaaren 50 Wortpaare für je fünf Sekunden dargeboten, im Test kamen 70 Wortpaare vor. Die Listen hatten damit eine vergleichbare Länge zu den ersten zwei

Experimenten dieser Arbeit. Auch die verschiedenen Wortkombinationen im Test wurden äquivalent zu den ersten beiden Experimenten zusammengestellt.⁴⁰ In der Testphase mussten die Probanden ihre Antworten über das sequentielle Ja-nein-Format abgeben (vgl. Experiment 1b, Kapitel 6). Vergleichbar zu den Validierungsexperimenten wurden die Antwortoptionen auf dem Bildschirm angezeigt. Die Antworten konnten über farblich korrespondierende Tasten abgegeben werden. Zum Ausgleich von Händigkeitseffekten wurde die Tastenbelegung ausbalanciert.

Nach der ersten Testphase wurden die Probanden gebeten, den Gesundheitsfragebogen auszufüllen. Dieser enthielt sowohl demographische Angaben als auch die Fragen zur objektiven Gesundheit. Des Weiteren wurde der Wortschatztest, die Synonym-Skala des SASKA-Tests (Riegel, 1967) durchgeführt. Danach folgten verschiedene kognitive Tests, die im Rahmen der Seniorenstudiumsevaluation durchgeführt wurden. Nach circa 25 Minuten fand die zweite Lern- und Testphase statt. Am Ende des Experiments wurden die Probanden über die Zielsetzung der Studie informiert, und sie hatten Gelegenheit, dem Versuchsleiter ihre Fragen zu stellen.

Material

Wie bereits erwähnt wurde, sollte für diese Studie sichergestellt werden, dass keine offensichtlich assoziierten Wortpaare in der Lernphase vorkommen. Daher wurden zunächst zwei randomisierte Lernlisten mit je 50 Wortpaaren sowie pro Lernliste zwei randomisierte Testlisten mit je 70 Wortpaaren aus den Wortlisten A und B (siehe Anhang C) erstellt. Diese Wortpaare wurden in einem Vortest von mehreren Probanden⁴¹ auf ihren semantischen, phonologischen und orthographischen Zusammenhang eingestuft. Die Teilnehmer sollten die Wortpaare auf einer 5-stufigen Skala spontan als „sehr schwach“, „eher schwach“, „neutral“, „eher stark“ oder „sehr stark“ assoziiert einschätzen. Alle Wortpaare, die über die Beurteiler hinweg durchschnittlich als „eher stark“ oder „sehr stark“ assoziiert bezeichnet wurden, wurden neu kombiniert und anschließend in weiteren Vortests überprüft. Es wurden drei Phasen des Vortests durchgeführt, an denen insgesamt 32 Personen teilnahmen. Die semantische Unabhängigkeit des endgültigen Wortmaterials wurde von allen 18 Probanden der letzten Testphase zwischen „sehr schwach assoziiert“ bis „neutral“ eingestuft. Das Wortmaterial kann Anhang E entnommen werden. Zwischen den Probanden wurde ausbalanciert, welche Liste als erste Lernliste verwendet wurde sowie welche der jeweils verfügbaren Testlisten genutzt wurde.

⁴⁰Es wurden 20 alte intakte Wortpaare, 20 neu kombinierte Wortpaare sowie jeweils 10 Target-Distraktor-Paare, Distraktor-Target-Kombinationen und Distraktor-Distraktor-Kombinationen getestet.

⁴¹Keiner der Probanden der Vorstudie nahm an Experiment 5 teil.

10.2 Ergebnisse

Der Ergebnisteil dieses Experiments gliedert sich in zwei Abschnitte. Zunächst wird die von Naveh-Benjamin und Kollegen vorgenommene varianzanalytische Auswertung berichtet. Hier ist vor allem von Interesse, ob es mit dem neuen Paradigma gelang, das typische assoziative Defizit zu replizieren. Danach wird die multinomiale Analyse berichtet. Dabei werden zunächst kurz die zugrundeliegenden Häufigkeiten besprochen, dann wird auf die Parametervergleiche eingegangen.

Ein Ergebnis der Auswertung soll an dieser Stelle bereits vorweggenommen werden, da es den Aufbau dieses Ergebnisteils beeinflusst: Das ADH-Modell konnte nicht auf die aggregierten Daten beider Lernlisten angewandt werden, dies führte zu einer schlechten Modellanpassung. An die einzelnen Listen zeigte das Modell aber eine gute Anpassung. Dieser Befund könnte auf eine Veränderung der Parameterschätzungen zwischen den Listen zurückgehen. Diese könnte zum Beispiel durch die Nutzung anderer Lernstrategien nach der Erfahrung mit der ersten Lernliste hervorgerufen werden. Eine genauere Erörterung dieser Befunde wird in der Diskussion vorgenommen. Im Moment soll festgehalten werden, dass aus diesem Grund die Auswertung und Ergebnisdarstellung für das Experiment 5 für die erste und zweite Liste getrennt berichtet wird.

Varianzanalytische Auswertung

In den bisher veröffentlichten Arbeiten zur assoziativen Rekognition wurden für die Itemgedächtnisleistung und für die assoziative Gedächtnisleistung jeweils Trefferraten (Hits: alte Items/Wortpaare, die korrekt als alt klassifiziert wurden) korrigiert um falsche Alarme (FA: neue Items/Wortpaare, die fälschlicherweise als alt klassifiziert wurden) als Maß der Gedächtnisleistung berichtet. Eine vergleichbare Analyse soll auch für das vorliegende Experiment durchgeführt werden, da zunächst über diese Auswertung das assoziative Defizit repliziert werden sollte.

Für den neuen kombinierten Item- und assoziativen Rekognitionstest kann die Auswertung der *assoziativen Rekognitionsleistung* analog zu den traditionellen Analysen erfolgen (Buchler et al., 2008; Castel & Craik, 2003; Naveh-Benjamin, 2000). Die korrekte Klassifikation eines alten Wortpaares wird als Treffer gewertet. Als falscher Alarm wird angesehen, wenn ein neu kombiniertes Wortpaar als altes, intaktes Wortpaar eingestuft wird.

Für die *Itemrekognition* kann die Auswertung nicht direkt vergleichbar zu den bisherigen Analysen erfolgen. In Anlehnung an Castel und Craik (2003) werden als Treffer für das linke Wort alle Antworten gewertet, die dieses Item als alt klassifizieren, unabhängig davon, wie das rechte Item oder die Assoziation eingestuft wird. Treffer für das linke Item sind also unter anderem diejenigen Antworten, die eine Target-Distraktor-Kombination entweder (1) als altes intaktes Wortpaar, (2) als neu kombiniertes Wortpaar oder (3) als Target-Distraktor-Kombination bewerten. Als falsche Alarme fließen unter anderem die

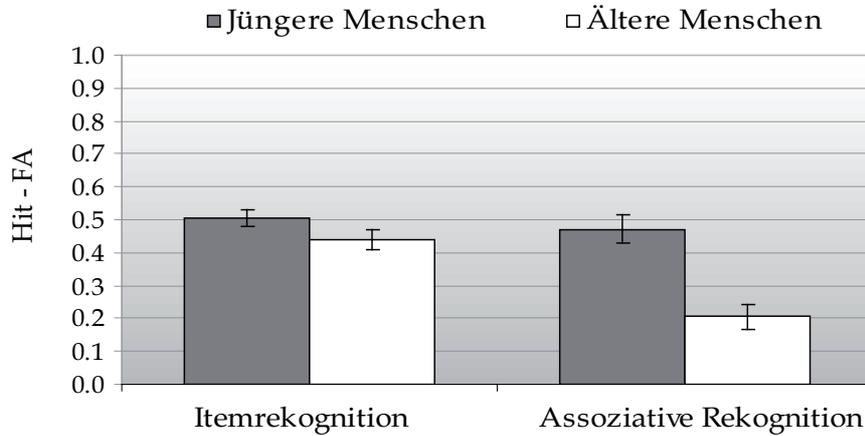


Abbildung 10.1: Als Maß für die Gedächtnisleistung in der ersten Liste sind hier die korrekten Klassifikationen abzüglich der falschen Alarme (Hit-FA) jeweils für den Itemtest und den assoziativen Rekognitionstest abgetragen. Die Fehlerbalken geben den Standardfehler des Mittelwertes an.

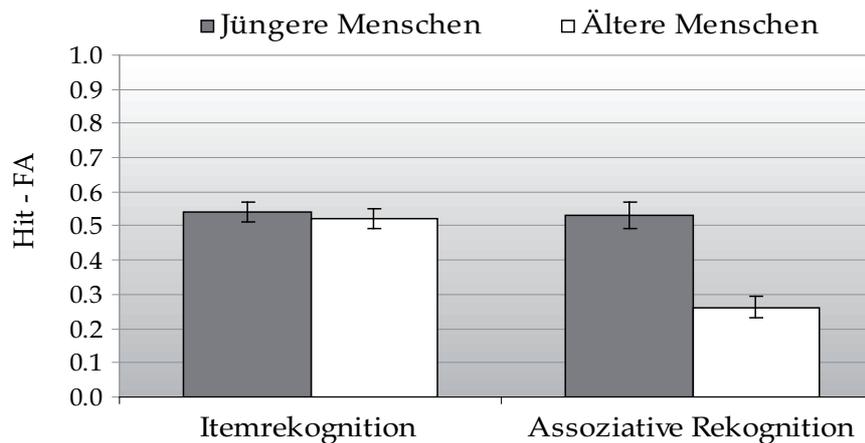


Abbildung 10.2: Als Maß für die Gedächtnisleistung in der zweiten Liste sind die korrekten Klassifikationen abzüglich der falschen Alarme (Hit-FA) jeweils für den Itemtest und den assoziativen Rekognitionstest abgetragen. Die Fehlerbalken geben den Standardfehler des Mittelwertes an.

Häufigkeiten ein, mit denen eine Distraktor-Distraktor-Kombination fälschlicherweise als Target-Distraktor-Kombination angesehen wird. Äquivalent wurde für das rechte Wort verfahren.

Abbildung 10.1 und 10.2 zeigen getrennt für die beiden Lern-Testlisten die Analyse der Treffer abzüglich der falschen Alarme (Hit-FA) jeweils für die assoziative Komponente der Aufgabe und die Itemgedächtnisleistung.

Zunächst werden die Ergebnisse der ersten Testliste besprochen: Bei den jüngeren Probanden ist deskriptiv kein Unterschied zwischen der Itemrekognitionsleistung ($M = .50, SEM = .03$) und der assoziativen Rekognitionsleistung ($M = .45, SEM = .04$) festzustellen. Dies entspricht dem Ergebnismuster der bisherigen Studien zum assoziativen Defizit. Im Itemrekognitionstest unterscheiden sich die älteren Probanden mit einem Mittelwert von $.44$ ($SEM = .03$) kaum von den jüngeren Teilnehmern ($.50$). Im assoziativen Test zeigt sich bei den älteren Menschen mit einem Wert von $.21$ ($SEM = .04$) ein starker Leistungseinbruch gegenüber den Studenten ($.45$) sowie gegenüber der Itemrekognitionsleistung ($= .44$). Dieses Muster deutet das assoziative Defizit älterer Menschen an.

In der zweiten Testliste bleibt dieses Muster bestehen. Jüngere Menschen zeigen in der Itemrekognitionsaufgabe ($M = .54, SEM = .03$) und dem assoziativen Rekognitionstest ($M = .53, SEM = .04$) kaum unterschiedliche Leistungen. Die Itemrekognitionsleistung der Älteren ist mit einem Mittelwert von $.52$ ($SEM = .03$) vergleichbar zur Rekognitionsleistung der Studenten ($.54$), fällt aber in der assoziativen Rekognitionsaufgabe deutlich ab ($M = .26, SEM = .03$).

Um auch inferenzstatistisch den direkten Vergleich zu den Studien von Naveh-Benjamin zu ziehen, wurde eine Varianzanalyse mit dem Zwischensubjektfaktor Alter und der Testart (Itemrekognition vs. assoziative Rekognition) als Messwiederholungsfaktor gerechnet. Für die erste Liste zeigte diese Analyse zwei signifikante Haupteffekte, sowohl für das Alter, $F(1, 66) = 17.20, MSE = 0.93, \eta_p^2 = .21, p = .00$, als auch für die Testart, $F(1, 66) = 21.25, MSE = 0.59, \eta_p^2 = .24, p = .00$. Viel wichtiger ist jedoch die signifikante Wechselwirkung zwischen den Faktoren, $F(1, 66) = 12.23, MSE = 0.34, \eta_p^2 = .16, p = .00$, denn diese zeigt auch inferenzstatistisch das klassische Assoziationsdefizit. Die Ergebnisse für die zweite Testliste fielen vergleichbar aus. Auch hier wurde ein signifikanter Haupteffekt des Alters, $F(1, 66) = 11.96, MSE = 0.70, \eta_p^2 = .15, p = .00$, ein signifikanter Haupteffekt der Testart, $F(1, 66) = 29.32, MSE = 0.61, \eta_p^2 = .31, p = .00$ sowie eine signifikante Wechselwirkung, $F(1, 66) = 25.24, MSE = 0.53, \eta_p^2 = .28, p = .00$ festgestellt.

Neben dieser klassischen zweifaktoriellen Varianzanalyse könnten mit den vorliegenden Daten noch spezifischere Hypothesen über geplante Kontraste gerechnet werden. Da die varianzanalytische Auswertung aber nur zur Replikation des typischen Effekts durchgeführt wurde, wird auf diese weitergehenden Analysen verzichtet, denn der eigentliche Fokus liegt auf der multinomialen Analyse.

Multinomiale Analyse

Die Häufigkeiten, die der multinomialen Analyse zugrunde liegen, können aus der Tabelle 10.1 entnommen werden. Im Altersvergleich deutet sich ein Unterschied in den korrekten Klassifikationsraten an. Diese liegen bei jüngeren Menschen in der ersten Liste für alte intakte Wortpaare bei .603, für ältere Menschen bei .423. In der zweiten Wortliste wird aus derselben Itemklasse ein Anteil von .621 Wörtern bzw. .506 von jüngeren respektive älteren Menschen richtig klassifiziert. Während also jüngere Menschen eine höhere korrekte Klassifikationsrate für die Assoziation zweier alter Wörter aufweisen, so findet sich das umgekehrte Muster in der Rate der falschen Alarme: Jüngere Menschen beurteilen neu kombinierte Wortpaare nur mit einer Rate von .131 in der ersten und .090 in der zweiten Liste fälschlicherweise als altes intaktes Wortpaar, bei älteren Menschen ist dies mit .217 bzw. .244 häufiger der Fall.

In den Antwortmustern für die Itemerkennung findet sich ebenfalls eine erhöhte Trefferrate, allerdings sind hier die Raten falscher Alarme kaum unterschiedlich zwischen den Altersgruppen. Beispielhaft wird dies an den Target-Distraktor-Kombinationen veranschaulicht: Für Target-Distraktor-Kombinationen liegt die strenge Rate richtiger Klassifikationen (d.h. als korrekt wird nur die richtige Klassifikation beider Wörter gewertet) für jüngere Menschen bei .442 (Liste 1) und .442 (Liste 2), bei älteren fallen die entsprechenden Werte mit .316 und .313 niedriger aus. Dafür sind die Raten der falschen Alarme in der ersten Liste bei Jüngeren (.167) kaum niedriger als bei Älteren (.184), in der zweiten Liste zeigen beide Altersgruppen mit einem Wert von .225 identische Raten falscher Alarme.

Im nächsten Auswertungsschritt wurden diese Häufigkeiten als Datengrundlage der multinomialen Modellierung verwendet. Für die aggregierten Daten beider Listen zeigte sich dabei eine schlechte Anpassung, die Modellanpassungsstatistik lag mit $G_{(18)}^2 = 52.82$ über dem kritischen Wert, $G_{krit,\alpha=.001}^2 = 42.31$. Die beiden Listen wurden daher getrennt ausgewertet, und mit $G_{(18)}^2 = 27.12$ für die erste Liste und $G_{(18)}^2 = 35.78$ für die zweite Liste konnte eine gute Anpassung gewährleistet werden. Eine mögliche Erklärung für die schlechte Anpassung der aggregierten Daten wird in der Diskussion besprochen.

Die Parameterschätzungen für die Gedächtnisparameter der ersten Liste sind in Abbildung 10.3 abgetragen. Deskriptiv sind zwischen allen vier Parameterschätzungen Altersunterschiede festzustellen, die Parameterwerte der älteren Stichprobe fallen tendenziell niedriger aus. Die Enkodierungsparameter unterscheiden sich mit $\Delta G_{(1)}^2 = 11.69$ signifikant zwischen den Altersgruppen, der Unterschied im Abrufparameter wird nicht signifikant, $\Delta G_{(1)}^2 = 3.30$. Für die beiden Itemerkennungsparameter ergibt sich ein uneinheitliches Bild. Während sich die Erkennungsleistung für das linke Wort signifikant zwischen den beiden Altersgruppen unterscheidet ($\Delta G_{(1)}^2 = 3.91$), können die Erkennungsparameter für das rechte Wort ohne eine signifikante Modellverletzung mit $\Delta G_{(1)}^2 = 0.90$ gleichgesetzt werden.

Tabelle 10.1: *Relative Antworthäufigkeiten der Probanden auf die fünf unterschiedlichen Itemkombinationen in der ersten und zweiten Testliste*

(a) Liste 1: Antworthäufigkeiten jüngerer Menschen

Wahre Itemkat.	Antwort der Versuchsperson				
	Altes WP	Neu komb. WP	Target Dis.	Dis. Target	Dis.Dis.
Altes WP	.603	.154	.069	.101	.072
Neu komb. WP	.131	.447	.179	.167	.076
Target Dis.	.058	.208	.422	.133	.178
Dis. Target	.064	.150	.086	.511	.189
Dis. Dis.	.019	.117	.167	.239	.458

(b) Liste 1: Antworthäufigkeiten älterer Menschen

Wahre Itemkat.	Antwort der Versuchsperson				
	Altes WP	Neu komb. WP	Target Dis.	Dis. Target	Dis.Dis.
Altes WP	.423	.170	.119	.164	.123
Neu komb. WP	.217	.277	.203	.188	.116
Target Dis.	.122	.169	.316	.163	.231
Dis. Target	.119	.178	.094	.416	.194
Dis. Dis.	.050	.081	.184	.256	.428

(c) Liste 2: Antworthäufigkeiten jüngerer Menschen

Wahre Itemkat.	Antwort der Versuchsperson				
	Altes WP	Neu komb. WP	Target Dis.	Dis. Target	Dis.Dis.
Altes WP	.621	.138	.083	.081	.078
Neu komb. WP	.090	.485	.178	.156	.092
Target Dis.	.056	.233	.442	.092	.178
Dis. Target	.081	.222	.100	.428	.169
Dis. Dis.	.033	.133	.225	.156	.453

(d) Liste 2: Antworthäufigkeiten älterer Menschen

Wahre Itemkat.	Antwort der Versuchsperson				
	Altes WP	Neu komb. WP	Target Dis.	Dis. Target	Dis.Dis.
Altes WP	.506	.200	.097	.122	.075
Neu komb. WP	.244	.369	.163	.152	.073
Target Dis.	.141	.281	.313	.128	.138
Dis. Target	.163	.234	.106	.353	.144
Dis. Dis.	.075	.197	.225	.216	.288

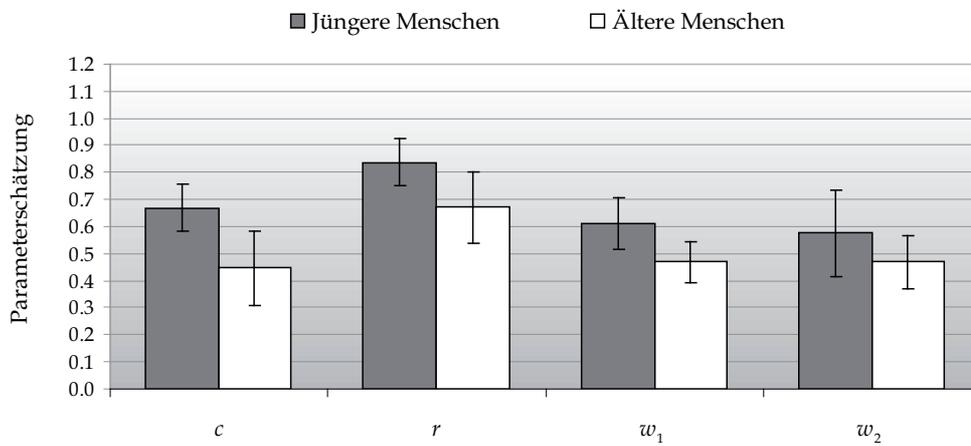


Abbildung 10.3: In dieser Abbildung sind die Schätzungen der Gedächtnisparameter des ADH-Modells für jüngere und ältere Menschen vergleichend abgetragen. Die Fehlerbalken geben die 95%-Konfidenzintervalle an. Die Werte beziehen sich auf die erste Lern- bzw. Testliste.

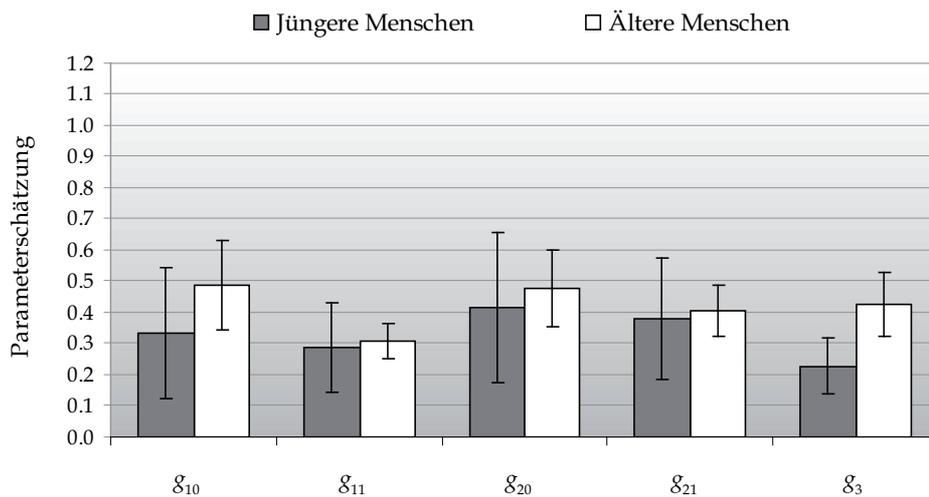


Abbildung 10.4: In dieser Grafik sind die Parameterschätzungen für die Rateparameter dargestellt. In Grau sind jeweils die Schätzungen für jüngere Menschen, in Weiß die Schätzungen für ältere Menschen abgebildet. Die Werte beziehen sich lediglich auf den ersten Lern-Testdurchgang. Die Fehlerbalken stellen die 95%-Konfidenzintervalle dar.

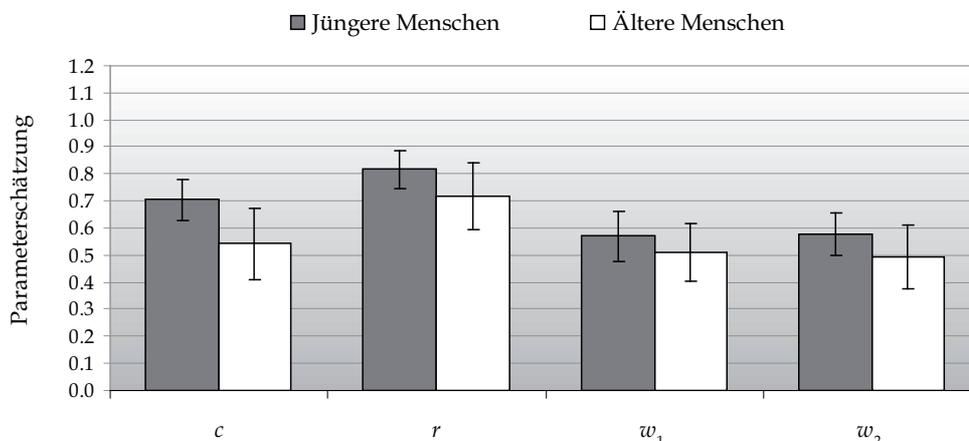


Abbildung 10.5: Diese Abbildung zeigt die Parameterschätzungen für die Gedächtnisparameter in Abhängigkeit der beiden Altersgruppen. Die 95%-Konfidenzintervalle sind als Fehlerbalken dargestellt.

In Abbildung 10.4 sind die Rateparameter des ADH-Modells dargestellt (Liste 1). Tendenziell zeigen die jüngeren Probanden konservativere Ratetendenzen als die älteren Menschen. Signifikant unterscheidet sich keiner der vier Itemrateparameter zwischen den beiden Gruppen (g_{10} : $\Delta G^2_{(1)} = 1.65$; g_{11} : $\Delta G^2_{(1)} = 0.25$; g_{20} : $\Delta G^2_{(1)} = 0.11$; g_{21} : $\Delta G^2_{(1)} = 0.10$). Der Rateparameter für die Assoziation hingegen fällt bei den jüngeren Menschen signifikant geringer aus als in der älteren Stichprobe (g_3 : $\Delta G^2_{(1)} = 17.72$). Alle Parameterschätzungen sowie die Konfidenzintervalle sind gemeinsam mit den Ergebnissen der zweiten Testliste in Tabelle 10.2 abgetragen.

Die Gedächtnisparameter der zweiten Liste weisen deskriptiv ein ähnliches Bild auf wie die Parameter der ersten Liste, wieder sind die Schätzungen für die jüngere Stichprobe etwas höher als die Schätzungen der älteren Stichprobe (siehe Abbildung 10.5). Vergleichbar zur ersten Liste ist ein signifikanter Alterseffekt im Enkodierungsparameter zu verzeichnen, $\Delta G^2_{(1)} = 6.63$. Auch in Übereinstimmung mit der ersten Liste führt die Gleichsetzung des Abrufparameters nicht zu einer signifikanten Verschlechterung der Modellanpassung ($\Delta G^2_{(1)} = 2.03$). Für die Itemrekognitionsparameter ergibt sich ebenfalls ein vergleichbares Ergebnismuster, ein Itemrekognitionsparameter unterscheidet sich knapp aber signifikant zwischen den Gruppen, der andere nicht. Allerdings wird in dieser Liste der Altersunterschied für den Rekognitionsparameter des rechten Wortes signifikant, $\Delta G^2_{(1)} = 3.98$, der Rekognitionsparameter des linken Wortes weist keinen signifikanten Unterschied auf, $\Delta G^2_{(1)} = 2.13$.

Die Rateparameter, dargestellt in Abbildung 10.6, zeigen mit einer konservativeren Ratetendenz bei den jüngeren Probanden ein vergleichbares Muster zu den Rateparametern der ersten Liste, wobei die Rateparameter insgesamt etwas höher ausfallen als in der ersten Liste. Dieser Effekt ist bei den älteren Probanden besonders deutlich. In dieser Al-

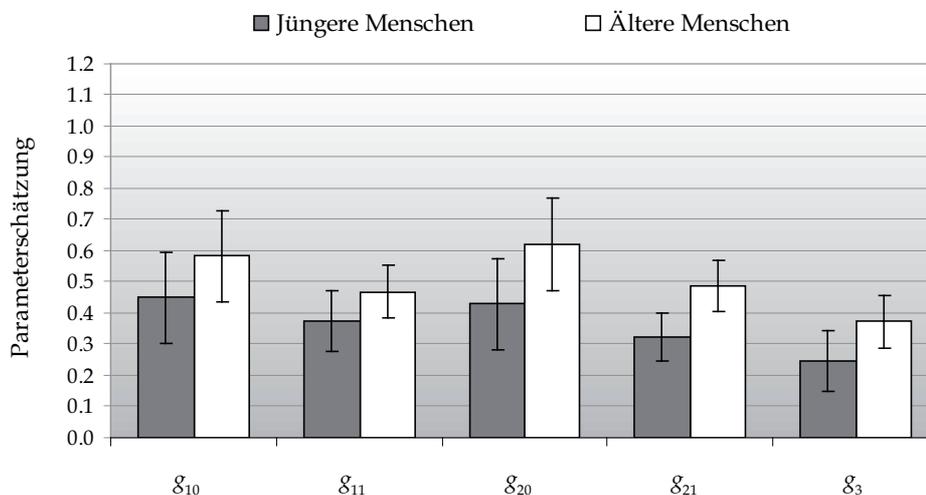


Abbildung 10.6: Die Rateparameter der zweiten Lern- bzw. Testliste sind in dieser Grafik jeweils für die beiden Altersgruppen dargestellt. Die Fehlerbalken geben die 95%-Konfidenzintervalle an.

tersgruppe sind die Rateparameter in der zweiten Liste durchschnittlich um 0.12 erhöht. Für den Vergleich zwischen den Altersgruppen hat dies zur Folge, dass die Gleichsetzung eines jeden Itemrateparameters zu einem signifikanten Effekt führt (g_{10} : $\Delta G_{(1)}^2 = 4.95$; g_{11} : $\Delta G_{(1)}^2 = 9.65$; g_{20} : $\Delta G_{(1)}^2 = 5.73$; g_{21} : $\Delta G_{(1)}^2 = 11.16$). Auch der Rateparameter für die Assoziation kann zwischen den Gruppen nicht gleichgesetzt werden, dies führt mit $\Delta G_{(1)}^2 = 6.18$ zu einer signifikant schlechteren Modellanpassung.

Die Parameter der Distraktorentdeckung wurden aufgrund der getrennten Auswertung der beiden Listen nur unreliabel geschätzt und weisen weder deskriptiv besondere Effekte auf, noch sind inferenzstatistisch signifikante Gruppenunterschiede festzustellen, (alle $\Delta G_{(1)}^2 < 3.84$). Genauer wird auf diese Parameterschätzungen daher nicht eingegangen.

Zusammenfassend kann festgehalten werden, dass über die varianzanalytische Auswertung das assoziative Defizit repliziert werden konnte. Vergleichbar zu den Experimenten von Naveh-Benjamin zeigte sich eine signifikante Wechselwirkung zwischen Altersgruppe und Testart. Dieser Effekt wird durch einen signifikant größeren Altersunterschied in der assoziativen Rekognitionsleistung hervorgerufen. In der multinomialen Analyse konnte dieses assoziative Defizit spezifischer analysiert und auf einen signifikanten Unterschied im assoziativen Enkodierungsparameter, nicht aber im Abrufparameter zurückgeführt werden. Des Weiteren unterschied sich das Rateverhalten der beiden Altersgruppen in der ersten Liste nicht signifikant, in der zweiten Liste tendierten die älteren Probanden zu einem liberaleren Rateverhalten als die jüngeren Teilnehmer.

Tabelle 10.2: Parameterschätzungen (Par. Schätzungen) des ADH-Modells sowie die 95%-Konfidenzintervalle (95% KI) in Abhängigkeit der beiden Altersgruppen für Liste 1 (Tabelle a) sowie 2 (Tabelle b)

(a) Liste 1

Parameter	Jüngere Menschen			Ältere Menschen	
	Par.Schätzung	95% KI		Par.Schätzung	95% KI
c	0.67	0.58 – 0.75	*	0.44	0.31 – 0.58
r	0.84	0.75 – 0.92		0.67	0.54 – 0.80
w_1	0.61	0.52 – 0.71	*	0.47	0.39 – 0.54
w_2	0.58	0.42 – 0.73		0.47	0.37 – 0.57
g_{10}	0.33	0.13 – 0.54		0.49	0.35 – 0.63
g_{11}	0.29	0.14 – 0.43		0.31	0.25 – 0.37
g_{20}	0.42	0.18 – 0.66		0.48	0.35 – 0.60
g_{21}	0.38	0.19 – 0.58		0.41	0.32 – 0.49
g_3	0.23	0.14 – 0.32	*	0.42	0.32 – 0.53
D	0.00	–0.55 – 0.55		0.00	–0.19 – 0.19
a	0.00	–0.29 – 0.29		0.00	–0.22 – 0.22

(b) Liste 2

Parameter	Jüngere Menschen			Ältere Menschen	
	Par.Schätzung	95% KI		Par.Schätzung	95% KI
c	0.71	0.63 – 0.78	*	0.54	0.41 – 0.67
r	0.82	0.75 – 0.88		0.72	0.59 – 0.84
w_1	0.57	0.48 – 0.66		0.51	0.40 – 0.62
w_2	0.58	0.50 – 0.66	*	0.49	0.37 – 0.61
g_{10}	0.45	0.30 – 0.59	*	0.58	0.44 – 0.73
g_{11}	0.37	0.28 – 0.47	*	0.47	0.39 – 0.55
g_{20}	0.43	0.28 – 0.58	*	0.62	0.47 – 0.77
g_{21}	0.32	0.24 – 0.40	*	0.49	0.40 – 0.57
g_3	0.25	0.15 – 0.34	*	0.37	0.29 – 0.46
D	0.00	–0.27 – 0.27		0.00	–0.18 – 0.18
a	0.23	–0.02 – 0.49		0.00	–0.21 – 0.21

*: Signifikante Effekte zwischen den Gruppen auf einem α -Niveau von .05.

10.3 Diskussion

In diesem Experiment wurde das kombinierte Item- und assoziative Rekognitionsparadigma zum ersten Mal in einer Altersstudie angewandt. Auch die Datenanalyse auf Basis des ADH-Modells erfolgte bisher noch nicht für den Altersvergleich. Das Experiment war im Hinblick auf das Wortmaterial (Wörter die keine offensichtliche Assoziation zueinander aufweisen) und die intentionale Lernsituation vergleichbar zu den meisten Experimenten von Naveh-Benjamin. Anderes als in manchen Experimenten aus dieser Forschergruppe lagen für beide Altersgruppen dieselben Lernbedingungen vor: Alle Teilnehmer hatten fünf Sekunden Zeit, jedes Wortpaar zu enkodieren. Insgesamt wurden zwei Listen gelernt und getestet.

Zunächst wurde eine Analyse der korrigierten Rekognitionsleistung vorgenommen. Dazu wurden die Raten falscher Alarme von den Trefferraten abgezogen. Dieses Maß wurde zum einen für die Itemrekognitionsleistung, zum anderen für die assoziative Rekognitionsleistung berechnet. Bei dieser Auswertung zeigte sich sowohl deskriptiv als auch inferenzstatistisch eine perfekte Replikation des assoziativen Defizits. Die Ergebnisse tragen damit zur Generalisierung dieses Defizits bei, da es in einem neuen Paradigma repliziert werden konnte.

Eine differentielle Analyse des Altersvergleichs von Treffen und falschen Alarmen zeigte, dass ältere Menschen in der assoziativen Rekognitionsaufgabe eine höhere Rate von falschen Alarmen aufwiesen als jüngere Menschen. Das Muster der Trefferraten war genau umgekehrt. Dieses Ergebnis ist vergleichbar mit den Analysen von Castel und Craik (2003), Naveh-Benjamin et al. (2009) und Shing et al. (2008): Auch sie berichten bei älteren Menschen neben der herabgesetzten Trefferrate vor allem ein Ansteigen in den falschen Alarmen.

In der multinomialen Datenanalyse zeigte sich zunächst eine schlechte Modellanpassungsstatistik. Aufgrund der Vermutung, dass sich bei älteren Probanden, die weniger Erfahrung mit Lernsituationen⁴² haben, zum einen positive Übungs- und Erfahrungseffekte oder zum anderen negative Interferenzeffekte in der zweiten Lernliste zeigen könnten, wurden daher die beiden Listen getrennt voneinander analysiert. Tatsächlich war für jede einzelne Lern-Testliste eine gute Modellanpassung vorhanden. Dies bedeutet, dass einige Parameter zwischen den Listen nicht gleichgesetzt werden können, ohne eine signifikant schlechtere Modellanpassung hervorzurufen. Diese Unterschiede könnten durch Interferenzeffekte der ersten Lernliste ausgelöst werden. Interferenzeffekte sollten zu einer schlechteren Gedächtnisleistung reflektiert durch die Parameter c , r , w_1 , w_2 führen. Andererseits ist es auch denkbar, dass die Probanden durch die vorangegangene Lernerfahrung in der zweiten

⁴²Zwar weisen die an dieser Studie teilnehmenden Seniorenstudenten sicherlich eine größere Erfahrung mit Lernsituationen auf als eine bevölkerungsrepräsentative Stichprobe, trotzdem kann davon ausgegangen werden, dass auch bei dieser selektiven Stichprobe die jüngeren Probanden mit Lernsituationen und Computereperimenten vertrauter sind.

Liste andere Strategien anwenden und dies zu einer besseren Gedächtnisleistung und einer Veränderung der Rateparameter führt. Ein Unterschied in den Rateparametern ist auch vorstellbar, da die Teilnehmer nach dem ersten Lern-Testdurchgang besser abschätzen können, welche Verhältnisse von alten zu neuen Wörtern den Testlisten zugrunde liegen.

Der Parametervergleich zwischen den Lernlisten zeigte, dass bei den jüngeren Probanden in den Gedächtnisparametern keinerlei Effekte auftraten, auch die Rateparameter wiesen nur leichte Schwankungen auf. Die älteren Probanden allerdings konnten ihre Enkodierungsleistung von $c = .44$ auf $c = .54$ steigern. Die anderen Gedächtnisparameter zeigten keine nennenswerten Veränderungen. Die Itemrateparameter wiesen in der zweiten Testliste insgesamt höhere Werte auf. Im Durchschnitt veränderten sich die Werte um $.12$, die kleinste Veränderung lag bei $.08$, die größte bei $.14$. Die älteren Probanden tendierten in der zweiten Liste also dazu, liberaler zu raten als in der ersten Liste. Dies könnte auf den hohen Anteil alter Wörter in der Testliste zurückzuführen sein, der eventuell von den älteren Probanden nicht erwartet wurde.

Die schlechte Modellanpassung beider Testlisten ist also zum einen auf eine Verbesserung der Enkodierungsleistung, zum anderen auf Unterschiede in den Rateparametern zurückzuführen. Dieses Muster spricht für strategische Lern- und Erfahrungseffekte der älteren Probanden.

In der spezifischeren multinomialen Analyse konnte das assoziative Defizit genauer analysiert werden. Dabei zeigte sich in der ersten Lernliste ein signifikanter Altersunterschied in der assoziativen Enkodierung, allerdings kein Alterseffekt im assoziativen Abruf. Dieses Muster konnte in der zweiten Testliste repliziert werden. Wieder war ein hoch signifikanter Effekt im Parameter c zu verzeichnen. Der Parameter r hingegen konnte zwischen den Gruppen gleichgesetzt, werden ohne eine signifikante Modellverletzung hervorzurufen.

Dieses Ergebnismuster ist im Einklang mit den Befunden von Mitchell et al. (2000), die ebenfalls ein Enkodierungsdefizit als Ursache des assoziativen Defizits ausmachen konnten. Allerdings wäre es eine Fehlinterpretation der vorliegenden Ergebnisse, das Enkodierungsdefizit zur alleinigen Ursache des Altersunterschieds zu erklären. Eine Teststärkenanalyse zeigte, dass in der ersten Liste eine ausreichende Power ($1 - \beta > .75$) nur vorlag, um einen Effekt von mindestens $.30$ im Abrufparameter zu entdecken. Für kleinere Effekte war die Teststärke nicht zufriedenstellend ($1 - \beta < .75$). Für die Teststärke in der zweiten Liste ergaben sich ähnlich Ergebnisse, auch hier war die Teststärke nur für Effekte von mindestens $.30$ befriedigend. Es muss daher berücksichtigt werden, dass aufgrund der vorliegenden Ergebnisse kein abschließendes Fazit über eine Abrufbeteiligung am assoziativen Defizit getroffen werden kann.

Die geringe Power für den Abrufparameter ist vor allem auf zwei Ursachen zurückzuführen. Zum einen konnten aufgrund einer schlechten Modellanpassung die beiden Testlisten nicht gemeinsam ausgewertet werden. Die getrennte Analyse der beiden Listen halbiert aber die Datenpunkte, die in die Auswertung einfließen und verursacht damit einen großen

Verlust in der Teststärke. Zum anderen war aufgrund des Enkodierungsdefizits die Schätzung des c -Parameters in der älteren Stichprobe nicht sehr hoch, so dass auch dadurch weniger Datenpunkte in die Schätzung des nachfolgenden r -Parameters einfließen.

Für zukünftige Studien ist es daher zu empfehlen, einen längeren Übungsdurchgang aufzunehmen und in der Testphase des Übungsdurchgangs bereits darauf zu achten, dass der Anteil alter Wörter im Test vergleichbar ist mit den folgenden Testlisten. Alternativ könnten nicht nur zwei Listen, sondern mehrere Listen verwendet werden. Dies sollte dazu führen, dass mindestens die zweite und dritte Liste aggregiert werden können. Für den Fall, dass keinerlei Hypothesen im Hinblick auf den Enkodierungsparameter vorliegen, kann über eine längere Präsentationsdauer auch der Enkodierungsparameter und damit der Anteil der Wörter, die in die Schätzung des Abrufparameters eingehen, erhöht werden. Bei diesem Vorgehen ist zu befürchten, dass für die jüngeren Probanden Deckeneffekte auftreten, daher sollte der Enkodierungsparameter bei einem solchen Experiment nur mit Vorsicht interpretiert werden.

Die Tatsache, dass in der Itemerkennung jeweils nur in einem Parameter ein signifikanter Unterschied zu verzeichnen war, soll in Kombination mit einer Teststärkeanalyse diskutiert werden. Für den linken Itemerkennungsparameter lag in der ersten Liste eine Teststärke von .83 für einen Effekt von .20 vor, in der zweiten Liste betrug die Teststärke für dieselbe Effektgröße .92. Die Werte für den rechten Itemerkennungsparameter fielen ähnlich aus. Es war also eine zufriedenstellende Teststärke für kleine Effekte gegeben. Daher kann der signifikante Altersunterschied in jeweils einem Parameter darauf zurückgeführt werden, dass es sich hierbei lediglich um geringe Alterseffekte handelt und diese wahrscheinlich mehr oder weniger zufällig einmal im rechten und einmal im linken Wort zur Signifikanz geführt haben. Dieser Befund steht in Einklang mit anderen Studien zum assoziativen Defizit und den Erkenntnissen zu Altersunterschieden in Erkognitionsexperimenten, denn die Alterseffekte in der Itemerkennung sind generell relativ schwach ausgeprägt.

Zusammenfassend kann dieses erste Altersexperiment, das mit dem kombinierten Item- und assoziativen Erkognitionsparadigma durchgeführt wurde und dessen Daten mit dem multinomialen ADH-Modell ausgewertet wurden, als Erfolg angesehen werden. Das assoziative Defizit konnte über die korrigierte Item- und assoziative Erkognitionleistung sowie die varianzanalytische Auswertung repliziert werden. Die spezifische multinomiale Auswertung zeigte darüber hinaus, dass ein Enkodierungsdefizit dem assoziativen Defizit zugrunde liegt. Ältere Menschen haben im Gegensatz zu jüngeren Menschen bereits in der Lernphase Schwierigkeiten, die beiden Wörter miteinander zu assoziieren. Inwieweit ein zusätzliches Abrufdefizit zu dem Altersunterschied beiträgt, konnte aufgrund dieses ersten Experiments nicht beantwortet werden. Die Daten deuten aber auf einen zusätzlichen Effekt des assoziativen Abrufs hin. Zukünftige Altersstudien sollten zunächst das Abrufdefizit genauer untersuchen. Des Weiteren wäre es dann aufschlussreich zu analysieren,

inwieweit strategische Trainingsphasen älteren Menschen helfen können, das assoziative Defizit zu überwinden. In diesem Kontext ist zu erwarten, dass sich der Alterseffekt zwar verringert, trotzdem aber erhalten bleibt, dies konnte von Shing et al. (2008) gezeigt werden. Spezifisch ermöglicht es die multinomiale Analyse allerdings zu zeigen, auf welche Prozesse das assoziative Defizit nach einem Strategietraining zurückzuführen ist. Auch im Hinblick auf einen Lebensspannenvergleich des assoziativen Defizits kann die multinomiale Analyse darüber Aufschluss geben, inwieweit dem Defizit bei kleinen Kindern dieselben Prozesse zugrunde liegen wie einer gerontologischen Stichprobe.

Teil IV

Diskussion

Kapitel 11

Abschließende Diskussion und Ausblick

Diese Arbeit beschäftigte sich mit der Assoziationsdefizit-Hypothese des kognitiven Alterns (Naveh-Benjamin, 2000) und der Frage nach der genauen Lokalisation dieses altersbedingten assoziativen Gedächtnisdefizits. Um herauszufinden, inwieweit dieses auf Enkodierungsprozesse und/oder Abrufprozesse zurückzuführen ist, wurde ein mathematisches Modell entwickelt, das es ermöglicht, die beiden kognitiven Prozesse zu dekomponieren. Die Modellparameter wurden in vier Experimenten erfolgreich validiert, danach wurde das Modell in einer Altersstudie angewandt, um auf die zentrale Lokalisationsfrage einzugehen. Es zeigte sich ein klares Enkodierungsdefizit, ein abschließendes Urteil über ein zusätzliches Abrufdefizit kann noch nicht getroffen werden.

In diesem Kapitel werden die wichtigsten Erkenntnisse, Ergebnisse und Implikationen dieser Arbeit zusammengefasst. Über die spezifische Diskussion der empirischen Validierung (Kapitel 9) und der gerontologischen Modellanwendung (Kapitel 10.3) hinaus wird die gesamte Arbeit kritisch betrachtet, Grenzen der Arbeit werden aufgezeigt und neue Forschungsperspektiven eröffnet.

11.1 Die Lokalisationsfrage des assoziativen Defizits

Assoziative Gedächtnisprozesse sind für die erfolgreiche Bewältigung verschiedenster kognitiver Aufgaben notwendig. Auch episodische Gedächtnisaufgaben beruhen auf assoziativen Prozessen, da eine Verbindung zu dem zeitlich räumlichen Kontext geschaffen werden muss. Das bei älteren Menschen auftretende Defizit in assoziativen Gedächtnisprozessen (Naveh-Benjamin, 2000) kann daher als Erklärung für deren schlechtere episodische Gedächtnisleistung herangezogen werden. Zahlreiche Experimente konnten die zentrale Rolle assoziativer Gedächtnisprozesse zur Erklärung von Altersunterschieden in episodischen Gedächtnisaufgaben bereits empirisch nachweisen (Bayen, Phelps & Spaniol, 2000; Chalfonte

& Johnson, 1996; Mitchell et al., 2000; Naveh-Benjamin, 2000). Naveh-Benjamin (2000) formulierte daraufhin die Assoziationsdefizit-Hypothese, deren zentrale Aussage folgendes Zitat am besten wiedergibt: „This hypothesis claimed that an important component of older adults’ poorer episodic memory is their deficiency in creating and retrieving links between single units of information. The basic units can be two items; an item and its context; two contextual elements; or, more generally, the representation of two mental codes.“ (Naveh-Benjamin, 2000, S. 1170).

Obwohl dieses assoziative Defizit als grundlegender Basismechanismus diskutiert wird, ist bisher unklar, ob assoziative Gedächtnisprozesse bereits während der Enkodierung scheitern und die Repräsentation einer Assoziation gar nicht angelegt wird, oder ob das Altersdefizit erst beim Abruf der assoziativen Informationen entsteht. Zwar wurde diese Frage schon häufig diskutiert (z.B. Cohn et al., 2008; Naveh-Benjamin, Hussain et al., 2003) und Naveh-Benjamin geht in der Definition des assoziativen Defizits von einem Problem in beiden Prozesse aus („creating and retrieving“). Bisher konnte aber kein Nachweis für diese Vermutung erbracht werden. Dies hängt unter anderem damit zusammen, dass es sich bei dieser Fragestellung um ein Dekomponierungsproblem zweier nicht beobachtbarer kognitiver Prozesse handelt (Erdfelder, 2000), und ein adäquater Zugang damit nur über die mathematische Modellierung der Prozesse ermöglicht wird.

In dieser Arbeit wurde daher ein multinomiales Modell entwickelt, welches erlaubt, die beiden Prozesse der assoziativen Enkodierung und des assoziativen Abrufs zu dekomponieren. Im Folgenden wird zunächst das Modell kritisch diskutiert, dann wird auf die Anwendungsgebiete des Modells sowie die gewonnen Erkenntnisse aus der gerontologischen Untersuchung eingegangen.

11.2 Das ADH-Modell: Eine kritische Evaluation

Kritisch kann man zunächst hinterfragen, ob die Entwicklung eines neuen Modells notwendig ist, um die vorliegende Forschungsfrage zu beantworten. Optimalerweise sollte versucht werden, auf existierende Ansätze zurückzugreifen und nicht für jede Fragestellung ein neues Modell zu entwerfen. Daher wurde in dieser Arbeit zunächst festgelegt, welchen Kriterien ein Modell genügen muss, um die Lokalisationsfrage zu beantworten. Anhand dieser Kriterien wurde sorgfältig geprüft, inwieweit existierende Modelle diese Voraussetzungen erfüllen. Leider lag kein Modell vor, das für ein Rekognitionsparadigma die Analyse der entsprechenden Prozesse erlaubt. Daher wurde ein neues Modell in Anlehnung an bereits existierende und erfolgreich angewandte Modelle (Enkodierung-Abruf-Modell, Batchelder & Riefer, 1986; Quellengedächtnismodelle, Batchelder & Riefer, 1990; Bayen et al., 1996; Meiser & Bröder, 2002) entwickelt.

Vergleichbar zu dem Enkodierung-Abruf-Modell enthält das Modell einen Parameter c , der die Wahrscheinlichkeit repräsentiert, dass ein Wortpaar in der Lernphase als Asso-

ziation (d.h. Cluster) enkodiert wird. Mit der Wahrscheinlichkeit r kann eine solche Assoziation auch erfolgreich abgerufen werden. Neben diesen assoziativen Parametern enthält das Modell auch zwei Itemerkennungparameter (w_1, w_2). Falls ein Wortpaar nicht als Assoziation enkodiert wurde, ist es möglich, dass beide Einzelwörter getrennt voneinander enkodiert, gespeichert und abgerufen werden, diese Wahrscheinlichkeit wird über w_1 für das linke Wort, über w_2 für das rechte Wort repräsentiert. Vergleichbar zu den Quellengedächtnismodellen (Batchelder & Riefer, 1990; Bayen et al., 1996; Meiser & Bröder, 2002), die ebenfalls für ein Erkennungsparadigma konstruiert wurden, sind auch Rateparameter, sowie Parameter der Distraktorentdeckung in dem Modell enthalten. Dabei wird zwischen der Distraktorentdeckung eines Items (D) und der Distraktorentdeckung einer Assoziation (a) unterschieden. Insgesamt besteht das ADH-Modell aus den vier oben genannten Gedächtnisparametern, zwei Distraktorentdeckungsparametern und fünf Rateparametern.

Jedes Modell basiert implizit auf bestimmten Annahmen, die kritisch hinterfragt und diskutiert werden können. Auch das ADH-Modell wurde aufbauend auf theoretischen Annahmen entworfen. So postuliert das Modell unter anderem, dass nach der erfolgreichen assoziativen Enkodierung der beiden Wörter entweder die Assoziation als Ganze, oder aber keines der beiden Wörter abgerufen werden kann. Gegen diese Art der Gedächtnisrepräsentation sprechen unter anderem die Ergebnisse von Buchler et al. (2008) sowie einige Modellierungsansätze (SAC-Modell, Reder et al., 2000; TODAM, Murdock, 1982). Andererseits wurde diese Annahme aus dem Enkodierung-Abruf-Modell übernommen und auch von theoretischer Seite findet das Konzept einer integrativen Speicherung zweier Wörter als Wortpaar Unterstützung (z.B. Fragment-Theorie von Jones, 1984). Während sich aus theoretischer Sicht sowohl Argumente für und gegen diese Art der Gedächtnisrepräsentation bzw. des Abrufs aus dem Gedächtnis finden lassen, so liegt mindestens ein empirisches Argument für die implizite Annahme des ADH-Modells vor: Die Modellanpassungst Statistik spricht für das validierte Modell und gegen ein alternatives Modell, in dem keine ‚Alles-oder-nichts‘-Annahme des Wortpaar-Abrufs enthalten ist. Die vorliegenden Daten sind also mit der ‚Alles-oder-Nichts‘-Annahme vereinbar und führen nicht nur im relativen Vergleich sondern auch absolut gesehen zu einer guten Modellanpassung. Die getroffene Annahme kann daher sowohl durch theoretische, als auch empirische Argumente gerechtfertigt werden.

Die vorliegenden Modellannahmen sollten systematisch gegenüber alternativen Ansätzen evaluiert werden, indem konkurrierende Modellvarianten entwickelt werden. Eine empirische Evaluation dieser Modellversionen kann Aufschluss darüber geben, welches Modell zu präferieren ist. Die Modellevaluation sollte dabei sowohl die Maße der Modellanpassung berücksichtigen als auch die Reaktion der Parameter auf verschiedene experimentelle Manipulationen. Nur über ein solches Vorgehen können die impliziten Annahmen sowie die Struktur des Modells überzeugend gerechtfertigt werden. Dies wurde in der vorliegenden Arbeit nicht vorgenommen, da weitere Experimente durchgeführt werden müssten, die

spezifisch auf die unterschiedlichen Modellannahmen zugeschnitten sind und diese kritisch überprüfen (vgl. Bayen et al., 1996).

Zur Verteidigung des vorliegenden Modells können daher hauptsächlich zwei Argumente angeführt werden: Zum einen wurden die impliziten theoretischen Annahmen soweit möglich aus anderen evaluierten, etablierten und erfolgreich angewandten Modellen übernommen. Zum anderen sollten sich grobe Fehlspezifikationen der Modellannahmen in einer schlechten Modellanpassung bemerkbar machen. Das ADH-Modell konnte in den Validierungsexperimenten anhand von insgesamt fünf Datensätzen empirisch getestet werden, lediglich für einen Datensatz war die Datenanpassung auf einem α -Niveau von .01 nicht gegeben. Alternative Modelle wiesen - soweit diese getestet wurden - eine schlechtere Modellanpassung auf. Darüber hinaus wird das Modell in dieser Arbeit pragmatisch als Messmodell angesehen und genutzt, es wird daher lediglich angenommen, dass die Modellannahmen eine plausible Approximation der zugrundeliegenden Gedächtnisprozesse widerspiegeln.

Die gerade angesprochene Tatsache, dass die Modellanpassung nicht für alle Datensätze perfekt war, wurde in dieser Arbeit als unkritisch eingestuft. In den fünf Validierungsdatensätzen war für einen Datensatz die Datenanpassung auf einem α -Niveau von .01 nicht gegeben. Lediglich auf einem α -Niveau von .001 wurde für *alle* Modellanwendungen eine Datenpassung erreicht. Dieser Befund kann mit der hohen Anzahl zugrundeliegender Datenpunkte erklärt werden. Diese führten zu einer sehr hohen Teststärke ($1 - \beta > .99$) und bereits kleinste und irrelevante Abweichungen rufen daher eine Verletzung der Modellanpassung hervor. Um diese Argumentation zu stützen, wurden zusätzlich *BIC*-Werte berichtet. Diese berücksichtigen die zugrundeliegende Datenbasis und konnten für alle Experimente eine zufriedenstellende Modellanpassung bestätigen. Inwieweit die zum Teil suboptimale Modellanpassung tatsächlich *nicht* auf problematische Modellannahmen zurückzuführen ist, kann wie erwähnt lediglich über weitere empirische Anwendungen und Evaluationen gegenüber anderen Modellvarianten abschließend geklärt werden.

Bei der Entwicklung des ADH-Modells wurde aufgezeigt, wie mit Identifizierbarkeitsproblemen umgegangen werden kann, wenn weniger unabhängige Datenpunkte als notwendige Parameter vorliegen. Bisher wurde das assoziative Defizit über den Vergleich eines Itemerkennungstests und eines assoziativen Rekognitionstests erfasst. Daraus resultieren allerdings nur zwei unabhängige Datenpunkte. Aus diesem Grund wurde ein neues Paradigma, das kombinierte Item- und assoziative Rekognitionsparadigma, entwickelt (vgl. Buchler et al., 2008). In diesem Paradigma werden alte intakte, sowie neu kombinierte Wortpaare aus dem assoziativen Test kombiniert mit Target-Distraktor-Kombinationen, sowie Distraktor-Target-Paaren aus dem Itemerkennungstest. Um eine vollkommene Randomisierung der Wortpaare zu ermöglichen, wird auch die Kombination von zwei Distraktoren aufgenommen.

Die Probanden müssen in der Testphase jeweils beurteilen, ob es sich bei den Items um alte oder neue Wörter handelt und inwieweit diese in der vorliegenden Kombination gelernt wurden. Damit stehen 20 unabhängige Datenpunkte für eine Modellierung zur Verfügung. Neben der Gewinnung von Freiheitsgraden weist dieses Paradigma den zusätzlichen Vorteil auf, dass Reihenfolgeeffekte in der Testdurchführung nicht mehr durch Ausbalancieren kontrolliert werden müssen. Nachteile des Paradigmas konnten nicht ausgemacht werden. Durch die Entwicklung des kombinierten Item- und assoziativen Rekognitionstests standen genug Freiheitsgrade zur Verfügung, um die Identifizierbarkeit prinzipiell zu ermöglichen: Für alle vorliegenden Datensätze konnte die lokale Identifizierbarkeit nachgewiesen werden.

Neben der nachgewiesenen Identifizierbarkeit und einer guten Modellanpassung konnten die Parameter des Modells in experimentellen Untersuchungen validiert werden:

In Experiment 1b wurden die Itemrekognitionsparameter (w_1 , w_2) über unterschiedliche Präsentationszeiten in der Lernphase validiert. Es wurde manipuliert, ob zunächst das rechte Wort für drei Sekunden alleine dargeboten und das linke Wort für nur eine Sekunde eingeblendet wurde (bzw. genau umgekehrt), oder ob beide Wörter gleichzeitig drei Sekunden lang enkodiert wurden. Diese Manipulation zeigte die erwarteten signifikanten Effekte auf die Itemrekognitionsparameter, und damit deren inhaltliche Bedeutung.

Im zweiten Experiment wurde über ein Vier-Gruppen-Enkodierung-Abruf-Design eine Enkodierungsmanipulation mit einer Abrufmanipulation gekreuzt, um die beiden assoziativen Parameter zu validieren. Der assoziative Enkodierungsparameter c sollte in einer flachen Enkodierungsbedingung (Vokalvergleich) signifikant niedriger ausfallen als in einer tiefen Verarbeitungsbedingung, in der die relationale Verarbeitung der beiden Wörter gefördert wird (Enkodierung des Wortpaares in einem sinnvollen Satz). Die Hälfte der gelernten Wörter wurde nach drei Minuten getestet, die andere Hälfte nach 24 Stunden. Diese Manipulation sollte sich im Parameter des assoziativen Abrufs niederschlagen. Wie erwartet wurde der Enkodierungsparameter nur von der Manipulation der Verarbeitungstiefe signifikant beeinflusst, der Abrufparameter spiegelte lediglich den Effekt des Testzeitpunktes wider. Damit konnte die erfolgreiche Dissoziation der zentralen Prozesse gezeigt werden.

Die Rateparameter wurden in Experiment 3 gegenüber den Gedächtnisparametern abgegrenzt. Eine Veränderung des Anteils alter Wörter im Test (70% vs. 30%) hatte keinen Effekt auf die Gedächtnisparameter, sondern beeinflusste lediglich die Rateprozesse. Die Positionsunterscheidung der Rateparameter für das linke bzw. rechte Wort wurde über unterschiedliche Auszahlungsmatrizen manipuliert. Die Validierung war hier ebenfalls erfolgreich und analog zum ersten Experiment kann nicht nur die Positionsunterscheidung für die Itemparameter sondern, auch für die Rateparameter als gerechtfertigt angesehen werden.

Zusammenfassend gelang es, die kognitionspsychologische Bedeutung der zentralen Parameter sicherzustellen. Die inhaltliche Validierung wurde dabei nicht nur anhand des

schwachen Kriteriums, sondern auch anhand des strengen Kriteriums (Erdfelder, 2000) als erfolgreich bewertet: Es lagen die jeweils konvergenten Effekte auf die Zielparame-ter vor, zugleich wurde sichergestellt, dass trotz einer ausreichenden Teststärke Effekte auf andere Parameter ausblieben. In Experiment 1a konnte darüber hinaus für alle Parameter die Unabhängigkeit von einem bestimmten Testparadigma nachgewiesen werden. Sowohl für einen sequentiellen Ja-nein-Test, als auch für einen simultanen 5AFC-Test zeigten sich keine signifikanten Unterschiede in den Parameterschätzungen. Eine detaillierte Diskussion der Validierungsexperimente kann Kapitel 9 entnommen werden.

Ausgehend von dem validierten Modell wurde eine Modellhierarchie aufgestellt, die unter bestimmten Voraussetzungen die Vereinfachung des Modells erlaubt. Durch Parameterrestriktionen kann das Ausgangsmodell mit 11 Parametern auf spezifische Submodelle restringiert werden, die sparsamste Version sieht sieben Parameter vor. Die Restriktionen beziehen sich zum einen auf die Itemrekognitionsparameter, zum anderen auf die Rateparameter. Eine Vereinfachung der Itemrekognitionsparameter sollte möglich sein, wenn beide Items aus derselben Itemklasse stammen, z.B. wenn es sich bei beiden Items um Wörter mit vergleichbaren linguistischen Merkmalen handelt und beide Wörter gleich lange verarbeitet werden können. Auch die Reduktion der Rateparameter ist von den verwendeten Materialien abhängig, da sich unterschiedliche Erinnerungsleistungen unter Umständen auch in den Rateparametern niederschlagen können. Des Weiteren ist das Verhältnis der unterschiedlichen Itemkombinationen im Test ausschlaggebend für eine Vereinfachung der Rateprozesse. Für die Validierung wurde das ausdifferenzierte 11-Parameter-Modell zugrunde gelegt, da eine spätere Restriktion der Parameter unproblematisch ist. Die Ausdifferenzierung eines validierten Modells erfordert allerdings eine erneute Validierung.

Das 11-Parameter-ADH-Modell wurde also im Hinblick auf alle wichtigen Kriterien der Modellevaluation einer kritischen Prüfung unterzogen, der es standhalten konnte.

11.3 Einsatzmöglichkeiten des ADH-Modells

Ein zentrales Kriterium bei der Entwicklung des ADH-Modells war es, einen möglichst breiten Einsatz sicherzustellen und daher keine besonderen Anforderungen an das Stimulusmaterial und die Testsituation zu stellen. Die Einschränkung auf semantisch assoziiertes Wortmaterial (Enkodierung-Abruf-Modell) ist im Kontext der Assoziationsdefizit-Hypothese problematisch, da das assoziative Defizit unter diesen Bedingungen geringer ausfällt (Naveh-Benjamin, 2000). Die Notwendigkeit, unterschiedliche Testformate zu verwenden, wie zum Beispiel der Vergleich freier und geförderter Reproduktionstests (Riefer & Rouder, 1992) bzw. der Vergleich von geförderter Reproduktion und Rekognition (Chechile & Meyer, 1976) hat den Nachteil, dass die unterschiedliche Beteiligung selbstinitiiertter Prozesse (Craik & Byrd, 1982) den Altersvergleich erschwert. Darüber hinaus werden dadurch die Stimulusmaterialien auf verbale Stimuli eingeschränkt, da visuelle Stimuli wie Bilder

oder Gesichter genau wie realistische Stimuli (z.B. ausgeführte Handlungen) nur für die Rekognition geeignet sind.

Für das vorliegende Modell sollte es möglich sein, assoziierte und nicht assoziierte Wörter zu verwenden, visuelle Materialien sollten genauso analysiert werden können wie realistische Assoziationen von Personen und ausgeführten Handlungen.

Aufgrund dieser Zielsetzung kann das ADH-Modell nicht nur im Kontext der Assoziationsdefizit-Hypothese verwendet werden und ist auch nicht auf entwicklungspsychologische Fragestellungen begrenzt. Vielmehr können auch in der allgemeinen Psychologie neue Forschungsfragen damit untersucht werden:

- (1) Das kombinierte Item- und assoziative Rekognitionsparadigma ist dem Quellengedächtnisparadigma relativ ähnlich. Es ist daher möglich, durch eine kleine Veränderung in der Testphase des Quellengedächtnisparadigmas dieses Modell auch zur Auswertung von Quellengedächtnisdaten einzusetzen.⁴³ Da das ADH-Modell eine etwas andere Analyse ermöglicht als die multinomialen Quellengedächtnismodelle (Batchelder & Riefer, 1990; Bayen & Murnane, 1996; Meiser & Bröder, 2002) kann es - abhängig von der zugrundeliegenden Forschungsfrage - aufschlussreich sein, das vorliegende Modell auf Quellengedächtnisdaten anzuwenden. Unter anderem ermöglicht das ADH-Modell die separate Erfassung von Quellenrekognitionsleistung und Assoziation von Quelle und zentraler Information. In den multinomialen Quellengedächtnismodellen wird mit dem Quellengedächtnisparameter d sowohl die Erinnerung an die Quelle als auch die Erinnerung an die Verbindung von Quelle und Iteminformation erfasst. Diese beiden Prozesse sind konfundiert. In den klassischen Quellengedächtnisstudien ist dies unproblematisch, da meist nur zwei Quellen verwendet werden. Es kann daher von einer perfekten Quellenrekognition ausgegangen werden. Sobald allerdings mehrere Quellen vorliegen z.B. 10 unterschiedliche Sprecher, ist es durchaus von Interesse, die Aspekte der Quellenerinnerung sowie die Assoziation von Quelle und Aussage zu trennen. Mit dem Einsatz des ADH-Modells ist diese Möglichkeit gegeben, und es könnten z.B. Interferenzeffekte bei unterschiedlichen Zuordnungsverhältnissen untersucht werden (siehe Kapitel 3.2.2).
- (2) Des Weiteren kann das ADH-Modell auch eingesetzt werden, um die Mechanismen, die assoziativen Lernstrategien zugrunde liegen, genauer zu erforschen. In Kapitel 3.2.3 wurden unterschiedliche assoziative Strategien diskutiert, z.B. das Generieren von Verbindungswörtern oder die Definition neuer Komposita. Kombiniert man den experimentellen Einsatz dieser Strategien mit einer Datenanalyse durch das ADH-

⁴³Der Quellengedächtnistest müsste dazu um die Frage nach der Rekognition der Quelle erweitert werden. In einem sequentiellen Test müsste also zunächst erfragt werden, ob es sich um eine gelernte Aussage handelt, dann müsste die Quelle als alte oder neue Quelle identifiziert werden. Wurden beide Stimuli als gelernte Stimuli eingestuft, würde abschließend die Frage gestellt werden, ob die spezifische Information auch von dieser Quelle dargeboten wurde.

Modell, so kann dies Aufschluss darüber geben, ob die Strategien ausschließlich die assoziative Enkodierung fördern oder auch den assoziativen Abruf. Damit ist man nicht mehr auf die bisher genutzten Selbstberichtsdaten angewiesen, und es werden alle Nachteile aufgehoben, die der Selbstbericht mit sich bringt (Dunlosky et al., 2005). Auch die Auswirkungen dieser assoziativen Strategien auf die Itemerinnerung können untersucht werden.

- (3) Da mit der Einführung des kombinierten Item- und assoziativen Rekognitionstests zwei bisher meist unabhängig voneinander untersuchte Paradigmen kombiniert wurden, ergeben sich durch diese Zusammenführung neue Forschungsfragen. So ist es unter anderem möglich, die Auswirkungen von sich überlappenden bzw. überschneidenden Wortpaaren in der Lernliste (AB, AC, AD) sowohl im Hinblick auf die assoziative Gedächtnisleistung als auch im Hinblick auf die Itemgedächtnisleistung in *einem* Paradigma und *einer* Analyse genauer zu untersuchen (Buchler et al., 2008).

Wie diese Beispiele zeigen, eröffnet das neue Paradigma sowie die neue Auswertungsmethode in unterschiedlichen Gebieten der Allgemeinen Psychologie interessante Forschungsvorhaben.

11.4 Entwicklungspsychologische Implikationen

Neben der Allgemeinen Psychologie ist ein wichtiges Anwendungsfeld des Modells in der Entwicklungspsychologie zu sehen. Bevor konkret erläutert wird welche Erkenntnisse das Modell in Bezug auf die Assoziationsdefizit-Hypothese ermöglicht, soll in diesem Abschnitt genauer begründet werden, warum gerade in der Entwicklungspsychologie der Einsatz von modellbasierten Verfahren besonders notwendig ist.

Im theoretischen Teil der Arbeit wurde zunächst die Kontextdefizit-Hypothese (Burke & Light, 1981) erläutert, da diese als Vorläufer der Assoziationsdefizit-Hypothese (Naveh-Benjamin, 2000) angesehen werden kann. Bei diesem Rückblick wurden vor allem die verwendeten Auswertungsmethoden kritisch evaluiert, und es konnte aufgezeigt werden, dass einige Experimente, die zunächst als empirische Belege für die Kontextdefizit-Hypothese aufgeführt wurden, auch alternative Interpretationen zulassen und nicht notwendigerweise die Kontextdefizit-Hypothese unterstützen. Manche Studien deuten eher auf ein Assoziationsdefizit als auf ein Kontextdefizit hin (siehe Kapitel 3.1). Es wurde verdeutlicht, dass eine fundierte Methodendiskussion und der Einsatz adäquater Analyseverfahren gerade in der Entwicklungspsychologie von ganz besonderer Bedeutung sind. Dies ist unabhängig von den berichteten Studien allgemein auf zwei Faktoren zurückzuführen: Erstens ist die Entwicklungspsychologie meist an Altersunterschieden sehr spezifischer Prozesse interessiert. Zum zweiten handelt es sich bei altersgruppenvergleichenden Experimenten notwendigerweise immer um Quasi-Experimente, eine perfekte Randomisierung ist daher

nicht möglich.⁴⁴ Die besonderen Herausforderungen, die sich durch die Kombination dieser beiden Aspekte ergeben, werden nun anhand einiger - über die Kontextdefizit-Hypothese hinausgehender - Beispiele verdeutlicht:

- (1) In gerontologischen Studien zum Rekognitionsgedächtnis stehen Unterschiede in den Gedächtnisprozessen im Fokus des Interesses, Alterseffekte in Antworttendenzen werden als Störeffekte angesehen. Es ist daher von Bedeutung, die Gedächtnisleistung unabhängig von den Antworttendenzen zu erfassen, denn nur auf diesen spezifischen Prozess sind die Altershypothesen ausgerichtet.
- (2) Bei Entwicklungsstudien zum prospektiven Gedächtnis wiederum werden retrospektive Gedächtnisunterschiede häufig als Störeffekte aufgefasst. Das Forschungsinteresse richtet sich auf Altersunterschiede in den spezifischen prospektiven Gedächtniskomponenten. Auch hier ist es also zentral, unterschiedliche Prozesse zu trennen und in diesem Fall eine Konfundierung von Alter und retrospektivem Gedächtnis zu vermeiden (Smith & Bayen, 2004).
- (3) Ein ähnliches Interesse liegt in der Quellengedächtnisforschung vor (siehe Kapitel 3.1). In diesen Experimenten soll die Itemrekognitionsleistung kontrolliert werden, da das Quellengedächtnis im Vordergrund steht. Bei einer mangelnden Kontrolle unterschiedlicher Itemgedächtnisleistungen könnten diese Effekte einen Unterschied im Quellengedächtnis vortäuschen, der in der Realität nicht vorhanden ist (Batchelder & Riefer, 1990; Bayen & Murnane, 1996).
- (4) Nicht nur in Gedächtnisstudien, sondern auch im Bereich des logischen Schlussfolgerns und Denkens ist es in entwicklungspsychologischen Studien relevant, unterschiedliche Teilprozesse zu dekomponieren. In der Vier-Karten-Aufgabe von Wason (1968) wäre es z.B. nicht nur aufschlussreich, Altersunterschiede per se aufzudecken, sondern gezielt zu analysieren, inwieweit ältere Menschen, die an die Aufgabe analytisch herangehen, weniger komplexe Vorgehensweisen nutzen als jüngere Menschen, die analytisch verfahren. Auch ist vorstellbar, dass bei einer traditionellen Analyse keine Altersunterschiede zu Tage treten, in einer spezifischen Analyse (Klauer et al., 2007) aber altersabhängig unterschiedliche Vorgehensweisen entdeckt werden. Ein unspezifisches Analyseverfahren könnte also zu dem falschen Schluss führen, dass keine Altersunterschiede vorliegen.
- (5) Unspezifische Analysen können Altersunterschiede nicht nur verschleiern, sondern auch suggerieren. In Altersstudien zum Rückschaufehler indizieren manche Auswertungsverfahren eine stärkere Antwortverzerrung älterer Menschen, obwohl diese nicht vorliegt (Bayen et al., 2006). Die mangelnde Berücksichtigung des unterschiedlichen

⁴⁴Natürlich gilt dies nicht nur für die Entwicklungspsychologie auch in der klinischen Psychologie liegt eine ähnliche Problematik vor.

globalen Erinnerungsniveaus der beiden Altersgruppen kann solche Scheinunterschiede hervorrufen. Selbst Auswertungsverfahren, die die Kontrolle der allgemeinen Erinnerungsleistung erlauben, sind zum Teil nicht in der Lage, Rückschlüsse auf die spezifischen Mechanismen des Rückschleifers zu ermöglichen. So findet sich unter Berücksichtigung der Gedächtnisleistung ein stärkerer Rückschleifer bei älteren Menschen und kleinen Kindern, dieser beruht aber auf unterschiedlichen Prozessen, wie eine multinomiale Auswertung aufzeigen konnte: Während bei älteren Menschen vor allem Rekonstruktionsprozesse relevant sind, ist bei Kindern auch eine signifikante Quellenverwechslung zu beobachten (Pohl & Auer, 2008). Diese spezifischen Erkenntnisse wurden erst durch eine multinomiale Analyse ermöglicht.

Diese Beispiele sowie die im theoretischen Teil der Arbeit angeführten Studien zeigen, dass bei vielen entwicklungspsychologischen Forschungsfragen sehr spezifische Prozesse von Interesse sind, die dekomponiert werden müssen. Ein Weg, diese Dekomponierung zu erreichen, besteht in designbasierten Vorgehensweisen. Unabhängig von den allgemeinen Problemen dieses Ansatzes bei Dekomponierungsproblemen (Brainerd, 1985; Erdfelder, 2000) kommt in der Entwicklungspsychologie verkomplizierend hinzu, dass es sich bei altersgruppenvergleichenden Experimenten immer um Quasi-Experimente handelt. Eine perfekte Randomisierung ist also nicht möglich, und der Einfluss von alterskonfundierten Faktoren kann nicht ausgeschlossen werden. Dies sollte verdeutlichen, dass eine fundierte und kritische Methodendiskussion sowie die Anwendung von elaborierten, modellbasierten Auswertungsverfahren gerade in der Entwicklungspsychologie unabdingbar sind. Diese Methoden erlauben es, konfundierende Faktoren zu modellieren und diese in der Datenauswertung (d.h. indirekt) zu erfassen. Der Einsatz mathematischer Modelle sollte daher in dieser Disziplin eine viel stärkere Verwendung finden, als es bisher der Fall ist (für einen Überblick siehe Spaniol & Bayen, 2004). Dies wurde nicht nur anhand theoretischer Argumente verdeutlicht, auch das praktische Vorgehen der Entwicklung modellbasierter Auswertungsverfahren wurde anhand des ADH-Modells aufgezeigt. Das ADH-Modell ermöglicht es, die mit dem kombinierten Item- und assoziativen Rekognitionsparadigma erfasste Gedächtnisleistung auf die Prozesse der Itemrekognition, assoziative Enkodierungs- und Abrufprozesse, Distraktorentdeckungsmechanismen sowie Ratetendenzen zurückzuführen. Damit kann klar zwischen einem Kontextdefizit und einem Assoziationsdefizit unterschieden werden. Ersteres sollte bei älteren Menschen zu einer schlechteren Rekognitionsleistung der Kontextkomponenten führen (Parameter w_1 bzw. w_2), letzteres sollte Altersunterschiede im assoziativen Enkodierungs- oder Abrufparameter hervorrufen.

11.5 Spezifizierung der Assoziationsdefizit-Hypothese

Im Theorieteil der Arbeit wurde das Klassifikationssystem von Salthouse (1991) genutzt, um die Assoziationsdefizit-Hypothese systematisch in das Gebiet der kognitiven Alters-

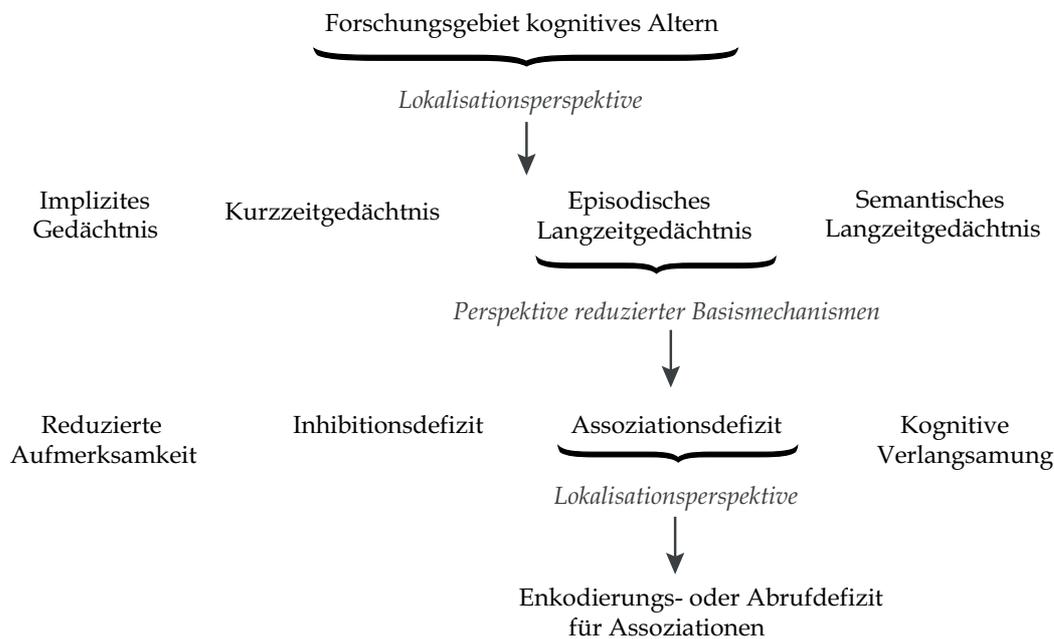


Abbildung 11.1: In dieser Abbildung wird veranschaulicht, wie die Forschungsfrage dieser Arbeit in das Feld der kognitiven Altersforschung eingeordnet werden kann. Zunächst wird das globale Forschungsfeld des kognitiven Alterns durch die Lokalisationsperspektive auf das Gebiet des episodischen Langzeitgedächtnisses eingeschränkt. Innerhalb des episodischen Langzeitgedächtnisses wird dann die Perspektive der reduzierten Verarbeitungsressourcen herangezogen und das Assoziationsdefizit als Erklärungsansatz fokussiert. Dieser Basismechanismus kann aus Sicht der Lokalisationsperspektive weiter unterteilt werden in ein assoziatives Defizit in der Enkodierungsphase und ein Defizit während des Abrufs.

forschung einzuordnen (siehe Abbildung 11.1). Des Weiteren wurden die theoretischen Perspektiven von Salthouse in einem zweiten Schritt auch auf die Assoziationsdefizit-Hypothese und deren empirische Befunde angewandt.

Hinsichtlich der Perspektive *qualitativ unterschiedlicher Verarbeitungsmechanismen* stellte sich heraus, dass strategische Prozesse substantiell zu dem Alterseffekt des assoziativen Defizits beitragen (Naveh-Benjamin, Brav & Levy, 2007; Shing et al., 2008). Selbst durch intensive und langfristig angelegte Strategietrainings können die Alterseffekte aber nicht *allein* auf diese qualitativ unterschiedlichen Prozesse zurückgeführt werden (Shing et al., 2008). Um die damit naheliegende Einordnung des *assoziativen Gedächtnisdefizits als Basismechanismus* zu rechtfertigen, wurden Experimente berichtet, die zeigen, dass alternative Basismechanismen den Alterseffekt des assoziativen Defizits nicht erklären können (Graef & Naveh-Benjamin, 2008; Naveh-Benjamin, 2008a; Naveh-Benjamin, Kilb & Hara, 2007). Daher kommt Naveh-Benjamin (2008a) zu der Schlussfolgerung, dem assoziativen Defizit die Rolle eines spezifischen Basismechanismus zuzuordnen. An dieser Stelle muss allerdings kritisch eingeräumt werden, dass zu dieser Position bisher nur wenige und hauptsächlich unveröffentlichte Studien vorliegen. Nach dem momentanen Forschungsstand ist

die Einordnung als Basismechanismus zwar gerechtfertigt, diese Schlussfolgerung muss allerdings durch weitere methodisch fundierte Experimente untermauert werden. Im Hinblick auf die *analytische Lokalisationsperspektive* des assoziativen Defizits liegen bisher keine überzeugenden Forschungsergebnisse vor. Naveh-Benjamin (2000) postuliert ein Defizit in Enkodierungs- sowie Abrufprozessen, wobei er dies theoretisch nicht genauer begründet und empirisch nicht belegt. Studien zu ähnlichen Paradigmen (Riefer & Batchelder, 1991; Bayen, 1990) legen ein Enkodierungs- sowie Abrufdefizit nahe. Die einzige Studie, die in einem vergleichbaren Rekognitionsparadigma durchgeführt wurde (Mitchell et al., 2000), konnte ein Enkodierungsdefizit nachweisen. Diese Forschungslücke wurde durch die vorliegende Arbeit geschlossen - zum einen aufgrund der Entwicklung des ADH-Modells, welches die Trennung von assoziativen Enkodierungs- und Abrufprozessen erlaubt, zum anderen aufgrund der durchgeführten Altersstudie.

Die vorliegenden Altersstudie hatte zum Ziel, empirische Hinweise auf die Lokalisation des assoziativen Defizit zu erlangen. Keinesfalls war daran die Erwartung geknüpft, dass mit diesem einen Experiment alle Aspekte des assoziativen Defizits beantwortet werden können. Diese Modellanwendung sollte vielmehr eine erste spezifische Analyse des Defizits ermöglichen und damit den Grundstein für weitere Altersstudien legen.

Die Ergebnisse zeigten ein klares Enkodierungsdefizit auf. Der Parameter der assoziativen Enkodierung c unterschied sich signifikant zwischen den Altersgruppen. Über ein Abrufdefizit kann aufgrund einer mangelnden Teststärke keine abschließende Aussage getroffen werden, die Daten deuten aber deskriptiv auf ein zusätzliches Abrufdefizit hin.

Die mangelnde Teststärke für den Abrufparameter ist auf zwei Faktoren zurückzuführen, die in zukünftigen Studien beachtet werden sollten: Die relativ niedrigen Werte des Enkodierungsparameters der älteren Stichprobe führten dazu, dass wenige Datenpunkte in die Schätzung des nachgeordneten Abrufparameter eingingen. In zukünftigen Experimenten sollte man daher versuchen, die Enkodierungsleistung (c) zu erhöhen. Dies kann durch eine längere Präsentationsdauer erreicht werden. Zwar ergibt sich daraus eventuell ein Deckeneffekt bei der jüngeren Stichprobe, sobald ein Experiment aber ausschließlich auf den Alterseffekt im Abrufparameter abzielt, sollte dies unproblematisch sein. Alternativ könnten mit den vorliegenden Lernbedingungen nicht nur zwei Listen, sondern mehrere Listen getestet werden, um die Anzahl der Datenpunkte zu erhöhen.

Dies knüpft an das zweite Problem der vorliegenden Altersstudie an: Die beiden Lern- bzw. Testlisten konnten nicht gemeinsam ausgewertet werden, da dies eine schlechte Datenanpassung des Modells zur Folge hatte. Die Anpassung an die Daten war allerdings gegeben, wenn beide Listen getrennt ausgewertet wurden. Innerhalb jeder Liste lag ein zufriedenstellender G^2 -Wert vor. Dieses Ergebnis weist darauf hin, dass zwischen den Listen Parameterveränderungen aufgetreten sind, die eine Aggregation der Daten nicht erlauben. Eventuell wendeten die Probanden durch die Lernerfahrung der ersten Liste im zweiten

Durchgang andere Strategien an. Natürlich könnten auch Interferenzeffekte - gerade bei den älteren Probanden - zu unterschiedlichen Parameterschätzungen in den beiden Listen geführt haben. Da sich die Gedächtnisleistung aber eher verbesserte und Unterschiede in den Rateprozessen zu verzeichnen waren, legt dies eine Veränderung der strategischen Prozesse nahe. Die getrennte Auswertung der beiden Listen hatte Einbußen in der Teststärke zur Folge. Um dies in weiteren Studien zu vermeiden, sollten entweder mehrere Listen getestet werden, damit mindestens die zweite und dritte Liste gemeinsam ausgewertet werden können, oder eine längere Übungsphase sollte insbesondere den älteren Menschen ausreichend Gelegenheit geben, Lernerfahrungen zu sammeln.

Obwohl aufgrund dieses Experiments keine abschließenden Aussagen bezüglich des Abrufdefizits getroffen werden können, wurde ein deutlicher Erkenntnisgewinn durch diese Altersstudie erreicht. Zum einen zeigte sich, dass das ADH-Modell auch bei älteren Stichproben erfolgreich eingesetzt werden kann. Zum anderen unterstützt das Enkodierungsdefizit die Schlussfolgerungen aus der Studie von Mitchell et al. (2000), die mit einem designbasierten Ansatz ebenfalls signifikante Alterseffekte in der Enkodierungsleistung fanden. Dem assoziativen Alterseffekt liegt also mindestens ein Enkodierungsdefizit der Senioren zugrunde.

Im Hinblick auf die Generalisierung des vorliegenden Altersexperiments muss darauf hingewiesen werden, dass es sich bei der älteren Stichprobe um Seniorenstudenten der Universität Mannheim mit einem Durchschnittsalter von 66.53 Jahren handelte, d.h. es lag eine sehr selektive, geistig aktive und relativ „junge“ ältere Stichprobe vor. Durch die universitäre Weiterbildung sind diese Probanden auch stärker mit Lernsituationen konfrontiert, als es bei einer bevölkerungsrepräsentativen Stichprobe zu erwarten wäre. Daher sollten die Ergebnisse - insbesondere im Hinblick auf das nicht signifikante Abrufdefizit - mit Vorsicht interpretiert werden. Es ist durchaus vorstellbar, dass eine bevölkerungsrepräsentative Stichprobe ein starkes Abrufdefizit aufweist. Die Ergebnisse des Enkodierungsdefizits sollten auch bei einer repräsentativen Stichprobe mindestens erhalten bleiben, wahrscheinlich würde dieser Alterseffekt eher größer ausfallen. Das vorliegende Experiment unterschätzt also aufgrund der geistig aktiven und „jungen“ Stichprobe eher die wahren Effekte. Positiv ist dies im Hinblick auf die Sensitivität des Paradigmas zu sehen, denn trotz dieser „erschweren“ Bedingungen konnte ein klares Enkodierungsdefizit festgestellt werden.

11.6 Ausblick

In weiteren Altersstudien sollte zunächst das Abrufdefizit genauer untersucht werden, um damit eine abschließende Antwort auf die Lokalisationsfrage zu geben. Hinweise, wie eine ausreichende Teststärke sichergestellt werden kann, wurden bereits angesprochen. Hinsichtlich der Generalisierung der weiteren Ergebnisse wäre es aufschlussreich, den kombi-

nierten Item- und assoziativen Rekognitionstest auch mit assoziierten Wortpaaren durchzuführen. Diese Resultate können in Verbindung mit den Ergebnissen aus der freien Reproduktion (Bayen, 1990; Riefer & Batchelder, 1991) zur Generalisierung eines Enkodierungs- und/oder Abrufdefizits im Alter beitragen.

Des Weiteren wäre es interessant, die Auswirkungen von Strategien mit einer modellbasierten Auswertung genauer zu spezifizieren. In designbasierten Studien wurde der Altersunterschied durch strategische Hinweise und ein strategisches Training verringert. Aufschlussreich wäre es nun herauszufinden, ob der Strategieeinsatz den Altersunterschied in Enkodierungs- und Abrufprozessen gleichermaßen verringert, oder ob sich selektive Effekte auf bestimmte Parameter zeigen. Eine solche Studie hätte auch direkt praktische Implikationen für den angewandten Trainings- und Interventionsbereich, da effiziente Strategien, die die defizitären Gedächtnisprozesse unterstützen, im Sinne von Kompensationsstrategien spezifisch eingesetzt werden könnten.

Ein weiterer eher für die Grundlagenforschung interessanter Aspekt sind die Auswirkungen des Zuordnungsverhältnisses auf assoziative Prozesse. In Kapitel 3.2.2 wurde erläutert, dass dem Aspekt des Zuordnungsverhältnisses bisher relativ wenig Aufmerksamkeit geschenkt wurde. Gerade im Hinblick auf Studien im entwicklungspsychologischen Kontext ist dieser Faktor allerdings von Bedeutung, da darüber die Interferenz der Aufgabe moderiert wird. Ältere Menschen sowie Kinder zeigen eine erhöhte Interferenzanfälligkeit, es wäre daher interessant, diesen Faktor auch im Kontext des assoziativen Defizits genauer zu untersuchen und die Interferenzauswirkungen auf *assoziative Gedächtnisprozesse* zu spezifizieren.

Ein letzter Aspekt, der genannt werden soll, ist der Einsatz des Modells an einer Stichprobe, die die gesamte Lebensspanne umfasst. Dabei können sowohl Parallelen als auch Unterschiede in der frühen Entwicklung und der späteren Degeneration von assoziativen Gedächtnisprozessen deutlich werden. Die bisher einzige Lebensspannenstudie von Shing et al. (2008) zeigte, dass bei Kindern im Alter von 10-12 sowie 13-15 Jahren das assoziative Defizit vor allem auf strategischen Prozessen beruht, da dies durch ein intensives Training aufgehoben werden konnte. Die Autoren vermuten allerdings, dass bei jüngeren Kindern - ähnlich wie bei älteren Menschen - die Gedächtniskomponente eine größere Rolle spielt. Inwieweit das assoziative Defizit in dieser Altersgruppe nicht nur global, sondern auch im Hinblick auf spezifische Prozesse mit dem Defizit der Älteren vergleichbar ist, könnte über eine Analyse mit dem ADH-Modell geklärt werden.

In dieser abschließenden Diskussion wurde aufgezeigt, wie wichtig ein adäquater Methodeinsatz gerade in der Entwicklungspsychologie ist und welche Vorteile modellbasierte Auswertungsverfahren mit sich bringen. Das ADH-Modell wurde kritisch diskutiert und Verfahren zur Untersuchung alternativer Modellvarianten für die zukünftige Forschung wurden aufgezeigt. Da das Modell alle Kriterien der Modellevaluierung erfüllen konnte,

kann es als Messmodell verwendet werden. Vielfältige Einsatzbereiche wurden diskutiert, wobei das Modell sicherlich neben den hier vorgestellten Forschungsfragen auch für weitere Projekte gewinnbringend eingesetzt werden kann.

11.7 Fazit

In dieser Arbeit wurde ein mathematisches Modell entwickelt, das es ermöglicht, die lang diskutierte Lokalisationsfrage des altersbedingten assoziativen Defizits zu beantworten. Dieses Modell erlaubt erstmals die getrennte Erfassung von Enkodierungs- und Abrufprozessen in einem Rekognitionsparadigma und ist somit für zahlreiche Forschungsfragen nicht nur in der Entwicklungspsychologie, sondern z.B. auch in der allgemeinen Psychologie gewinnbringend einsetzbar. In vier Experimenten wurden die Modellparameter sowohl im Hinblick auf ihre konvergente als auch unter Berücksichtigung der diskriminanten Validität erfolgreich validiert. Des Weiteren zeigte das Modell eine gute Datenanpassung. Die Anwendung des Modells im gerontologischen Bereich trug zunächst zur Generalisierung des assoziativen Defizits bei, da es auch in einem neuen Paradigma repliziert werden konnte. Als zentraler und innovativer Befund des Altersexperiments konnte aufgezeigt werden, dass ältere Menschen ein Enkodierungsdefizit für assoziative Informationen aufweisen.

In dieser Arbeit wurde erfolgreich gezeigt *wie* die Lokalisationsfrage adäquat beantwortet werden kann, darüber hinaus konnte ein Enkodierungsdefizit älterer Menschen als Ursache des assoziativen Defizits identifiziert werden.

Teil V

Literatur

Literaturverzeichnis

- Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, 6(4), 451-474.
- Anderson, J. R. & Bower, G. H. (1973). *Human associative memory*. Washington: Winston and Sons.
- Balgur, R. (1968). *List of basic words for school*. Tel Aviv, Israel: Otsar Hamoreh.
- Ball, K., Berch, D. B., Helmers, K. F., Jobe, J. B., Leveck, M. D., Marsiske, M. et al. (2002). Effects of cognitive training interventions with older adults: A randomized controlled trial. *Journal of the American Medical Association*, 288(18), 2271-2281.
- Balota, D. A., Duchek, J. M. & Paullin, R. (1989). Age-related differences in the impact of spacing, lag, and retention interval. *Psychology and Aging*, 4(1), 3-9.
- Baltes, P. B. & Kliegl, R. (1992). Further testing of limits of cognitive plasticity: Negative age differences in a mnemonic skill are robust. *Developmental Psychology*, 28(1), 121-125.
- Baltes, P. B. & Lindenberger, U. (1997). Emergence of a powerful connection between sensory and cognitive functions across the adult life span: A new window to the study of cognitive aging? *Psychology and Aging*, 12(1), 12-21.
- Bamber, D. & Santen, J. P. H. van. (2000). How to assess a model's testability and identifiability. *Journal of Mathematical Psychology*, 44(1), 20-40.
- Bastin, C. & Van der Linden, M. (2005). Memory for temporal context: Effects of ageing, encoding instructions, and retrieval strategies. *Memory*, 13(1), 95-109.
- Bastin, C. & Van der Linden, M. (2006). The effects of aging on the recognition of different types of associations. *Experimental Aging Research*, 32(1), 61-77.
- Batchelder, W. H. & Batchelder, E. (2008). Metacognitive guessing strategies in source monitoring. In J. Dunlosky & R. A. Bjork (Hrsg.), *Handbook of metamemory and memory* (S. 211-244). New York: Psychology Press.
- Batchelder, W. H. & Riefer, D. M. (1980). Separation of storage and retrieval factors in free recall of clusterable pairs. *Psychological Review*, 87(4), 375-397.
- Batchelder, W. H. & Riefer, D. M. (1986). The statistical analysis of a model for storage and retrieval processes in human memory. *British Journal of Mathematical and Statistical Psychology*, 39(2), 129-149.
- Batchelder, W. H. & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review*, 97(4), 548-564.
- Batchelder, W. H. & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, 6(1), 57-86.
- Batchelder, W. H., Riefer, D. M. & Hu, X. (1994). Measuring memory factors in source monitoring: Reply to Kinchla. *Psychological Review*, 101(1), 172-176.
- Bäuml, K.-H. (1996). A Markov model for measuring storage loss and retrieval failure in

- retroactive inhibition. *Acta Psychologica*, 92(3), 231-250.
- Bayen, U. J. (1990). *Zur Lokalisation von Altersdifferenzen im episodischen Gedächtnis Erwachsener: Eine Querschnittsuntersuchung auf der Basis eines mathematischen Modells*. Unveröffentlichte Diplomarbeit, Psychologisches Institut der Albert-Ludwigs-Universität Freiburg i.Br.
- Bayen, U. J., Erdfelder, E., Bearden, J. N. & Lozito, J. P. (2006). The interplay of memory and judgment processes in effects of aging on hindsight bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(5), 1003-1018.
- Bayen, U. J. & Murnane, K. (1996). Aging and the use of perceptual and temporal information in source memory tasks. *Psychology and Aging*, 11(2), 293-303.
- Bayen, U. J., Murnane, K. & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(1), 197-215.
- Bayen, U. J., Nakamura, G. V., Dupuis, S. E. & Yang, C.-L. (2000). The use of schematic knowledge about sources in source monitoring. *Memory & Cognition*, 28(3), 480-500.
- Bayen, U. J., Phelps, M. P. & Spaniol, J. (2000). Age-related differences in the use of contextual information in recognition memory: A global matching approach. *Journals of Gerontology: Series B: Psychological Sciences and Social Sciences*, 55B(3), 131-141.
- Begg, I. (1978). Imagery and organization in memory: Instructional effects. *Memory & Cognition*, 6(2), 174-183.
- Bentin, S., Moscovitch, M. & Nirhod, O. (1998). Levels of processing and selective attention effects on encoding in memory. *Acta Psychologica*, 98(2-3), 311-341.
- Birren, J. E., Riegel, K. F. & Morrison, D. F. (1962). Age differences in response speed as a function of controlled variations of stimulus conditions: Evidence of a general speed factor. *Gerontologia*, 6(1), 1-18.
- Bower, G. H. (1970). Imagery as a relational organizer in associative learning. *Journal of Verbal Learning and Verbal Behavior*, 9(5), 529-533.
- Bowles, N. L. & Poon, L. W. (1982). An analysis of the effect of aging on recognition memory. *Journal of Gerontology*, 37(2), 212-219.
- Brainerd, C. J. (1985). Model-based approaches to storage and retrieval development. In C. J. Brainerd & M. Pressley (Hrsg.), *Basic processes in memory development: Progress in cognitive development research* (S. 143-208). New York: Springer.
- Brehmer, Y., Li, S.-C., Straube, B., Stoll, G., von Oertzen, T., Müller, V. et al. (2008). Comparing memory skill maintenance across the life span: Preservation in adults, increase in children. *Psychology and Aging*, 23(2), 227-238.
- Brockmole, J. R., Parra, M. A., Della Sala, S. & Logie, R. H. (2008). Do binding deficits account for age-related decline in visual working memory? *Psychonomic Bulletin &*

Review, 15(3), 543-547.

- Bröder, A., Herwig, A., Teipel, S. & Fast, K. (2008). Different storage and retrieval deficits in normal aging and mild cognitive impairment: A multinomial modeling analysis. *Psychology and Aging*, 23(2), 353-365.
- Bröder, A. & Meiser, T. (2007). Measuring source memory. *Zeitschrift für Psychologie/ Journal of Psychology*, 215(1), 52-60.
- Brown, G. D., Preece, T. & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, 107(1), 127-181.
- Brown, J., Lewis, V. J. & Monk, A. F. (1977). Memorability, word frequency and negative recognition. *The Quarterly Journal of Experimental Psychology*, 29(3), 461-473.
- Buchler, N. G., Light, L. L. & Reder, L. M. (2008). Memory for items and associations: Distinct representations and processes in associative recognition. *Journal of Memory and Language*, 59(2), 183-199.
- Buchner, A., Erdfelder, E. & Vaterrodt-Plünnecke, B. (1995). Toward unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *Journal of Experimental Psychology: General*, 124(2), 137-160.
- Buckner, R. L., Head, D. & Lustig, C. (2006). Brain changes in aging: A lifespan perspective. In E. Bialystok & F. I. M. Craik (Hrsg.), *Lifespan cognition: Mechanisms of change* (S. 27-42). New York, NY: Oxford University Press.
- Bugelski, B. R. (1962). Presentation time, total time, and mediation in paired-associate learning. *Journal of Experimental Psychology*, 63(4), 409-412.
- Bugelski, B. R., Kidd, E. & Segmen, J. (1968). Image as a mediator in one-trial paired-associate learning. *Journal of Experimental Psychology*, 76(1), 69-73.
- Burke, D. M. & Light, L. L. (1981). Memory and aging: The role of retrieval processes. *Psychological Bulletin*, 90(3), 513-546.
- Cabeza, R. (2006). Prefrontal and mediotemporal lobe contributions to relational memory in young and older adults. In H. D. Zimmer, M. A. & U. Lindenberger (Hrsg.), *Handbook of binding and memory: Perspectives from cognitive neuroscience* (S. 595-626). New York: Oxford University Press.
- Calfee, R. C. & Anderson, R. (1971). Presentation rate effects in paired-associate learning. *Journal of Experimental Psychology*, 88(2), 239-245.
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105.
- Carroll, J. B. & Burke, M. L. (1965). Parameters of paired-associate verbal learning: Length of list, meaningfulness, rate of presentation, and ability. *Journal of Experimental Psychology*, 69(6), 543-553.
- Carstensen, L. L. (1992). Social and emotional patterns in adulthood: Support for socioemotional selectivity theory. *Psychology and Aging*, 7, 331-338.
- Carstensen, L. L. (1995). Evidence for a life-span theory of socioemotional selectivity.

Current Directions in Psychological Science, 4, 151-156.

- Carstensen, L. L., Fung, H. H. & Charles, S. T. (2003). Socioemotional selectivity theory and the regulation of emotion in the second half of life. *Motivation and Emotion*, 27(2), 103-123.
- Carstensen, L. L. & Mikels, J. A. (2005). At the intersection of emotion and cognition: Aging and the positivity effect. *Current Directions in Psychological Science*, 14(3), 117-121.
- Castel, A. D. & Craik, F. I. M. (2003). The effects of aging and divided attention on memory for item and associative information. *Psychology and Aging*, 18(4), 873-885.
- Cerella, J. (1985). Information processing rates in the elderly. *Psychological Bulletin*, 40, 153-194.
- Chalfonte, B. L. & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory & Cognition*, 24(4), 403-416.
- Challis, B. H., Velichovsky, B. M. & Craik, F. I. M. (1996). Levels-of-processing effects on a variety of memory tasks: New findings and theoretical implications. *Consciousness and Cognition: An International Journal*, 5(1-2), 142-164.
- Charles, S. T., Mather, M. & Carstensen, L. L. (2003). Aging and emotional memory: The forgettable nature of negative images for older adults. *Journal of Experimental Psychology: General*, 132(2), 310-324.
- Chechile, R. A. (2004). New multinomial models for the Chechile-Meyer task: Comment. *Journal of Mathematical Psychology*, 48(6), 364-384.
- Chechile, R. A. & Meyer, D. L. (1976). A Bayesian procedure for separately estimating storage and retrieval components of forgetting. *Journal of Mathematical Psychology*, 13(3), 269-295.
- Christensen, H. & Henderson, A. S. (1991). Is age kinder to the initially more able? A study of eminent scientists and academics. *Psychological Medicine*, 21(4), 935-946.
- Clark, S. E. & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, 3(1), 37-60.
- Cleary, A. M. (2002). Recognition with and without identification: Dissociative effects of meaningful encoding. *Memory & Cognition*, 30(5), 758-767.
- Cohen, G. & Faulkner, D. (1989). Age differences in source forgetting: Effects on reality monitoring and on eyewitness testimony. *Psychology and Aging*, 4(1), 10-17.
- Cohn, M., Emrich, S. M. & Moscovitch, M. (2008). Age-related deficits in associative memory: The influence of impaired strategic retrieval. *Psychology and Aging*, 23(1), 93-103.
- Cohn, M. & Moscovitch, M. (2007). Dissociating measures of associative memory: Evidence and theoretical implications. *Journal of Memory and Language*, 57(3), 437-454.
- Compton, D. M., Bachman, L. D., Brand, D. & Avet, T. L. (2000). Age-associated chan-

- ges in cognitive function in highly educated adults: Emerging myths and realities. *International Journal of Geriatric Psychiatry*, 15(1), 75-85.
- Cooper, E. H. & Pantle, A. J. (1967). The total-time hypothesis in verbal learning. *Psychological Bulletin*, 68(4), 221-234.
- Craik, F. I. M. (1982). Selective changes in encoding as a function of reduced processing capacity. In F. Klix, S. Hoffman & E. Van der Meer (Hrsg.), *Cognitive research in psychology* (S. 152-161). Berlin: Deutscher Verlag der Wissenschaften.
- Craik, F. I. M. (1983). On the transfer of information from temporary to permanent memory. *Philosophical Transactions of the Royal Society of London, Series B*, 302, 341-359.
- Craik, F. I. M. (1986). A functional account of age differences in memory. In F. Klix & H. Hagendorf (Hrsg.), *Human memory and cognitive capabilities: Mechanisms and performances* (S. 409-422). North Holland: Elsevier.
- Craik, F. I. M. (2002). Levels of processing: Past, present ... and future? *Memory*, 10(5-6), 305-318.
- Craik, F. I. M. & Byrd, M. (1982). Aging and cognitive deficits: The role of attentional resources. In F. I. M. Craik & S. Trehub (Hrsg.), *Aging and cognitive processes* (Bd. 8, S. 191-211). New York: Plenum Press.
- Craik, F. I. M. & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671-684.
- Craik, F. I. M. & McDowd, J. M. (1987). Age differences in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(3), 474-479.
- Craik, F. I. M. & Rabinowitz, J. C. (1985). The effects of presentation rate and encoding task on age-related memory deficits. *Journal of Gerontology*, 40(3), 309-315.
- Craik, F. I. M. & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104(3), 268-294.
- Criss, A. H. & Shiffrin, R. M. (2004). Pairs do not suffer interference from other types of pairs or single items in associative recognition. *Memory & Cognition*, 32(8), 1284-1297.
- Curran, T., DeBuse, C. & Leynes, P. A. (2007). Conflict and criterion setting in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(1), 2-17.
- D'Argembeau, A. & Van der Linden, M. (2004). Identity but not expression memory for unfamiliar faces is affected by ageing. *Memory*, 12(5), 644-654.
- Denney, N. W., Dew, J. R. & Kihlstrom, J. F. (1992). An adult developmental study of the encoding of spatial location. *Experimental Aging Research*, 18(1-2), 25-32.
- Denney, N. W., Miller, B. V., Dew, J. R. & Levav, A. L. (1991). An adult developmental study of contextual memory. *Journals of Gerontology*, 46(2), 44-50.
- Dew, I. T. Z., Bayen, U. J. & Giovanello, K. S. (2007). Implicit relational memory in

- young and older adults. *Journal of Psychology*, 215(1), 25-34.
- Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In D. T. Stuss & R. T. Knight (Hrsg.), *Principles of frontal lobe function* (S. 466-503). New York, NY: Oxford University Press.
- Diana, R. A., Yonelinas, A. P. & Ranganath, C. (2008). The effects of unitization on familiarity-based source memory: testing a behavioral prediction derived from neuroimaging data. *Journal of experimental psychology. Learning, memory, and cognition*, 34(4), 730-40.
- Dickinson, C. M. & Rabbitt, P. M. A. (1991). Simulated visual impairment: effects on text comprehension and reading speed. *Clinical Vision Sciences*, 6(4), 301-308.
- Dodson, C. S. & Johnson, M. K. (1993). Rate of false source attributions depends on how questions are asked. *American Journal of Psychology*, 106(4), 541-557.
- Dumas, J. A. & Hartman, M. (2003). Adult age differences in temporal and item memory. *Psychology and Aging*, 18(3), 573-586.
- Dunlosky, J. & Hertzog, C. (1998). Aging and deficits in associative memory: What is the role of strategy production? *Psychology and Aging*, 13(4), 597-607.
- Dunlosky, J. & Hertzog, C. (2001). Measuring strategy production during associative learning: The relative utility of concurrent versus retrospective reports. *Memory & Cognition*, 29(2), 247-253.
- Dunlosky, J., Hertzog, C. & Powell-Moman, A. (2005). The contribution of mediator-based deficiencies to age differences in associative learning. *Developmental Psychology*, 41(2), 389-400.
- Dyne, A. M., Humphreys, M. S., Bain, J. D. & Pike, R. (1990). Associative interference effects in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(5), 813-824.
- Ebbinghaus, H. (1885). *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie*. Leipzig: K. Buehler.
- Ecker, U. K. H., Zimmer, H. D. & Groh-Bordin, C. (2007). Color and context: An ERP study on intrinsic and extrinsic feature binding in episodic memory. *Memory & Cognition*, 35(6), 1483-1501.
- Eichenbaum, H. (2002). *The cognitive neuroscience of memory: An introduction*. New York: Oxford University Press.
- Eldridge, L. L., Sarfatti, S. & Knowlton, B. J. (2002). The effect of testing procedure on remember-know judgments. *Psychonomic Bulletin & Review*, 9(1), 139-145.
- Erdfelder, E. (1987). *Die Entwicklung psychometrischer Intelligenz über die Lebensspanne*. Frankfurt am Main: Verlag Peter Lang GmbH.
- Erdfelder, E. (2000). *Multinomiale Modelle in der kognitiven Psychologie*. Universität Bonn: Habilitationsschrift.

- Erdfelder, E. & Buchner, A. (1998). Decomposing the hindsight bias: A multinomial processing tree model for separating recollection and reconstruction in hindsight. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(2), 387-414.
- Erdfelder, E., Cüpper, L., Auer, T.-S. & Undorf, M. (2007). The four-states model of memory retrieval experiences. *Zeitschrift für Psychologie/Journal of Psychology*, *215*(1), 61-71.
- Erdfelder, E. & Undorf, M. (2003). Quantitative Methoden in der Entwicklungspsychologie. In H.-H. Uslucan & A. Born (Hrsg.), *Studentexte Entwicklungspsychologie* (S. 293-321). Köln: Kölner Studien Verlag.
- Estes, W. K. & Maddox, W. T. (1995). Interactions of stimulus attributes, base rates, and feedback in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1075-1095.
- Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191.
- Ferguson, S. A., Hashtroudi, S. & Johnson, M. K. (1992). Age differences in using source-relevant cues. *Psychology and Aging*, *7*(3), 443-452.
- Flynn, J. R. (1984). The mean IQ of americans: Massive gains 1932 to 1978. *Psychological Bulletin*, *95*(1), 29-51.
- Francis, W. N. & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston, Massachusetts: Houghton Mifflin.
- Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition*, *16*(4), 309-313.
- Geiselman, R. E. & Crawley, J. M. (1983). Incidental processing of speaker characteristics: Voice as connotative information. *Journal of Verbal Learning and Verbal Behavior*, *22*(1), 15-23.
- Gilbert, J. G. (1941). Memory loss in senescence. *The Journal of Abnormal and Social Psychology*, *36*(1), 73-86.
- Giovanello, K. S., Keane, M. M. & Verfaellie, M. (2006). The contribution of familiarity to associative memory in amnesia. *Neuropsychologia*, *44*(10), 1859-65.
- Glanzer, M. & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, *13*(1), 8-20.
- Glanzer, M., Adams, J. K., Iverson, G. J. & Kim, K. (1993). The regularities of recognition memory. *Psychological Review*, *100*(3), 546-567.
- Glanzer, M. & Bowles, N. (1976). Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(1), 21-31.
- Glover, J. A., Rankin, J., Langner, N., Toderò, C. & Dinnel, D. (1985). Memory for

- sentences and prose: Levels-of-processing or transfer-appropriate-processing? *Journal of Reading Behavior*, 17(3), 215-234.
- Goshen-Gottstein, Y. & Moscovitch, M. (1995). Repetition priming effects for newly formed associations are perceptually based: Evidence from shallow encoding and format specificity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(5), 1249-1262.
- Graef, D. & Naveh-Benjamin, M. (2008). Sensory decline as a mediating factor in age-related differences in associative memory for objects. Poster presented at the Cognitive Aging Conference: Atlanta.
- Graf, P. & Schacter, D. L. (1989). Unitization and grouping mediate dissociations in memory for new associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(5), 930-940.
- Greene, R. L. (1996). Mirror effect in order and associative information: Role of response strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(3), 687-695.
- Greeno, J. G. (1968). Identifiability and statistical properties of two-stage learning with no successes in the initial stage. *Psychometrika*, 33(2), 173-215.
- Guttentag, R. E. & Hunt, R. R. (1988). Adult age differences in memory for imagined and performed actions. *Journals of Gerontology*, 43(4), P107-P108.
- Habib, R. & Nyberg, L. (1997). Incidental retrieval processes influence explicit test performance with data-limited cues. *Psychonomic Bulletin & Review*, 4(1), 130-133.
- Hager, W. & Hasselhorn, M. (1994). *Handbuch deutschsprachiger Wortnormen*. Göttingen: Hogrefe.
- Hasher, L. & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, 108(3), 356-388.
- Hasher, L. & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Hrg.), *The psychology of learning and motivation: Advances in research and theory* (Bd. 22, S. 193-225). San Diego, CA: Academic Press.
- Hasher, L., Zacks, R. T. & May, C. P. (1999). Inhibitory control, circadian arousal, and age. In D. Gopher & A. Koriat (Hrsg.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (S. 653-675). Cambridge, MA: MIT Press.
- Hashtroudi, S., Johnson, M. K. & Chrosniak, L. D. (1989). Aging and source monitoring. *Psychology and Aging*, 4(1), 106-112.
- Healy, A. F. & Kubovy, M. (1978). The effects of payoffs and prior probabilities on indices of performance and cutoff location in recognition memory. *Memory & Cognition*, 6(5), 544-553.

- Healy, M. R., Light, L. L. & Chung, C. (2005). Dual-process models of associative recognition in young and older adults: Evidence from receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(4), 768-788.
- Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.
- Heit, E., Brockdorff, N. & Lamberts, K. (2003). Adaptive changes of response criterion in recognition memory. *Psychonomic Bulletin & Review*, 10(3), 718-723.
- Hertzog, C. & Dixon, R. A. (1996). Methodological issues in research on cognition and aging. In F. Blanchard-Fields & T. M. Hess (Hrsg.), *Perspectives on cognitive change in adulthood and aging* (S. 66-121). New York: McGraw-Hill.
- Hess, T. M. & Slaughter, S. J. (1990). Schematic knowledge influences on memory for scene information in young and older adults. *Developmental Psychology*, 26(5), 855-865.
- Hicks, J. L. & Marsh, R. L. (1999). Remember-know judgments can depend on how memory is tested. *Psychonomic Bulletin & Review*, 6(1), 117-122.
- Hilford, A., Glanzer, M. & Kim, K. (1997). Encoding, repetition, and the mirror effect in recognition memory: Symmetry in motion. *Memory & Cognition*, 25(5), 593-605.
- Hirshman, E. (1995). Decision processes in recognition memory: Criterion shifts and the list-strength paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(2), 302-313.
- Hirshman, E. & Henzler, A. (1998). The role of decision processes in conscious recollection. *Psychological Science*, 9(1), 61-65.
- Hockley, W. E. (1991). Recognition memory for item and associative information: A comparison of forgetting rates. In W. E. Hockley & S. Lewandowsky (Hrsg.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (S. 227-248). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.
- Hockley, W. E. (1992). Item versus associative information: Further comparisons of forgetting rates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(6), 1321-1330.
- Hockley, W. E. (1994). Reflections of the mirror effect for item and associative recognition. *Memory & Cognition*, 22(6), 713-722.
- Hockley, W. E. & Consoli, A. (1999). Familiarity and recollection in item and associative recognition. *Memory & Cognition*, 27(4), 657-664.
- Hockley, W. E. & Niewiadomski, M. W. (2007). Strength-based mirror effects in item and associative recognition: Evidence for within-list criterion changes. *Memory & Cognition*, 35(4), 679-688.
- Howard, M. W. & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46(3), 269-299.
- Howard, M. W., Kahana, M. J. & Wingfield, A. (2006). Aging and contextual binding:

- Modeling recency and lag recency effects with the temporal context model. *Psychonomic Bulletin & Review*, 13(3), 439-445.
- Howe, M. L. (1988). Measuring memory development in adulthood: A model-based approach to disentangling storage-retrieval contributions. In M. L. Howe & C. J. Brainerd (Hrsg.), *Cognitive development in adulthood* (S. 39-64). New York, NY: Springer.
- Howe, M. L. & Brainerd, C. J. (1989). Development of children's long-term retention. *Developmental Review*, 9(4), 301-340.
- Howe, M. L. & Hunter, M. A. (1985). Adult age differences in storage-retrieval processes: A stages-of-learning analysis of developmental interactions in concreteness effects. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, 39(1), 130-150.
- Howe, M. L. & Hunter, M. A. (1986). Long-term memory in adulthood: An examination of the development of storage and retrieval processes at acquisition and retention. *Developmental Review*, 6(4), 334-364.
- Hu, X. & Batchelder, W. H. (1994). The statistical analysis of engineering processing tree models with the EM algorithm. *Psychometrika*, 59(1), 21-47.
- Hulicka, I. M. & Grossman, J. L. (1967). Age-group comparisons for the use of mediators in paired-associate learning. *The Journals of Gerontology*, 22, 46-51.
- Humphreys, M. S. (1976). Relational information and the context effect in recognition memory. *Memory & Cognition*, 4(2), 221-232.
- Humphreys, M. S. (1978). Item and relational information: A case for context independent retrieval. *Journal of Verbal Learning and Verbal Behavior*, 17(2), 175-187.
- Hyde, T. S. & Jenkins, J. J. (1973). Recall for words as a function of semantic, graphic, and syntactic orienting tasks. *Journal of Verbal Learning and Verbal Behavior*, 12(5), 471-480.
- Intraub, H. & Hoffman, J. E. (1992). Reading and visual memory: Remembering scenes that were never seen. *American Journal of Psychology*, 105(1), 101-114.
- Jack, W. H. (1968). Paired-associate learning with simultaneous and sequential presentations. *Journal of Experimental Psychology*, 76(4), 574-578.
- Jacoby, L. L., Shimizu, Y., Daniels, K. A. & Rhodes, M. G. (2005). Modes of cognitive control in recognition and source memory: Depth of retrieval. *Psychonomic Bulletin & Review*, 12(5), 852-857.
- Jäger, T. & Mecklinger, A. (2009). Familiarity supports associative recognition memory for face stimuli that can be unitized: Evidence from receiver operating characteristics. *European Journal of Cognitive Psychology*, 21(1), 35-60.
- Jäger, T., Mecklinger, A. & Kipp, K. H. (2006). Intra- and inter-item associations doubly dissociate the electrophysiological correlates of familiarity and recollection. *Neuron*, 52(3), 535-545.
- James, L. E., Fogler, K. A. & Tauber, S. K. (2008). Recognition memory measures yield

- disproportionate effects of aging on learning face-name associations. *Psychology and Aging*, 23(3), 657-664.
- Java, R. I. (1992). Priming and aging: Evidence of preserved memory function in an anagram solution task. *American Journal of Psychology*, 105(4), 541-548.
- Java, R. I., Gregg, V. H. & Gardiner, J. M. (1997). What do people actually remember (and know) in „remember/know“ experiments? *European Journal of Cognitive Psychology*, 9(2), 187-197.
- Jenkins, J. R., Neale, D. C. & Deno, S. L. (1967). Differential memory for picture and word stimuli. *Journal of Educational Psychology*, 58(5), 303-307.
- Jones, G. V. (1984). Fragment and schema models for recall. *Memory & Cognition*, 12(3), 250-263.
- Kahneman, D., Treisman, A. & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, 24(2), 175-219.
- Kane, M. J. & Hasher, L. (1995). Interference. In G. Maddox (Hrg.), *Encyclopedia of aging* (2. Aufl., S. 514-516). New York, NY: Springer Publishing Co.
- Kane, M. J., Hasher, L., Stoltzfus, E. R. & Zacks, R. T. (1994). Inhibitory attentional mechanisms and aging. *Psychology and Aging*, 9(1), 103-112.
- Kapucu, A., Rotello, C. M., Ready, R. E. & Seidl, K. N. (2008). Response bias in „remembering“ emotional stimuli: A new perspective on age differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(3), 703-711.
- Kausler, D. H. (1994). *Learning and memory in normal aging*. New York: Academic Press.
- Kausler, D. H., Lichty, W. & Davis, R. T. (1985). Temporal memory for performed activities: Intentionality and adult age differences. *Developmental Psychology*, 21(6), 1132-1138.
- Kausler, D. H. & Phillips, P. L. (1988). Instructional variation and adult age differences in activity memory. *Experimental Aging Research*, 14(4), 195-199.
- Kausler, D. H. & Puckett, J. M. (1980). Adult age differences in recognition memory for a nonsemantic attribute. *Experimental Aging Research*, 6(4), 349-355.
- Kausler, D. H. & Puckett, J. M. (1981a). Adult age differences in memory for modality attributes. *Experimental Aging Research*, 7(2), 117-125.
- Kausler, D. H. & Puckett, J. M. (1981b). Adult age differences in memory for sex of voice. *Journal of Gerontology*, 36(1), 44-50.
- Kausler, D. H., Salthouse, T. A. & Sauls, J. S. (1988). Temporal memory over the adult lifespan. *American Journal of Psychology*, 101(2), 207-215.
- Kelley, R. & Wixted, J. T. (2001). On the nature of associative information in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(3), 701-722.
- Kersten, A. W., Earles, J. L., Curtayne, E. S. & Lane, J. C. (2008). Adult age differences

- in binding actors and actions in memory for events. *Memory & Cognition*, *36*(1), 119-131.
- Kilb, A. & Naveh-Benjamin, M. (2007). Paying attention to binding: Further studies assessing the role of reduced attentional resources in the associative deficit of older adults. *Memory & Cognition*, *35*(5), 1162-1174.
- Kim, S., Hasher, L. & Zacks, R. T. (2007). Aging and benefit of distractibility. *Psychonomic Bulletin & Review*, *14*(2), 301-305.
- Kim, S., Healey, M. K., Goldstein, D., Hasher, L. & Wiprzycka, U. J. (2008). Age differences in choice satisfaction: A positivity effect in decision making. *Psychology and Aging*, *23*(1), 33-38.
- Klauer, K. C. (2006). Hierarchical multinomial processing tree models: A latent-class approach. *Psychometrika*, *71*(1), 7-31.
- Klauer, K. C., Musch, J. & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review*, *107*(4), 852-884.
- Klauer, K. C., Stahl, C. & Erdfelder, E. (2007). The abstract selection task: New data and an almost comprehensive model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 680-703.
- Klauer, K. C. & Wegener, I. (1998). Unraveling social categorization in the „Who said what?“ paradigm. *Journal of Personality and Social Psychology*, *75*(5), 1155-1178.
- Klauer, K. C., Wegener, I. & Ehrenberg, K. (2002). Perceiving minority members as individuals: The effects of relative group size in social categorization. *European Journal of Social Psychology*, *32*(2), 223-245.
- Kliegl, R. & Lindenberger, U. (1993). Modeling intrusions and correct recall in episodic memory: Adult age differences in encoding of list context. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 617-637.
- Kliegl, R., Smith, J. & Baltes, P. B. (1989). Testing-the-limits and the study of adult age differences in cognitive plasticity of a mnemonic skill. *Developmental Psychology*, *25*(2), 247-256.
- Köhler, W. (1947). *Gestalt Psychology*. New York: Liveright.
- Labouvie-Vief, G. (1985). Intelligence and cognition. In J. E. Birren & K. W. Schaie (Hrsg.), *Handbook of the psychology of aging* (2. Aufl., S. 500-530). New York, NY: Van Nostrand Reinhold Co.
- Lehman, E. B. & Mellinger, J. C. (1984). Effects of aging on memory for presentation modality. *Developmental Psychology*, *20*(6), 1210-1217.
- Li, K. Z. H. & Lindenberger, U. (2002). Relations between aging sensory/sensorimotor and cognitive functions. *Neuroscience & Biobehavioral Reviews*, *26*(7), 777-783.
- Li, S.-C. & Baltes, P. B. (2006). Cognitive developmental research from lifespan perspectives: The challenge of integration. In E. Bialystok & F. I. M. Craik (Hrsg.), *Lifespan cognition: Mechanisms of change* (S. 344-363). New York, NY: Oxford University

Press.

- Light, L. L. (1991). Memory and aging: Four hypotheses in search of data. In M. R. Rosenzweig & L. W. Porter (Hrsg.), *Annual review of psychology* (Bd. 42, S. 333-376). Palo Alto, CA: Annual Reviews.
- Light, L. L. & Berger, D. E. (1976). Are there long-term literal copies of visually presented words? *Journal of Experimental Psychology: Human Learning and Memory*, 2(6), 654-662.
- Light, L. L., Patterson, M. M., Chung, C. & Healy, M. R. (2004). Effects of repetition and response deadline on associative recognition in young and older adults. *Memory & Cognition*, 32(7), 1182-1193.
- Light, L. L., Prull, M. W., La Voie, D. J. & Healy, M. R. (2000). Dual-process theories of memory in old age. In T. J. Perfect & E. A. Maylor (Hrsg.), *Models of cognitive aging* (S. 238-300). New York, NY: Oxford University Press.
- Light, L. L. & Singh, A. (1987). Implicit and explicit memory in young and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(4), 531-541.
- Light, L. L., Singh, A. & Capps, J. L. (1986). Dissociation of memory and awareness in young and older adults. *Journal of Clinical and Experimental Neuropsychology*, 8(1), 62-74.
- Light, L. L. & Zelinski, E. M. (1983). Memory for spatial information in young and old adults. *Developmental Psychology*, 19(6), 901-906.
- Lindenberger, U. & Baltes, P. B. (1994). Sensory functioning and intelligence in old age: A strong connection. *Psychology and Aging*, 9(3), 339-355.
- Lindenberger, U., Scherer, H. & Baltes, P. B. (2001). The strong connection between sensory and cognitive performance in old age: Not due to sensory acuity reductions operating during cognitive assessment. *Psychology and Aging*, 16(2), 196-205.
- Lindsay, D. S. & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition*, 17(3), 349-358.
- Lipman, P. D. & Caplan, L. J. (1992). Adult age differences in memory for routes: Effects of instruction and spatial diagram. *Psychology and Aging*, 7(3), 435-442.
- Lövden, M., Li, S.-C., Shing, Y. L. & Lindenberger, U. (2007). Within-person trial-to-trial variability precedes and predicts cognitive decline in old and very old age: Longitudinal data from the berlin aging study. *Neuropsychologia*, 45(12), 2827-2838.
- Lövden, M. & Lindenberger, U. (2005). Development of intellectual abilities in old age: From age gradients to individuals. In O. Wilhelm & R. W. Engle (Hrsg.), *Handbook of understanding and measuring intelligence* (S. 203-221). Thousand Oaks, CA: Sage Publications, Inc.
- Lustig, C., Hasher, L. & Tonev, S. T. (2001). Inhibitory control over the present and the past. *European Journal of Cognitive Psychology*, 31(1-2), 107-122.

- Lustig, C., May, C. P. & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General*, *130*(2), 199-207.
- Lyle, K. B., Bloise, S. M. & Johnson, M. K. (2006). Age-related binding deficits and the content of false memories. *Psychology and Aging*, *21*(1), 86-95.
- Macho, S. (2004). Modeling associative recognition: A comparison of two-high-threshold, two-high-threshold signal detection, and mixture distribution models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 83-97.
- Macmillan, N. A. & Creelman, C. D. (1990). Response bias: Characteristics of detection theory, threshold theory, and „nonparametric“ indexes. *Psychological Bulletin*, *107*(3), 401-413.
- Maddox, W. T. & Bohil, C. J. (1998). Base-rate and payoff effects in multidimensional perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1459-1482.
- Marsh, R. L. & Hicks, J. L. (1998). Test formats change source-monitoring decision processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(5), 1137- 1151.
- Martin, C. J., Boersma, F. J. & Cox, D. L. (1965). A classification of associative strategies in paired-associate learning. *Psychonomic Science*, *3*(10), 455-456.
- Mata, R. (2007). Understanding the aging decision maker. *Human Development*, *50*(6), 359-366.
- Mata, R., Schooler, L. J. & Rieskamp, J. (2007). The aging decision maker: Cognitive aging and the adaptive selection of decision strategies. *Psychology and Aging*, *22*(4), 796-810.
- Mather, M. & Carstensen, L. L. (2005). Aging and motivated cognition: The positivity effect in attention and memory. *Trends in Cognitive Sciences*, *9*(10), 496-502.
- Mather, M. & Johnson, M. K. (2000). Choice-supportive source monitoring: Do our decisions seem better to us as we age? *Psychology and Aging*, *15*(4), 596-606.
- Mather, M., Knight, M. & McCaffrey, M. (2005). The allure of the alignable: Younger and older adults' false memories of choice features. *Journal of Experimental Psychology: General*, *134*(1), 38-51.
- Max Planck Institute for Psycholinguistics. (1991). *The CELEX lexical database (German Version 2.0) [CD-ROM]*. Nijmegen, The Netherlands: Autor.
- May, C. P., Hasher, L. & Kane, M. J. (1999). The role of interference in memory span. *Memory & Cognition*, *27*(5), 759-767.
- May, C. P., Zacks, R. T., Hasher, L. & Multhaup, K. S. (1999). Inhibition in the processing of garden-path sentences. *Psychology and Aging*, *14*(2), 304-313.
- McCabe, J. & Hartman, M. (2008). An analysis of age differences in perceptual speed. *Memory & Cognition*, *36*(8), 1495-1508.
- McDaniel, M. A. & Kearney, E. M. (1984). Optimal learning strategies and their spon-

- taneous use: The importance of task-appropriate processing. *Memory & Cognition*, 12(4), 361-373.
- McDougall, S. & Velmans, M. (1993). Encoding strategy dynamics: When relationships between words determine strategy use. *British Journal of Psychology*, 84(2), 227-248.
- Mecklinger, A. (2000). Interfacing mind and brain: A neurocognitive model of recognition memory. *Psychophysiology*, 37(5), 565-582.
- Meiser, T. (2005). A hierarchy of multinomial models for multidimensional source monitoring. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 1(1), 2-17.
- Meiser, T. & Bröder, A. (2002). Memory for multidimensional source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(1), 116-137.
- Meiser, T. & Hewstone, M. (2004). Cognitive processes in stereotype formation: The role of correct contingency learning for biased group judgments. *Journal of Personality and Social Psychology*, 87(5), 599-614.
- Meiser, T. & Hewstone, M. (2006). Illusory and spurious correlations: Distinct phenomena or joint outcomes of exemplar-based category learning? *European Journal of Social Psychology*, 36(3), 315-336.
- Meiser, T., Sattler, C. & von Hecker, U. (2007). Metacognitive inferences in source memory judgements: The role of perceived differences in item recognition. *The Quarterly Journal of Experimental Psychology*, 60(7), 1015-1040.
- Miller, E. K. & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167-202.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.
- Miller, P. H. (2000). How best to utilize a deficiency. *Child Development*, 71(4), 1013-1017.
- Mitchell, K. J., Johnson, M. K., Raye, C. L., Mather, M. & D'Esposito, M. (2000). Aging and reflective processes of working memory: Binding and test load deficits. *Psychology and Aging*, 15(3), 527-541.
- Montague, W. E. & Wearing, A. J. (1967). The complexity of natural language mediators and its relation to paired-associate learning. *Psychonomic Science*, 7(4), 135-136.
- Morris, C. D., Bransford, J. D. & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, 16(5), 519-533.
- Moshagen, M. (2009). *multiTree: A software program for the analysis of multinomial processing tree models (Version 0.37) [Computer Software]*. Beziehbar über Lehrstuhl für Psychologie III, Universität Mannheim: <http://psycho3.uni-mannheim.de/uploads/software/multiTree/multiTree.html>.

- Mueller, D. J. & Travers, R. M. (1965). Temporal relations and meaningfulness in paired-associate learning. *Psychological Reports*, *17*(2), 491-497.
- Mulligan, N. W. & Hirshman, E. (1997). Measuring the bases of recognition memory: An investigation of the process-dissociation framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(2), 280-304.
- Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, *89*(6), 609-626.
- Murnane, K., Phelps, M. P. & Malmberg, K. (1999). Context-dependent recognition memory: The ice theory. *Journal of Experimental Psychology: General*, *128*(4), 403-415.
- Murphy, D. R., Craik, F. I. M., Li, K. Z. H. & Schneider, B. A. (2000). Comparing the effects of aging and background noise of short-term memory performance. *Psychology and Aging*, *15*(2), 323-334.
- Myerson, J., Hale, S., Wagstaff, D., Poon, L. & Smith, G. (1990). The information-loss model: A mathematical theory of age-related cognitive slowing. *Psychological Review*, *97*, 475-487.
- Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1170-1187.
- Naveh-Benjamin, M. (2006). Binding of memories: Adult-age differences and the effects of divided attention in young adults on episodic memory. In H. D. Zimmer, M. A. & U. Lindenberger (Hrsg.), *Handbook of binding and memory: Perspectives from cognitive neuroscience* (S. 627-656). New York: Oxford University Press.
- Naveh-Benjamin, M. (2008a). Assessing potential factors that mediate age-related associative deficits in episodic memory. *International Journal of Psychology*, *43*(3/4), 377.
- Naveh-Benjamin, M. (2008b). Persönliche Mitteilung. Cognitive Aging Conference: Atlanta.
- Naveh-Benjamin, M., Brav, T. K. & Levy, O. (2007). The associative memory deficit of older adults: The role of strategy utilization. *Psychology and Aging*, *22*(1), 202-208.
- Naveh-Benjamin, M., Guez, J., Kilb, A. & Reedy, S. (2004). The associative memory deficit of older adults: Further support using face-name associations. *Psychology and Aging*, *19*(3), 541-546.
- Naveh-Benjamin, M., Guez, J. & Marom, M. (2003). The effects of divided attention at encoding on item and associative memory. *Memory & Cognition*, *31*(7), 1021-1035.
- Naveh-Benjamin, M., Guez, J. & Shulman, S. (2004). Older adults' associative deficit in episodic memory: Assessing the role of decline in attentional resources. *Psychonomic Bulletin & Review*, *11*(6), 1067-1073.
- Naveh-Benjamin, M., Hussain, Z., Guez, J. & Bar-On, M. (2003). Adult age differences

- in episodic memory: Further support for an associative-deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(5), 826-837.
- Naveh-Benjamin, M., Kilb, A. & Hara, Y. (2007). Is older adult's associative memory deficit mediated by age-related sensory decline? In *Abstracts of the 48th Annual Meeting of the Psychonomic Society* (S. 41). Long Beach, California: Psychonomic Society.
- Naveh-Benjamin, M., Shing, Y. L., Kilb, A., Werkle-Bergner, M., Lindenberger, U. & Shu-ChenLi. (2009). Adult age differences in memory for name-face associations: The effects of intentional and incidental learning. *Memory*, 17(2), 220-232.
- Newman, M. C., Allen, J. J. B. & Kaszniak, A. W. (2001). Tasks for assessing memory for temporal order versus memory for items in aging. *Aging, Neuropsychology, and Cognition*, 8(1), 72-78.
- Nyberg, L., Olofsson, U., Gardiner, J. M. & Nilsson, L.-G. (1997). Assessment of retrieval strategy in incidental, intentional, and inclusion tests with word-fragment cues. *Psychologische Forschung*, 59(4), 231-239.
- Ofen, N., Kao, Y.-C., Sokol-Hessner, P., Kim, H., Whitfield-Gabrieli, S. & Gabrieli, J. D. E. (2007). Development of the declarative memory system in the human brain. *Nature Neuroscience*, 10(9), 1198-1205.
- Old, S. R. & Naveh-Benjamin, M. (2008a). Differential effects of age on item and associative measures of memory: A meta-analysis. *Psychology and Aging*, 23(1), 104-118.
- Old, S. R. & Naveh-Benjamin, M. (2008b). Memory for people and their actions: Further evidence for an age-related associative deficit. *Psychology and Aging*, 23(2), 467-472.
- Opitz, B. & Cornell, S. (2006). Contribution of familiarity and recollection to associative recognition memory: Insights from event-related potentials. *Journal of Cognitive Neuroscience*, 18(9), 1595-1605.
- Paivio, A. & Csapo, K. (1969). Concrete image and verbal memory codes. *Journal of Experimental Psychology*, 80(2), 279-285.
- Paivio, A., Yuille, J. C. & Smythe, P. C. (1966). Stimulus and response abstractness, imagery, and meaningfulness, and reported mediators in paired-associate learning. *Canadian Journal of Psychology*, 20(4), 362-377.
- Palfai, T., Halperin, S. & Hoyer, W. J. (2003). Age inequalities in recognition memory: Effects of stimulus presentation time and list repetitions. *Aging, Neuropsychology, and Cognition*, 10(2), 134-140.
- Pantelis, P. C., Van Vugt, M. K., Sekuler, R., Wilson, H. R. & Kahana, M. J. (2008). Why are some people's names easier to learn than others? The effects of face similarity on memory for face-name associations. *Memory & Cognition*, 36(6), 1182-1195.
- Park, D. C. & Puglisi, J. T. (1985). Older adults' memory for the color of pictures and words. *Journal of Gerontology*, 40(2), 198-204.
- Park, D. C., Puglisi, J. T. & Lutz, R. (1982). Spatial memory in older adults: Effects of

- intentionality. *Journal of Gerontology*, *37*(3), 330-335.
- Park, D. C., Puglisi, J. T. & Sovacool, M. (1984). Picture memory in older adults: Effects of contextual detail at encoding and retrieval. *Journal of Gerontology*, *39*(2), 213-215.
- Paus, T. (2005). Mapping brain maturation and cognitive development during adolescence. *Trends in Cognitive Sciences*, *9*(2), 60-68.
- Penick, S. & Solomon, P. R. (1991). Hippocampus, context, and conditioning. *Behavioral neuroscience*, *105*(5), 611-7.
- Phillips, R. G. & LeDoux, J. E. (1992). Differential contribution of amygdala and hippocampus to cued and contextual fear conditioning. *Behavioral neuroscience*, *106*(2), 274-85.
- Pohl, R. F. & Auer, T.-S. (2008). Rückschaufehler über die lebensspanne. Kolloquium Lehrstuhl Psychologie III: Mannheim.
- Potter, M. C. & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology*, *81*(1), 10-15.
- Primoff, E. (1938). Backward and forward association as an organizing act in serial and in paired associate learning. *Journal of Psychology: Interdisciplinary and Applied*, *5*, 375-395.
- Prior, A. & Bentin, S. (2003). Incidental formation of episodic associations: The importance of sentential context. *Memory & Cognition*, *31*(2), 306-316.
- Provyn, J. P., Sliwinski, M. J. & Howard, M. W. (2007). Effects of age on contextually mediated associations in paired associate learning. *Psychology and Aging*, *22*(4), 846-857.
- Psychological Software Tools, I. (2000). *Eprime (Version 1.2) [Computer Software]*. Pittsburgh, PA: Autor.
- Puglisi, J. T., Park, D. C., Smith, A. D. & Hill, G. W. (1985). Memory for two types of spatial location: Effects of instructions, age, and format. *American Journal of Psychology*, *98*, 101-118.
- Quamme, J. R., Yonelinas, A. P. & Norman, K. A. (2007). Effect of unitization on associative recognition in amnesia. *Hippocampus*, *17*(3), 192-200.
- Rabbitt, P. (1966). Recognition: Memory for words correctly heard in noise. *Psychonomic Science*, *6*(8), 383-384.
- Rabinowitz, J. C. (1989). Judgments of origin and generation effects: Comparisons between young and elderly adults. *Psychology and Aging*, *4*(3), 259-268.
- Rabinowitz, J. C., Craik, F. I. M. & Ackerman, B. P. (1982). A processing resource account of age differences in recall. *Canadian Journal of Psychology*, *36*(2), 325-344.
- Radvansky, G. A., Zacks, R. T. & Hasher, L. (1996). Fact retrieval in younger and older adults: The role of mental models. *Psychology and Aging*, *11*(2), 258-271.
- Rahhal, T. A., Hasher, L. & Colcombe, S. J. (2001). Instructional manipulations and age differences in memory: Now you see them, now you don't. *Psychology and Aging*,

- 16(4), 697-706.
- Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition*, 21(1), 89-102.
- Ratcliff, R., Sheu, C.-F. & Gronlund, S. D. (1992). Testing global memory models using roc curves. *Psychological Review*, 99(3), 518-535.
- Raz, N. (2000). Aging of the brain and its impact on cognitive performance: Integration of structural and functional findings. In F. I. M. Craik & T. A. Salthouse (Hrsg.), *The handbook of aging and cognition* (2. Aufl., S. 1-90). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Raz, N. (2005). The aging brain observed in vivo: Differential changes and their modifiers. In R. Cabeza, L. Nyberg & D. Park (Hrsg.), *Cognitive neuroscience of aging: Linking cognitive and cerebral aging* (S. 19-57). New York, NY: Oxford University Press.
- Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P. & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(2), 294-320.
- Reed, H. B. (1918a). Associative aids II: Their relation to practice and the transfer of training. *Psychological Review*, 25, 257-285.
- Reed, H. B. (1918b). Associative aids I: Their relation to learning, retention, and other associations. *Psychological Review*, 25, 128-155.
- Rhodes, M. G., Castel, A. D. & Jacoby, L. L. (2008). Associative recognition of face pairs by younger and older adults: The role of familiarity-based processing. *Psychology and Aging*, 23(2), 239-249.
- Rhodes, S. M. & Donaldson, D. I. (2007). Electrophysiological evidence for the influence of unitization on the processes engaged during episodic retrieval: Enhancing familiarity based remembering. *Neuropsychologia*, 45(2), 412-424.
- Rhodes, S. M. & Donaldson, D. I. (2008). Electrophysiological evidence for the effect of interactive imagery on episodic memory: Encouraging familiarity for non-unitized stimuli during associative recognition. *Neuro Image*, 39(2), 873-884.
- Richardson, J. T. E. (1998). The availability and effectiveness of reported mediators in associative learning: A historical review and an experimental investigation. *Psychonomic Bulletin & Review*, 5(4), 597-614.
- Riefer, D. M. & Batchelder, W. H. (1991). Age differences in storage and retrieval: A multinomial modeling analysis. *Bulletin of the Psychonomic Society*, 29(5), 415-418.
- Riefer, D. M. & Batchelder, W. H. (1995). A multinomial modeling analysis of the recognition-failure paradigm. *Memory & Cognition*, 23(5), 611-630.
- Riefer, D. M., Chien, Y. & Reimer, J. F. (2007). Positive and negative generation effects in source monitoring. *The Quarterly Journal of Experimental Psychology*, 60(10),

1389-1405.

- Riefer, D. M., Knapp, B. R., Batchelder, W. H., Bamber, D. & Manifold, V. (2002). Cognitive psychometrics: Assessing storage and retrieval deficits in special populations with multinomial processing tree models. *Psychological Assessment*, *14*(2), 184-201.
- Riefer, D. M. & Rouder, J. N. (1992). A multinomial modeling analysis of the mnemonic benefits of bizarre imagery. *Memory & Cognition*, *20*(6), 601-611.
- Riegel, K. F. (1967). *Der sprachliche Leistungstest SASKA: Synonym- Antonym- Selektions- Klassifikations- Analogietests*. Göttingen: Hogrefe.
- Roberts, W. A. (1972). Free recall of word lists varying in length and rate of presentation: A test of total-time hypotheses. *Journal of Experimental Psychology*, *92*(3), 365-372.
- Rogers, W. A., Hertzog, C. & Fisk, A. D. (2000). An individual differences analysis of ability and strategy influences: Age-related differences in associative learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(2), 359-394.
- Rotello, C. M. & Heit, E. (1999). Two-process models of recognition memory: Evidence for recall-to-reject? *Journal of Memory and Language*, *40*, 432-453.
- Rotello, C. M. & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition*, *28*(6), 907-922.
- Rotello, C. M., Macmillan, N. A., Hicks, J. L. & Hautus, M. J. (2006). Interpreting the effects of response bias on remember-know judgments using signal detection and threshold models. *Memory & Cognition*, *34*(8), 1598-1614.
- Rothkegel, R. (1999). Appletree: A multinomial processing tree modeling program for macintosh computers. *Behavior Research Methods, Instruments & Computers*, *31*(4), 696-700.
- Rouder, J. & Batchelder, W. (1998). Multinomial models for measuring storage and retrieval processes in paired associate learning. In C. E. Dowling, F. S. Roberts & T. P. (Hrsg.), *Recent progress in mathematical psychology*. NJ: Lawrence Erlbaum Associates.
- Rovee-Collier, C. & Dufault, D. (1991). Multiple contexts and memory retrieval at three months. *Developmental Psychobiology*, *24*(1), 39-49.
- Rowe, E. J. & Rogers, T. B. (1975). Effects of concurrent auditory shadowing on free recall and recognition of pictures and words. *Journal of Experimental Psychology: Human Learning and Memory*, *1*(4), 415-422.
- Russo, R., Cullis, A. M. & Parkin, A. J. (1998). Consequences of violating the assumption of independence in the process dissociation procedure: A word fragment completion study. *Memory & Cognition*, *26*(4), 617-632.
- Salthouse, T. A. (1982). *Adult cognition: An experimental psychology of human aging*. New York: Springer.
- Salthouse, T. A. (1991). *Theoretical perspectives on cognitive aging*. NJ, England: La-

- wrence Erlbaum Associates, Inc.
- Salthouse, T. A. (1992). Working memory mediation of adult age differences in integrative reasoning. *Memory & Cognition*, 20, 413-423.
- Salthouse, T. A. (1993). Speed mediation of adult age differences in cognition. *Developmental Psychology*, 29(4), 722-739.
- Salthouse, T. A. (1995). Selective influences on age and speed on associative memory. *The American Journal of Psychology*, 108(3), 381-396.
- Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, 103(3), 403-428.
- Salthouse, T. A. (2006). Mental exercise and mental aging. *Perspectives on Psychological Science*, 1(1), 68-87.
- Salthouse, T. A. (2007). Reply to Schooler: Consistent is not conclusive. *Perspectives on Psychological Science*, 2(1), 30-32.
- Schacter, D. L., Kaszniak, A. W., Kihlstrom, J. F. & Valdiserri, M. (1991). The relation between source memory and aging. *Psychology and Aging*, 6(4), 559-568.
- Schacter, D. L., Osowiecki, D., Kaszniak, A. W., Kihlstrom, J. F. & Valdiserri, M. (1994). Source memory: Extending the boundaries of age-related deficits. *Psychology and Aging*, 9(1), 81-89.
- Schaie, K. W. (1965). A general model for the study of developmental problems. *Psychological Bulletin*, 64(2), 92-107.
- Schaie, K. W. (1975). Age changes in adult intelligence. In D. Wooderuff & J. E. Birren (Hrsg.), *Aging: Scientific perspectives and social issues* (S. 111-124). New York: Van Nostrand Reinhold.
- Schaie, K. W. (1994). Developmental designs revisited. In S. H. Cohen & H. W. Reese (Hrsg.), *Life-span developmental psychology: Methodological contributions* (S. 45-64). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.
- Schaie, K. W. (2005). *Developmental influences on adult intelligence: The Seattle longitudinal study*. New York: Oxford University Press.
- Schmitter-Edgecombe, M. & Simpson, A. L. (2001). Effects of age and intentionality on content memory and temporal memory for performed activities. *Aging, Neuropsychology, and Cognition*, 8(2), 81-97.
- Schneider, B. A. & Pichora-Fuller, M. K. (2000). Implications for perceptual deterioration for cognitive aging research. In F. I. M. Craik & T. A. Salthouse (Hrsg.), *The handbook of aging and cognition* (2. Aufl., S. 155-219). New Jersey: Erlbaum.
- Schooler, C. (2007). Use it—and keep it, longer, probably: A reply to Salthouse (2006). *Perspectives on Psychological Science*, 2(1), 24-29.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.
- Searcy, J. H. & Bartlett, J. C. (1996). Inversion and processing of component and spatial-relational information in faces. *Journal of Experimental Psychology: Human Perception*

- tion and Performance*, 22(4), 904-915.
- Shimamura, A. P., Berry, J. M., Mangels, J. A., Rusting, C. L. & Jurica, P. (1995). Memory and cognitive abilities in university professors: Evidence for successful aging. *Psychological Science*, 6(5), 271-277.
- Shing, Y. L., Werkle-Bergner, M., Li, S.-C. & Lindenberger, U. (2008). Associative and strategic components of episodic memory: A life-span dissociation. *Journal of Experimental Psychology: General*, 137(3), 495-513.
- Simons, J. S., Dodson, C. S., Bell, D. & Schacter, D. L. (2004). Specific- and partial-source memory: Effects of aging. *Psychology and Aging*, 19(4), 689-694.
- Simons, J. S. & Spiers, H. J. (2003). Prefrontal and medial temporal lobe interactions in long-term memory. *Nature Reviews Neuroscience*, 4(8), 637-648.
- Slamecka, N. J. (1976). An analysis of double-function lists. *Memory & Cognition*, 4(5), 581-585.
- Slamecka, N. J. & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, 4(6), 592-604.
- Smith, R. E. & Bayen, U. J. (2004). A multinomial model of event-based prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(4), 756-777.
- Snodgrass, J. G. & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6(2), 174-215.
- Spaniol, J. & Bayen, U. J. (2004). Formal modeling in research on episodic memory and aging. *Psychology Science*, 46(4), 477-513.
- Spaniol, J. & Bayen, U. J. (2005). Aging and conditional probability judgments: A global matching approach. *Psychology and Aging*, 20(1), 165-181.
- Spencer, W. D. & Raz, N. (1995). Differential effects of aging on memory for content and context: A meta-analysis. *Psychology and Aging*, 10(4), 527-539.
- Speranza, F., Daneman, M. & Schneider, B. A. (2000). How aging affects the reading of words in noisy backgrounds. *Psychology and Aging*, 15(2), 253-258.
- Stahl, C. & Klauer, K. C. (2007). HMMTree: A computer program for latent-class hierarchical multinomial processing tree models. *Behavior Research Methods*, 39(2), 267-273.
- Starns, J. J., Hicks, J. L., Brown, N. L. & Martin, B. A. (2008). Source memory for unrecognized items: Predictions from multivariate signal detection theory. *Memory & Cognition*, 36(1), 1-8.
- Strack, F. & Forster, J. (1995). Reporting recollective experiences: Direct access to memory systems? *Psychological Science*, 6(6), 352-358.
- Sward, K. (1945). Age and mental ability in superior men. *American Journal of Psychology*, 58, 443-479.

- Thapar, A. & Greene, R. L. (1994). Effects of level of processing on implicit and explicit tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(3), 671-679.
- Thompson, R. A. & Nelson, C. A. (2001). Developmental science and the media: Early brain development. *American Psychologist*, 56(1), 5-15.
- Toth, J. P. (1996). Conceptual automaticity in recognition memory: Levels-of-processing effects on familiarity. *Canadian Journal of Experimental Psychology*, 50(1), 123-138.
- Treat, N. J. & Reese, H. W. (1976). Age, pacing, and imagery in paired-associate learning. *Developmental Psychology*, 12(2), 119-124.
- Treisman, A. (1992). Perceiving and re-perceiving objects. *American Psychologist*, 47(7), 862-875.
- Treisman, A. (2006). Object tokens, binding, and visual memory. In H. D. Zimmer, M. A. & U. Lindenberger (Hrsg.), *Handbook of binding and memory: perspectives from cognitive neuroscience* (S. 315-338). New York: Oxford University Press.
- Treisman, A. & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12, 97-136.
- Tulving, E. (1983). *Elements of episodic memory*. Oxford: Oxford University Press.
- Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, 26(1), 1-12.
- Tulving, E. & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80(5), 352-373.
- Van Mondfrans, A. & Travers, R. M. (1965). Paired-associate learning within and across sense modalities and involving simultaneous and sequential presentations. *American Educational Research Journal*, 2, 89-99.
- Van Petten, C., Senkfor, A. J. & Newberg, W. M. (2000). Memory for drawings in locations: Spatial source memory and event-related potentials. *Psychophysiology*, 37(4), 551-564.
- Verde, M. F. (2004). Associative interference in recognition memory: A dual-process account. *Memory & Cognition*, 32(8), 1273-1283.
- Verhaeghen, P. & Salthouse, T. A. (1997). Meta-analyses of age-cognition relations in adulthood: Estimates of linear and non-linear age effects and structural models. *Psychological Bulletin*, 122, 231-249.
- von Hecker, U. & Meiser, T. (2005). Defocused attention in depressed mood: Evidence from source monitoring. *Emotion*, 5(4), 456-463.
- Voss, J. F., Vesonder, G. T., Post, T. A. & Ney, L. G. (1987). Was the item recalled and if so by whom? *Journal of Memory and Language*, 26(4), 466-479.
- Wason, P. C. (1968). Reasoning about a rule. *The Quarterly Journal of Experimental Psychology*, 20(3), 273-281.
- Waters, H. S. (2000). Memory strategy development: Do we need yet another deficiency? *Child Development*, 71(4), 1004-1012.

- Waugh, N. C. (1967). Presentation time and free recall. *Journal of Experimental Psychology*, *73*(1), 39-44.
- Wegesin, D. J., Friedman, D., Varughese, N. & Stern, Y. (2002). Age-related changes in course memory retrieval: An ERP replication and extension. *Cognitive Brain Research*, *13*(3), 323-338.
- Wegesin, D. J., Jacobs, D. M., Zubin, N. R., Ventura, P. R. & Stern, Y. (2000). Source memory and encoding strategy in normal aging. *Journal of Clinical and Experimental Neuropsychology*, *22*(4), 455-464.
- Westermann, R. (1987). *Strukturalistische Theorienkonzeption und empirische Forschung in der Psychologie. Eine Fallstudie*. Berlin: Springer.
- Wiley, J. G. & Kausler, D. H. (1993). Adult age differences in temporal memory for cyclic actions. *Experimental Aging Research*, *19*(4), 251-365.
- Wilkinson, A. C. & Koestler, R. (1983). Repeated recall: A new model and tests of its generality from childhood to old age. *Journal of Experimental Psychology: General*, *112*(3), 423-451.
- Xu, J. & Malmberg, K. J. (2007). Modeling the effects of verbal and nonverbal pair strength on associative recognition. *Memory & Cognition*, *35*(3), 526-544.
- Yonelinas, A. P. (1997). Recognition memory rocs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition*, *25*(6), 747-763.
- Yonelinas, A. P., Dobbins, I., Szymanski, M. D., Dhaliwal, H. S. & King, L. (1996). Signal-detection, threshold, and dual-process models of recognition memory: Rocs and conscious recollection. *Consciousness and Cognition: An International Journal*, *5*(4), 418-441.
- Yonelinas, A. P., Kroll, N. E. A., Dobbins, I. G. & Soltani, M. (1999). Recognition memory of faces: When familiarity supports associative recognition judgments. *Psychonomic Bulletin & Review*, *6*(4), 654-661.
- Young, R. K. (1961). Paired-associate learning when the same items occur as stimuli and responses. *Journal of Experimental Psychology*, *61*(4), 315-318.
- Zacks, R. T., Hasher, L. & Li, K. Z. H. (2000). Human memory. In F. I. M. Craik & T. A. Salthouse (Hrsg.), *The handbook of aging and cognition* (2. Aufl., S. 293-357). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Zimmer, H., Mecklinger, A. & Lindenberger, U. (2006). Levels of binding: types, mechanisms, and functions of binding in remembering. In H. D. Zimmer, M. A. & U. Lindenberger (Hrsg.), *Handbook of binding and memory: perspectives from cognitive neuroscience* (S. 3-22). New York: Oxford University Press.

Teil VI

Anhang

Anhang A

Verarbeitungsbaummodelle des AHD-Modells

Die auf den folgenden Seiten abgebildeten Grafiken zeigen die Verarbeitungsbaummodelle des ADH-Modells.

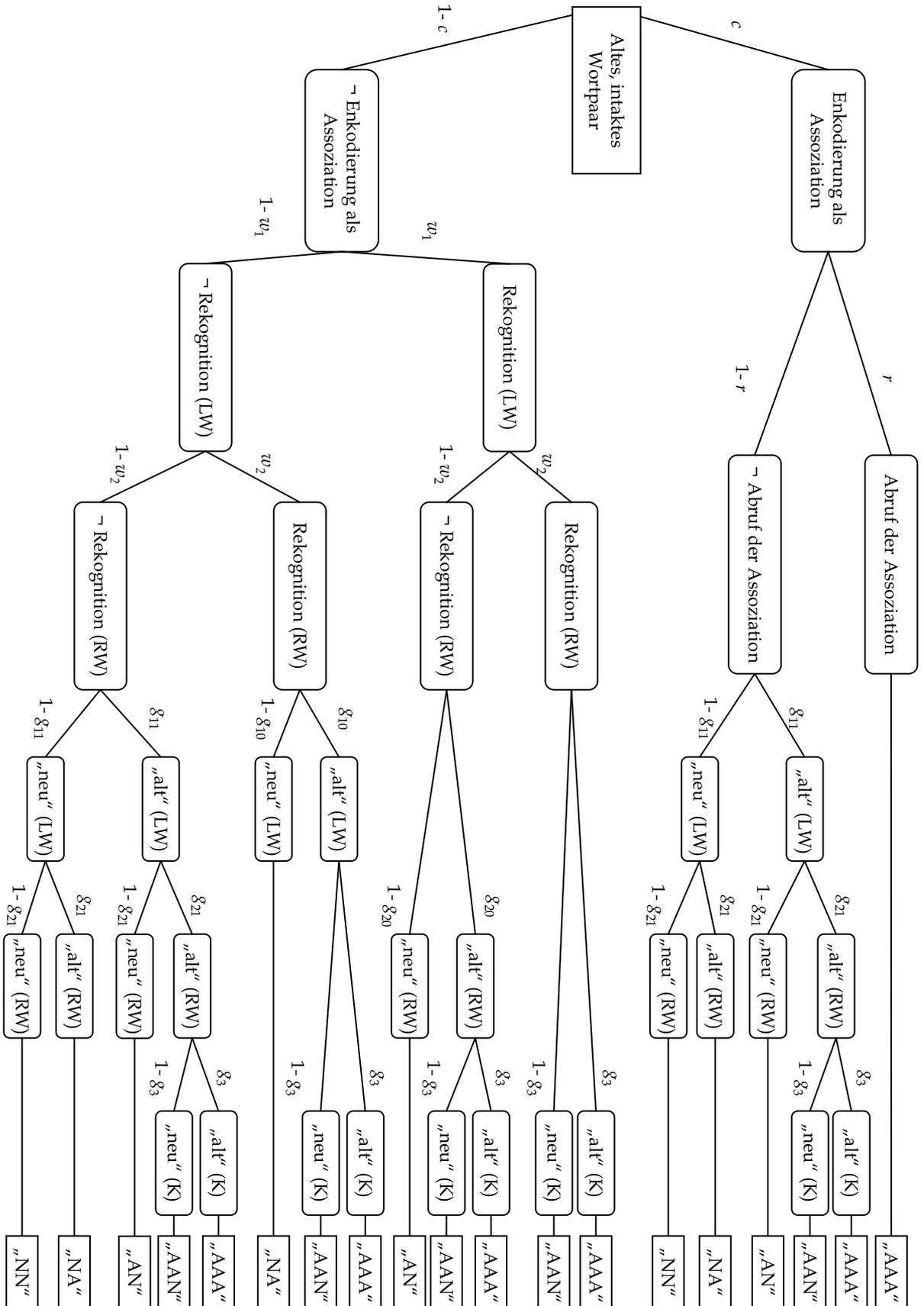


Abbildung A.1: Verarbeitungsbaummodell für alte, intakte Wortpaare

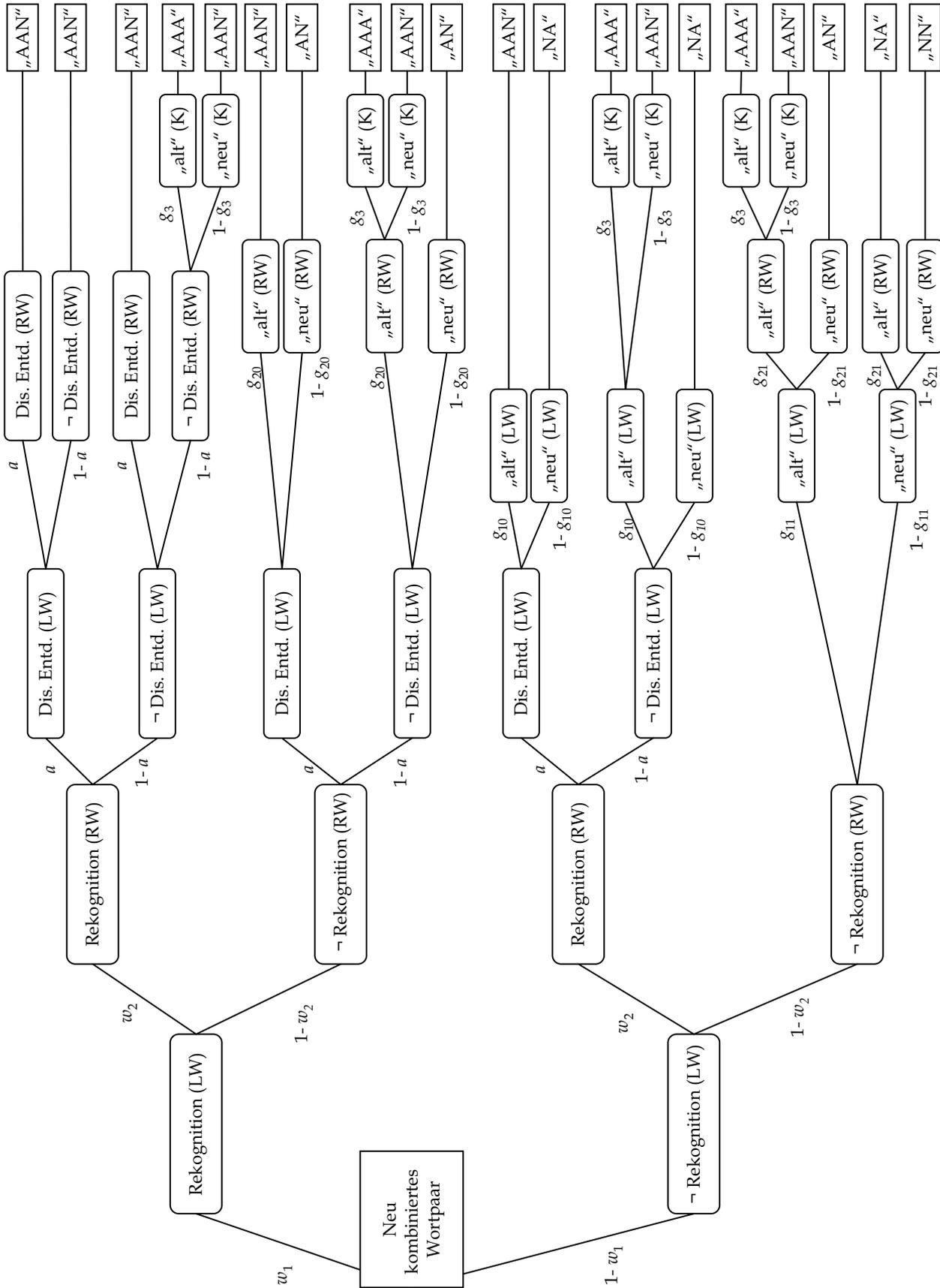


Abbildung A.2: Verarbeitungsbaummodell für neu kombinierte Wortpaare

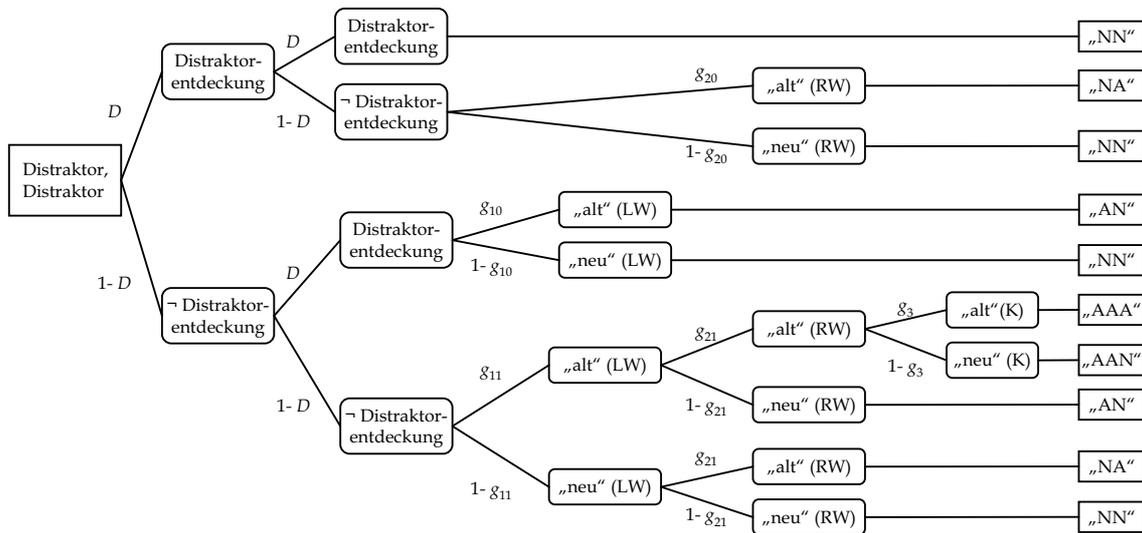


Abbildung A.5: Verarbeitungsbaummodell für zwei Distraktoren

Anhang B

eqn-Datei

Die Modellgleichungen für die fünf Bäume des ADH-Modells sind in folgender eqn-Datei enthalten.

75

```
1 1 c*r
1 1 c*(1-r)*g11*g21*g3
1 2 c*(1-r)*g11*g21*(1-g3)
1 3 c*(1-r)*g11*(1-g21)
1 4 c*(1-r)*(1-g11)*g21
1 5 c*(1-r)*(1-g11)*(1-g21)
1 1 (1-c)*w1*w2*g3
1 2 (1-c)*w1*w2*(1-g3)
1 1 (1-c)*w1*(1-w2)*g20*g3
1 2 (1-c)*w1*(1-w2)*g20*(1-g3)
1 3 (1-c)*w1*(1-w2)*(1-g20)
1 1 (1-c)*(1-w1)*w2*g10*g3
1 2 (1-c)*(1-w1)*w2*g10*(1-g3)
1 4 (1-c)*(1-w1)*w2*(1-g10)
1 1 (1-c)*(1-w1)*(1-w2)*g11*g21*g3
1 2 (1-c)*(1-w1)*(1-w2)*g11*g21*(1-g3)
1 3 (1-c)*(1-w1)*(1-w2)*g11*(1-g21)
1 4 (1-c)*(1-w1)*(1-w2)*(1-g11)*g21
1 5 (1-c)*(1-w1)*(1-w2)*(1-g11)*(1-g21)
2 7 w1*w2*a*a
2 7 w1*w2*a*(1-a)
2 7 w1*w2*(1-a)*a
2 6 w1*w2*(1-a)*(1-a)*g3
```

```

2 7 w1*w2*(1-a)*(1-a)*(1-g3)
2 7 w1*(1-w2)*a*g20
2 8 w1*(1-w2)*a*(1-g20)
2 6 w1*(1-w2)*(1-a)*g20*g3
2 7 w1*(1-w2)*(1-a)*g20*(1-g3)
2 8 w1*(1-w2)*(1-a)*(1-g20)
2 7 (1-w1)*w2*a*g10
2 9 (1-w1)*w2*a*(1-g10)
2 6 (1-w1)*w2*(1-a)*g10*g3
2 7 (1-w1)*w2*(1-a)*g10*(1-g3)
2 9 (1-w1)*w2*(1-a)*(1-g10)
2 6 (1-w1)*(1-w2)*g11*g21*g3
2 7 (1-w1)*(1-w2)*g11*g21*(1-g3)
2 8 (1-w1)*(1-w2)*g11*(1-g21)
2 9 (1-w1)*(1-w2)*(1-g11)*g21
2 10 (1-w1)*(1-w2)*(1-g11)*(1-g21)
3 13 w1*D
3 12 w1*(1-D)*a*g20
3 13 w1*(1-D)*a*(1-g20)
3 11 w1*(1-D)*(1-a)*g20*g3
3 12 w1*(1-D)*(1-a)*g20*(1-g3)
3 13 w1*(1-D)*(1-a)*(1-g20)
3 13 (1-w1)*D*g10
3 15 (1-w1)*D*(1-g10)
3 11 (1-w1)*(1-D)*g11*g21*g3
3 12 (1-w1)*(1-D)*g11*g21*(1-g3)
3 13 (1-w1)*(1-D)*g11*(1-g21)
3 14 (1-w1)*(1-D)*(1-g11)*g21
3 15 (1-w1)*(1-D)*(1-g11)*(1-g21)
4 19 w2*D
4 17 w2*(1-D)*a*g10
4 19 w2*(1-D)*a*(1-g10)
4 16 w2*(1-D)*(1-a)*g10*g3
4 17 w2*(1-D)*(1-a)*g10*(1-g3)
4 19 w2*(1-D)*(1-a)*(1-g10)
4 19 (1-w2)*D*g20
4 20 (1-w2)*D*(1-g20)
4 16 (1-w2)*(1-D)*g11*g21*g3
4 17 (1-w2)*(1-D)*g11*g21*(1-g3)

```

4 18 $(1-w2)*(1-D)*g11*(1-g21)$
4 19 $(1-w2)*(1-D)*(1-g11)*g21$
4 20 $(1-w2)*(1-D)*(1-g11)*(1-g21)$
5 25 $D*D$
5 24 $D*(1-D)*g20$
5 25 $D*(1-D)*(1-g20)$
5 23 $(1-D)*D*g10$
5 25 $(1-D)*D*(1-g10)$
5 21 $(1-D)*(1-D)*g11*g21*g3$
5 22 $(1-D)*(1-D)*g11*g21*(1-g3)$
5 23 $(1-D)*(1-D)*g11*(1-g21)$
5 24 $(1-D)*(1-D)*(1-g11)*g21$
5 25 $(1-D)*(1-D)*(1-g11)*(1-g21)$

Anhang C

Wortmaterial

Im Folgenden ist das verwendete Wortmaterial aufgelistet. Aus der Celex Datenbank (Max Planck Institute for Psycholinguistics, 1991) sind die Identifikationsnummern (IDNum), die Häufigkeit des Wortes bezogen auf einen Korpus von 1 Mio. Wörter (MannMln) und das Wort selbst (Head) aufgeführt. Ergänzt wurde die Listenzugehörigkeit zu den parallelisierten Listenversionen (Liste 1 bzw. 2).

Liste 1			Liste 2		
IDNum	MannMln	Head	IDNum	MannMln	Head
556	37	Absatz	108	133	Abend
740	47	Abstand	394	36	Ablauf
891	27	Abwehr	684	72	Absicht
1045	20	Achtung	951	34	Abzug
1429	84	Aktie	1342	22	Ahnung
1608	20	Allee	1436	84	Aktion
1828	53	Alter	1745	19	Alltag
2174	20	Anblick	2305	171	Anfang
2443	74	Angriff	2449	102	Angst
2575	20	Ankunft	2529	20	Ankauf
2790	32	Ansatz	2865	143	Ansicht
3086	84	Antrag	2982	96	Anteil
3112	162	Antwort	3689	36	Aspekt
3123	26	Anwalt	3938	113	Aufbau
3158	39	Anzahl	4746	19	Ausfuhr
3278	34	Appell	5031	33	Ausmaß
3462	23	Ärger	5590	61	Autor
3532	68	Armee	5729	38	Bahnhof

Liste 1			Liste 2		
IDNum	MannMln	Head	IDNum	MannMln	Head
4292	20	Aufruf	5963	55	Basis
4306	19	Aufsatz	6294	34	Befehl
4441	101	Auftrag	6438	131	Begriff
4463	25	Aufwand	6533	66	Beifall
4591	48	Ausbau	6865	169	Bereich
4952	108	Ausland	7212	169	Besuch
5464	34	Auswahl	7262	341	Betrieb
5480	22	Ausweg	7352	62	Beweis
6186	40	Bedarf	7417	108	Bezirk
6407	119	Beginn	7422	31	Bezug
6611	103	Beitrag	7481	39	Bilanz
6894	31	Bergbau	7546	22	Bindung
6911	192	Bericht	7591	80	Bischof
6932	87	Beruf	7696	27	Blech
7096	61	Besitz	7884	153	Boden
7145	28	Bestand	8068	31	Branche
7250	51	Betrag	8073	28	Brand
7305	19	Beute	8153	28	Breite
7498	81	Bildung	8193	142	Brief
7613	22	Bitte	8411	38	Bündnis
7665	73	Blatt	8435	36	Bursche
7734	176	Blick	8495	118	Chance
7806	32	Blume	8507	50	Chemie
7937	30	Bombe	8533	27	Christ
7980	70	Börse	9060	74	Dichter
8128	20	Braut	9487	33	Drohung
8271	89	Bruder	9501	51	Druck
8301	43	Brust	10248	27	Einfuhr
8374	57	Bühne	10254	19	Eingang
8425	117	Bürger	10784	21	Eisen
8467	19	Butter	10939	44	Empfang
8685	66	Dasein	11013	32	Engel
8702	34	Datum	11370	70	Entwurf
9091	104	Dienst	12402	34	Fabrik
9293	87	Doktor	12473	58	Fahrer
9397	27	Drama	12583	68	Farbe
9479	21	Dritte	12691	19	Feder

Liste 1			Liste 2		
IDNum	MannMln	Head	IDNum	MannMln	Head
10315	174	Einheit	12809	90	Fenster
10492	79	Einsatz	13120	60	Finger
10714	25	Einwand	13167	34	Fisch
11563	286	Erfolg	13206	31	Fläche
11803	93	Ernst	13381	40	Flucht
11811	24	Ernte	13399	28	Flügel
11863	17	Ersatz	13468	166	Folge
12000	56	Ertrag	13540	24	Formel
12337	44	Export	13830	188	Frieden
12461	28	Fahne	13861	19	Frist
12491	65	Fahrt	13891	85	Front
12507	38	Faktor	13968	104	Führung
12524	33	Falle	14034	23	Furcht
12672	37	Faust	14225	54	Garten
12717	57	Fehler	14307	301	Gebiet
12756	50	Feind	14340	25	Geburt
13055	43	Figur	14384	19	Geduld
13153	127	Firma	14390	180	Gefahr
13231	18	Flamme	14444	98	Gefühl
13242	36	Flasche	14457	40	Gegend
13286	35	Fleisch	14513	11	Gehalt
13815	208	Freund	14583	115	Geist
13912	24	Frucht	14680	24	Geltung
13971	30	Fülle	14856	57	Gerät
14052	21	Fürst	14876	108	Gericht
14368	45	Gedicht	14909	24	Gerücht
14505	94	Gegner	15011	102	Gesicht
14998	176	Gesetz	15141	127	Gewalt
15051	71	Gestalt	15172	46	Gewicht
15134	20	Gewähr	15411	95	Glück
15178	83	Gewinn	15655	191	Grenze
15274	19	Glanz	15796	441	Grund
15296	44	Glaube	16112	87	Hälfte
15367	24	Glied	16120	37	Halle
15684	21	Griff	16188	118	Handel
15724	75	Größe	16578	63	Heimat
15784	20	Grube	16669	32	Hektar

Liste 1			Liste 2		
IDNum	MannMln	Head	IDNum	MannMln	Head
15863	201	Gruppe	16683	23	Helfer
15877	36	Gruß	16954	28	Herzog
16020	46	Hafen	17041	217	Hilfe
16024	19	Hafer	17275	79	Hinweis
16161	106	Haltung	17683	48	Hörer
17066	81	Himmel	17735	118	Hotel
17870	28	Hütte	18377	67	Inhalt
18115	25	Import	19110	65	Junge
18138	21	Impuls	19339	44	Kammer
18549	71	Insel	19795	37	Keller
18598	23	Instanz	19976	213	Kirche
18886	37	Irrtum	20008	22	Kiste
19096	143	Jugend	20039	28	Klage
19203	53	Kaffee	20259	28	Klima
19229	49	Kaiser	20283	17	Klinke
19351	342	Kampf	20315	36	Kloster
19565	48	Karte	20725	24	Komplex
19615	52	Kasse	20837	108	König
19693	26	Käufer	20881	21	Können
19867	23	Kette	21074	23	Konzern
20104	118	Klasse	21294	350	Kraft
20169	41	Kleid	21526	318	Krieg
20278	19	Klinik	21673	69	Küche
20507	21	Koffer	21749	77	Kultur
20518	22	Kohle	21773	21	Kumpel
20954	68	Kontakt	21783	5	Kunde
21068	19	Konzept	21896	24	Kurve
21076	37	Konzert	22021	73	Lager
21176	73	Körper	22160	20	Landung
21249	99	Kosten	22181	33	Länge
21364	26	Kranke	22656	25	Leiden
21447	6	Kredit	22809	60	Leser
21468	276	Kreis	22869	183	Leute
21497	40	Kreuz	22914	104	Licht
21580	63	Krise	23148	36	Lippe
21598	85	Kritik	23166	28	Liste
21621	27	Krone	23175	32	Liter

Liste 1			Liste 2		
IDNum	MannMln	Head	IDNum	MannMln	Head
21698	23	Kugel	23610	19	Lyrik
21804	111	Kunst	23627	175	Macht
21840	28	Kupfer	23826	42	Mangel
21909	27	Kürze	23886	26	Mantel
21951	37	Küste	24293	246	Meinung
22557	59	Lehre	24345	47	Meldung
22603	25	Leiche	24392	86	Menge
22738	65	Leiter	24529	57	Metall
22963	114	Liebe	24645	33	Miete
23081	18	Linde	24677	43	Milch
23101	111	Linie	25059	202	Mittel
23384	171	Lösung	25174	57	Modell
23645	227	Mädchen	25282	25	Mönch
23780	33	Maler	25682	120	Musik
23949	104	Markt	25730	22	Muster
24033	69	Masse	26407	24	Nebel
24314	92	Meister	26614	37	Neubau
24476	23	Merkmal	26922	33	Niveau
24496	36	Messe	27179	46	Nummer
24573	211	Meter	27214	38	Nutzen
25052	134	Mitte	27580	112	Opfer
25146	23	Möbel	27941	75	Papier
25261	39	Moment	28069	451	Partei
25274	330	Monat	28107	57	Partner
25395	21	Moral	28243	43	Pause
25415	26	Mörder	28419	151	Person
25436	77	Morgen	28505	30	Pfarrer
25507	45	Motiv	28541	52	Pfennig
25511	94	Motor	28546	65	Pferd
25757	101	Mutter	28646	50	Pfund
25836	44	Nachbar	28802	48	Physik
26004	161	Nacht	28871	27	Pilot
26282	111	Nation	28919	18	Plakat
26330	119	Natur	28942	63	Planung
26504	22	Neigung	29305	26	Posten
27315	43	Objekt	29539	93	Presse
27553	24	Onkel	29596	65	Prinz

Liste 1			Liste 2		
IDNum	MannMln	Head	IDNum	MannMln	Head
27655	125	Ordnung	29600	98	Prinzip
27659	55	Organ	29658	20	Probe
28610	56	Pflicht	29867	48	Protest
28694	38	Phase	29898	30	Provinz
28981	260	Platz	29932	51	Prüfung
29014	22	Plenum	30054	268	Punkt
29475	71	Praxis	30197	26	Quartal
29493	266	Preis	30883	261	Recht
29677	322	Problem	30977	27	Redner
29688	47	Produkt	31040	85	Regel
29760	49	Projekt	31050	48	Regen
30231	47	Quelle	31073	23	Regie
30450	118	Rahmen	31085	30	Regime
30631	30	Rathaus	31153	71	Reich
31012	46	Reform	31255	46	Reise
31180	154	Reihe	31514	36	Rente
31319	24	Reiter	32026	20	Roggen
31374	32	Rekord	32097	59	Roman
31490	40	Rennen	32515	29	Runde
31690	27	Rettung	32635	262	Sache
31818	56	Richter	33131	44	Schaden
32076	170	Rolle	33448	19	Scheibe
32714	31	Saison	33584	52	Schicht
33471	22	Schein	33666	97	Schiff
33832	25	Schlaf	34208	36	Schmerz
33866	51	Schlag	34654	179	Schritt
34360	41	Schnee	34726	55	Schuld
34580	22	Schrank	35075	25	Schwein
34637	38	Schrift	35342	71	Seele
34707	33	Schuh	35779	60	Sendung
34755	184	Schule	35813	20	Senkung
34869	67	Schutz	36058	40	Sicht
35600	35	Sektor	36133	49	Sieger
35761	60	Senat	36158	21	Signal
36181	26	Silber	36369	96	Sitzung
36771	109	Sorge	36560	117	Soldat
37212	38	Spiegel	36687	91	Sonne

Liste 1			Liste 2		
IDNum	MannMln	Head	IDNum	MannMln	Head
37505	147	Sprache	36788	22	Sorte
37661	32	Sprung	36996	31	Spaß
37734	498	Staat	37229	204	Spiel
37835	411	Stadt	37386	45	Spitze
37900	35	Stahl	37445	78	Sport
37996	79	Stand	37936	21	Stall
38317	60	Stein	37946	35	Stamm
38370	267	Stelle	38085	43	Stärke
38470	71	Stern	38108	51	Start
38500	35	Steuer	38131	34	Station
38600	20	Stille	38167	19	Status
38626	213	Stimme	38177	22	Staub
38663	27	Stirn	38675	21	Stock
38693	51	Stoff	38844	25	Strand
38716	23	Stolz	38865	269	Straße
38758	24	Störung	38942	46	Streik
38791	33	Strafe	39089	121	Student
39016	55	Strom	39112	52	Stufe
39078	142	Stück	39117	38	Stuhl
39146	318	Stunde	39326	41	Summe
39167	38	Sturm	39447	20	Symbol
39179	26	Sturz	39466	22	Symptom
39249	25	Suche	39701	86	Tagung
39532	129	System	40309	38	Tempo
39549	41	Szene	40350	40	Termin
39831	33	Tante	40484	135	Thema
39928	36	Tasche	40591	35	Tiefe
39979	29	Täter	40922	75	Tonne
40134	136	Technik	41152	31	Träger
40318	187	Tendenz	41397	38	Traum
40342	24	Teppich	41441	20	Treffer
40384	18	Terror	41700	59	Trommel
40431	22	Teufel	41841	72	Truppe
40538	36	These	43154	83	Umstand
40711	101	Tisch	43228	35	Umwelt
40743	85	Titel	43628	52	Unfall
40772	97	Tochter	45808	63	Verkehr

Liste 1			Liste 2		
IDNum	MannMln	Head	IDNum	MannMln	Head
41191	42	Trainer	45918	100	Verlag
41226	32	Träne	46543	162	Versuch
41496	28	Treppe	46951	20	Villa
42864	63	Umfang	47340	76	Vorgang
42893	19	Umgang	47401	81	Vorjahr
43066	153	Umsatz	47491	20	Vorrat
43856	30	Unglück	47610	40	Vortrag
43880	26	Unheil	47799	81	Waffe
44090	35	Unrecht	47850	32	Wähler
44915	46	Urlaub	47924	21	Währung
44955	97	Urteil	48186	85	Wasser
45101	236	Vater	48281	20	Weber
45256	82	Verband	48531	30	Weile
45331	31	Verbot	48849	23	Werbung
45485	58	Verein	48923	110	Wesen
45799	63	Verkauf	48944	31	Weste
45937	66	Verlauf	48978	66	Wetter
45996	62	Verlust	49230	21	Wiese
46593	152	Vertrag	49513	83	Wissen
46925	22	Viertel	49551	378	Woche
47016	39	Vogel	49681	152	Wohnung
47257	39	Vorbild	49709	23	Wolke
47558	24	Vorsitz	49720	24	Wolle
47605	60	Vorteil	49829	46	Wunder
47671	58	Vorwurf	49859	168	Wunsch
47688	20	Vorzug	49872	23	Würde
47820	144	Wagen	50006	25	Zahlung
48095	20	Wärme	50190	19	Zeile
48133	19	Warnung	50239	178	Zeitung
48293	30	Wechsel	50269	26	Zelle
48686	28	Weizen	50514	27	Zeuge
48702	31	Welle	50520	28	Zeugnis
49279	78	Wille	50616	133	Zimmer
49438	85	Wirkung	50712	25	Zitat
49530	19	Witwe	50931	26	Zugang
50130	24	Zebra	51030	195	Zukunft
50170	79	Zeichen	51403	125	Zustand

Liste 1			Liste 2		
IDNum	MannMln	Head	IDNum	MannMln	Head
50342	51	Zentrum	51501	23	Zwang
50904	33	Zufall	51513	56	Zweck
51535	73	Zweifel	51542	20	Zweig

Anhang D

Wortmaterial Experiment 2

Als Wortmaterial wurden für Experiment 2 Wörter zufällig zu Wortpaaren zusammengestellt. Für diese Wortpaare wurden dann Sätze gebildet, die beide Wörter enthalten. Im Folgenden sind jeweils die Wörter, sowie die integrativen Sätze aufgelistet.

linkes Wort	rechtes Wort	Verbindender Satz
ABEND	PRESSE	Am ABEND hat die PRESSE bereits alles erfahren.
ABSICHT	FRIEDEN	Ihre ABSICHT ist es, FRIEDEN zu stiften.
ANFANG	KRAFT	Am ANFANG dachte sie, ihre KRAFT würde ausreichen.
ANGST	STRAND	Voll ANGST verlassen sie den überfluteten STRAND.
BEREICH	STUDENT	Für den gesamten BEREICH ist der STUDENT verantwortlich.
BLECH	STAUB	Das BLECH ist dick mit STAUB bedeckt.
BODEN	BAHNHOF	Der steinerne BODEN im BAHNHOF ist sehr alt.
BRIEF	ABSATZ	Der kurze BRIEF enthält nur einen ABSATZ.
BURSCHE	MIETE	Der junge BURSCHE bezahlt die MIETE selten pünktlich.
CHANCE	PROTEST	Die CHANCE auf einen wirksamen PROTEST ist gering.
DRAMA	FORMEL	Um ein DRAMA zu schreiben, kann man sich an keine FORMEL halten.
ENGEL	WISSEN	Über ENGEL gibt es wenig WISSEN.
FINGER	ABZUG	Der FINGER liegt gefährlich nah am ABZUG.
FLUCHT	REGEL	Eine überstürzte FLUCHT endet in der REGEL katastrophal.
GARTEN	DICHTER	Der wunderschöne GARTEN inspiriert den DICHTER.
GEBURT	FABRIK	Bald nach der GEBURT geht sie wieder in der FABRIK arbeiten.
GEFÜHL	PROVINZ	Sie hat das GEFÜHL, mitten in der PROVINZ zu sein.

linkes Wort	rechtes Wort	Verbindender Satz
GEIST	KULTUR	Sein GEIST beeinflusst noch immer die KULTUR.
GELTUNG	ANSICHT	Er verschafft sich GELTUNG mit seiner provokanten ANSICHT.
GERICHT	GEDULD	Das GERICHT vertagt sich und bittet um GEDULD.
GESICHT	BEIFALL	Ihr GESICHT strahlt bei dem tosenden BEIFALL.
GEWICHT	ABLAUF	Zusätzliches GEWICHT behindert den reibungslosen ABLAUF.
GRENZE	MÖNCH	An der GRENZE zögert der fliehende MÖNCH kurz.
HALLE	SCHADEN	In der überfluteten HALLE entstand ein großer SCHADEN.
HILFE	KAMMER	Sie ruft nach HILFE, da es in der KAMMER plötzlich dunkel wird.
HÖRER	VILLA	Der treue HÖRER hat eine VILLA gewonnen.
HOTEL	WIRKUNG	Das gemütliche HOTEL hat eine beruhigende WIRKUNG.
KIRCHE	BRAND	Die KIRCHE ist nach dem BRAND vollständig zerstört.
KLINKE	BRUDER	Die KLINKE hat der BRUDER wütend abgebrochen.
KOMPLEX	WOHNUNG	Im ganzen KOMPLEX ist nur eine WOHNUNG frei.
KÖNNEN	ALLTAG	Sein KÖNNEN muss er im ALLTAG beweisen.
KONZERN	WÄHLER	Der feindliche KONZERN versucht, die WÄHLER zu beeinflussen.
KUMPEL	BEZUG	Zu seinem alten KUMPEL hat er keinen BEZUG mehr.
LAGER	ZEILE	Im LAGER angekommen, beginnt er die erste ZEILE zu schreiben.
LANDUNG	VORGANG	Die begonnene LANDUNG erweist sich als ein komplizierter VORGANG.
LÄNGE	REGIE	Die endgültige LÄNGE bestimmt die REGIE.
LEIDEN	INHALT	Das LEIDEN sieht man ihm an, als er den grauenvollen INHALT liest.
LESER	FENSTER	Der vertiefte LESER sitzt am offenen FENSTER.
LEUTE	NUTZEN	Fähige LEUTE sind überall von großem NUTZEN.
LICHT	BETRIEB	Das LICHT ist außer BETRIEB.
LISTE	PARTNER	In der LISTE stehen alle geschäftlichen PARTNER.
MELDUNG	BEZIRK	Die wichtige MELDUNG geht eilig an den ganzen BEZIRK.
METALL	FLÄCHE	Mit geschmolzenem METALL soll die marode FLÄCHE überzogen werden.
MILCH	FARBE	Frische MILCH hat eine weiße FARBE.
PAUSE	BREITE	Er genießt die PAUSE in voller BREITE.
PFARRER	VERSUCH	Der PFARRER gibt nicht auf, auch wenn der VERSUCH aussichtslos erscheint.

linkes Wort	rechtes Wort	Verbindender Satz
PFERD	QUARTAL	Das PFERD kann erst nach dem QUARTAL verkauft werden.
PLAKAT	LIPPE	Auf dem PLAKAT ist eine große rote LIPPE abgebildet.
POSTEN	LEICHE	Als er seinen POSTEN bezieht, entdeckt er die LEICHE.
PRINZ	SIGNAL	Der PRINZ gibt schließlich das entscheidende SIGNAL.
PRINZIP	WERBUNG	Das wirksamste PRINZIP ist die großflächige WERBUNG.
PRÜFUNG	WIESE	Nach der PRÜFUNG liegt er faul auf der WIESE.
RECHT	SORTE	Er hat RECHT, die edle SORTER zu wählen.
REICH	GEWALT	Im ganzen REICH wütet GEWALT.
RENTE	ZUKUNFT	Sie träumt von ihrer RENTE und einer gesicherten ZUKUNFT.
ROGGEN	AUSMASS	Der ROGGEN wächst in ungewohntem AUSMASS.
RUNDE	FLÜGEL	Sie spielt lebhaft eine RUNDE auf dem FLÜGEL.
SACHE	GRUND	Der SACHE muss man auf den GRUND gehen.
SCHEIBE	CHEMIE	Die dicke SCHEIBE schützt vor spritzender CHEMIE.
SCHICHT	BILANZ	Nach der letzten SCHICHT wird BILANZ gezogen.
SCHIFF	BEWEIS	Auf dem SCHIFF wird der entscheidende BEWEIS gesucht.
SCHMERZ	KELLER	Gepeinigt vor SCHMERZ stürmt sie aus dem KELLER.
SCHRITT	MUTTER	Mit energischem SCHRITT kommt die MUTTER herein.
SCHWEIN	KLOSTER	Das geschlachtete SCHWEIN stammt aus dem nahen KLOSTER.
SENDUNG	AUTOR	In der literarischen SENDUNG spricht heute der AUTOR persönlich.
SITZUNG	ZEUGNIS	In der SITZUNG wird ihr ZEUGNIS besprochen.
STALL	SICHT	Der STALL versperrt die weite SICHT.
STATION	GEFAHR	Die strategisch wichtige STATION befindet sich in größter GEFAHR.
STRASSE	KLIMA	Auf der überfüllten STRASSE ist das schwüle KLIMA unerträglich.
STUFE	ZUSTAND	Nach der letzten STUFE ist sie in einem erschöpften ZUSTAND.
SYMBOL	WÄHRUNG	Ein bestimmtes SYMBOL kennzeichnet jede WÄHRUNG.
SYMPTOM	FURCHT	Das SYMPTOM ist bei starker FURCHT zu beobachten.
TAGUNG	MUSIK	Die TAGUNG wird von leiser MUSIK umrahmt.
TERMIN	MEINUNG	Bis zu diesem TERMIN muss eine klare MEINUNG vorliegen.
TONNE	MACHT	Die überfüllte TONNE wird mit aller MACHT verschlossen.
TROMMEL	SIEGER	Die TROMMEL kündigt den SIEGER an.
TRUPPE	NEBEL	Die TRUPPE schleicht vorsichtig durch den NEBEL.

linkes Wort	rechtes Wort	Verbindender Satz
UNFALL	DRUCK	Seit dem UNFALL lastet großer DRUCK auf ihm.
VERKEHR	BESUCH	Der starke VERKEHR hindert den BESUCH, pünktlich zu sein.
WEILE	EMPFANG	Es dauert eine WEILE, ehe der EMPFANG beginnt.
WESTE	STOCK	Die schicke WESTE passt gut zum eleganten STOCK.
WETTER	PLANUNG	Das WETTER muss bei der PLANUNG berücksichtigt werden.
WUNDER	VERLAUF	Wie durch ein WUNDER bessert sich der VERLAUF.
WUNSCH	VORTRAG	Es war ihr WUNSCH, einen VORTRAG zu halten.
ZAHLUNG	BRANCHE	Eine sofortige ZAHLUNG ist in der BRANCHE üblich.
ZIMMER	WOCHE	Sie mietet das ZIMMER für eine ganze WOCHE.
ZITAT	PARTEI	Dieses ZITAT ist typisch für die PARTEI.
ZWANG	TRAUM	Der elterliche ZWANG hält ihn davon ab, seinen TRAUM zu erfüllen.
ZWECK	MANTEL	Für diesen feierlichen ZWECK wählt sie den MANTEL.

Anhang E

Wortmaterial Experiment 5

Dieser Anhang enthält drei Tabellen. In der ersten Tabelle sind die Wortpaare für die erste und zweite Lernliste aufgelistet. Die zweite Tabelle enthält die beiden Testversionen A und B für die erste Lernliste, die dritte Tabelle gibt die beiden alternativen Testlisten für die zweite Lernliste wieder.

E.1 Wortpaare der ersten und zweiten Lernliste

Liste 1		Liste 2	
linkes Wort	rechtes Wort	linkes Wort	rechtes Wort
Abstand	Nation	Ärger	Ansatz
Alter	Zebra	Aufsatz	Sektor
Antrag	Zelle	Ausbau	Himmel
Antwort	Strom	Bergbau	Blatt
Pilot	Schuh	Besitz	Kugel
Blick	Stunde	Beute	Treppe
Bürger	Kupfer	Erfolg	Kürze
Butter	Umfang	Falle	Urteil
Dasein	Kohle	Flügel	Sprache
Doktor	Objekt	Fülle	Kreuz
Drama	Vertrag	Gestalt	Schein
Fenster	Hektar	Gruppe	Teppich
Gewähr	Verlauf	Haltung	Plenum
Glaube	Markt	Hütte	Meinung
Halle	Dritte	Instanz	Anblick
Kette	Stille	Kaffee	Trainer
Klasse	Maler	Klinik	Sprung

Liste 1		Liste 2	
linkes Wort	rechtes Wort	linkes Wort	rechtes Wort
Kleid	Fehler	Kontakt	Weizen
Klima	Praxis	Konzept	Kreis
Klinke	Gedicht	Leiche	Frucht
Körper	Oberst	Liebe	Keller
Krone	Börse	Linie	Auftrag
Kunst	Pflicht	Lösung	Sturm
Kurve	Ausland	Masse	Umgang
Länge	Fürst	Merkmal	Ankunft
Licht	Gesetz	Moral	Verband
Lippe	Störung	Motor	Mitte
Messe	Vogel	Provinz	Datum
Moment	Meister	Rathaus	Rettung
Nacht	Insel	Rekord	Gegend
Organ	Welle	Schule	Wärme
Phase	Braut	Schwung	Verlust
Prinzip	Kranke	Silber	Schutz
Projekt	Küste	Stahl	Feind
Quelle	Impuls	Stoff	Brust
Rolle	Linde	Stufe	Träne
Runde	Irrtum	System	Allee
Schrank	Reiter	Teufel	Leiter
Staub	Verein	Tiefe	Käufer
Stimme	Ausweg	Unheil	Suche
Stirn	Achtung	Unrecht	Tisch
Stück	Anwalt	Verkauf	Stern
Szene	These	Vorzug	Kritik
Technik	Staat	Weile	Figur
Träger	Sorge	Weste	Reihe
Traktor	Platz	Wille	Nachbar
Verlag	Bitte	Zeichen	Ernte
Warnung	Spitze	Zentrum	Reform
Woche	Geruch	Zitat	Ersatz
Wolke	Blume	Zweifel	Dienst

E.2 Testversionen A und B für die erste Lernliste

Liste 1, Testversion A		Liste 1, Testversion B	
linkes Wort	rechtes Wort	linkes Wort	rechtes Wort
Antwort	Strom	Antrag	Zelle
Blick	Stunde	Bürger	Kupfer
Bürger	Kupfer	Butter	Umfang
Butter	Umfang	Doktor	Objekt
Kleid	Fehler	Gewähr	Verlauf
Krone	Börse	Glaube	Markt
Kurve	Ausland	Klasse	Maler
Länge	Fürst	Klinke	Gedicht
Lippe	Störung	Körper	Oberst
Messe	Vogel	Kunst	Pflicht
Moment	Meister	Kurve	Ausland
Nacht	Insel	Projekt	Küste
Organ	Welle	Rolle	Linde
Projekt	Küste	Schrank	Reiter
Rolle	Linde	Staub	Verein
Stimme	Ausweg	Stimme	Ausweg
Szene	These	Stirn	Achtung
Traktor	Platz	Szene	These
Verlag	Bitte	Technik	Staat
Wolke	Blume	Warnung	Spitze
Gewähr	Verein	Abstand	Irrtum
Alter	Impuls	Antwort	Anwalt
Dasein	Braut	Dasein	Börse
Drama	Markt	Drama	Vogel
Halle	Oberst	Verlag	Braut
Klasse	Kranke	Kleid	Stille
Klinke	Achtung	Klima	Fürst
Körper	Irrtum	Prinzip	Kohle
Kunst	Staat	Licht	Fehler
Licht	Verlauf	Lippe	Meister
Phase	Gedicht	Halle	Störung
Prinzip	Reiter	Nacht	Dritte
Antrag	Nation	Phase	Hektar
Staub	Sorge	Quelle	Stunde
Stirn	Gesetz	Runde	Geruch

Liste 1, Testversion A		Liste 1, Testversion B	
linkes Wort	rechtes Wort	linkes Wort	rechtes Wort
Stück	Vertrag	Kette	Gesetz
Pilot	Kohle	Pilot	Platz
Träger	Anwalt	Traktor	Nation
Klima	Hektar	Woche	Sorge
Woche	Schuh	Wolke	Vertrag
Abstand	Fleisch	Alter	Fleisch
Technik	Schlaf	Länge	Vorsitz
Doktor	Bereich	Blick	Problem
Quelle	Schlag	Moment	Freund
Glaube	Status	Stück	Bereich
Fenster	Problem	Krone	Morgen
Kette	Einwand	Messe	Wechsel
Warnung	Möbel	Organ	Möbel
Runde	Bedarf	Fenster	Titel
Schrank	Tendenz	Träger	Status
Fahne	Stille	Beruf	Bitte
Flamme	Zebra	Grube	Impuls
Gruß	Geruch	Gruß	Insel
Mädchen	Spitze	Natur	Praxis
Regel	Dritte	Regel	Zebra
Senat	Objekt	Richter	Blume
Spiegel	Pflicht	Spiegel	Kranke
Viertel	Praxis	Unglück	Schuh
Vorteil	Zelle	Vorteil	Welle
Vorwurf	Maler	Vorwurf	Strom
Beruf	Freund	Flamme	Schlag
Blech	Stolz	Blech	Stolz
Grube	Wechsel	Fahne	Schlaf
Koffer	Firma	Koffer	Firma
Krise	Glanz	Krise	Glanz
Natur	Titel	Mädchen	Tendenz
Richter	Morgen	Senat	Einwand
Sturz	Bildung	Sturz	Bildung
Unglück	Vorsitz	Viertel	Bedarf
Zufall	Boden	Zufall	Boden

E.3 Testversionen A und B für die zweite Lernliste

Liste 2, Testversion A		Liste 2, Testversion B	
linkes Wort	rechtes Wort	linkes Wort	rechtes Wort
Ärger	Ansatz	Ausbau	Himmel
Ausbau	Himmel	Bergbau	Blatt
Beute	Treppe	Falle	Urteil
Erfolg	Kürze	Gruppe	Teppich
Hütte	Meinung	Hütte	Meinung
Klinik	Sprung	Kaffee	Trainer
Liebe	Keller	Leiche	Frucht
Linie	Auftrag	Liebe	Keller
Lösung	Sturm	Linie	Auftrag
Masse	Umgang	Masse	Umgang
Motor	Mitte	Merkmal	Ankunft
Rathaus	Rettung	Motor	Mitte
Rekord	Gegend	Rekord	Gegend
Schule	Wärme	Schwung	Verlust
Stoff	Brust	Stufe	Träne
Teufel	Leiter	Unheil	Suche
Unheil	Suche	Unrecht	Tisch
Vorzug	Kritik	Weile	Figur
Zeichen	Ernte	Weste	Reihe
Zweifel	Dienst	Wille	Nachbar
Aufsatz	Teppich	Silber	Sprung
Bergbau	Figur	Aufsatz	Kreis
Weile	Verband	Beute	Brust
Fülle	Reihe	Flügel	Feind
Gestalt	Träne	Fülle	Rettung
Gruppe	Frucht	Gestalt	Ernte
Konzept	Kreuz	Moral	Leiter
Instanz	Ersatz	Instanz	Sprache
Kontakt	Anblick	Klinik	Anblick
Leiche	Weizen	Besitz	Treppe
Verkauf	Allee	Konzept	Reform
Moral	Kugel	Stahl	Ansatz
Provinz	Reform	Stoff	Datum
Schwung	Käufer	Teufel	Kürze
Tiefe	Blatt	Zitat	Schein

Liste 2, Testversion A		Liste 2, Testversion B	
linkes Wort	rechtes Wort	linkes Wort	rechtes Wort
System	Ankunft	Verkauf	Dienst
Flügel	Feind	Vorzug	Ersatz
Wille	Trainer	Zeichen	Allee
Zentrum	Schein	Provinz	Stern
Zitat	Urteil	Zweifel	Sektor
Besitz	Neigung	Ärger	Flasche
Falle	Flasche	Erfolg	Schrift
Kaffee	Konzert	Lösung	Vorbild
Weste	Wagen	Zentrum	Gegner
Silber	Aufwand	Tiefe	Anzahl
Stufe	Karte	Rathaus	Rennen
Merkmal	Rennen	Schule	Regime
Unrecht	Kaiser	Kontakt	Meter
Haltung	Schrift	System	Aufwand
Stahl	Vorbild	Haltung	Wagen
Bericht	Nachbar	Bericht	Kugel
Bestand	Sprache	Bestand	Schutz
Ernst	Schutz	Ernst	Wärme
Hafer	Plenum	Hafer	Plenum
Jugend	Datum	Liter	Kritik
Rahmen	Verlust	Produkt	Sturm
Stock	Kreis	Rahmen	Weizen
Strafe	Stern	Saison	Kreuz
Tasche	Sektor	Tante	Käufer
Urlaub	Tisch	Tasche	Verband
Auswahl	Lehre	Auswahl	Lehre
Faust	Fahrt	Faust	Fahrt
Inhalt	Schnee	Inhalt	Schnee
Liter	Regime	Jugend	Neigung
Produkt	Meter	Sicht	Aufruf
Saison	Anzahl	Stein	Entwurf
Sicht	Aufruf	Stock	Karte
Stein	Entwurf	Strafe	Konzert
Tante	Gegner	Urlaub	Kaiser
Witwe	Griff	Witwe	Griff

Anhang F

mdt-Dateien

Dieser Anhang enthält die mdt-Dateien mit den absoluten Häufigkeiten für die Experimente 1 bis 5.

F.1 Experiment 1a

Exp1a; Ja-nein-Bedingung dann 5AFC-Bedingung; Liste 1+2

1	603
2	78
3	55
4	61
5	43
6	44
7	212
8	72
9	69
10	21
11	17
12	76
13	221
14	30
15	76
16	19
17	75
18	38
19	225

20	63
21	20
22	74
23	189
24	184
25	373
26	597
27	79
28	48
29	39
30	37
31	97
32	409
33	128
34	115
35	51
36	18
37	97
38	209
39	18
40	58
41	18
42	102
43	35
44	173
45	72
46	13
47	50
48	75
49	64
50	198

=====

F.2 Experiment 1b

Exp.1b; Darbietungszeit gleich lang, kurz lang, lang kurz; Liste 1+2

1	860
2	124

3	83
4	76
5	57
6	65
7	289
8	103
9	97
10	30
11	19
12	113
13	333
14	38
15	97
16	22
17	117
18	51
19	325
20	85
21	29
22	120
23	273
24	253
25	525
26	425
27	79
28	32
29	58
30	36
31	19
32	58
33	19
34	24
35	10
36	14
37	68
38	149
39	22
40	46
41	20

42	63
43	20
44	169
45	27
46	360
47	81
48	61
49	36
50	32
51	24
52	64
53	35
54	21
55	4
56	18
57	56
58	180
59	20
60	27
61	20
62	61
63	27
64	142
65	51

=====

F.3 Experiment 2

Exp.2; Satz, 3min RI; Vokal, 3min RI; Satz, 24h RI; Vokal, 24h RI;

1	304
2	72
3	36
4	31
5	17
6	26
7	245
8	74
9	93
10	22

11	9
12	49
13	126
14	24
15	22
16	8
17	48
18	20
19	120
20	34
21	10
22	27
23	55
24	49
25	89
26	133
27	114
28	51
29	67
30	45
31	99
32	124
33	70
34	75
35	42
36	19
37	42
38	79
39	16
40	49
41	20
42	41
43	38
44	76
45	30
46	21
47	29
48	48
49	55

50	52
51	144
52	96
53	67
54	85
55	38
56	32
57	167
58	99
59	87
60	45
61	11
62	51
63	73
64	35
65	45
66	14
67	46
68	22
69	84
70	49
71	5
72	42
73	39
74	58
75	71
76	80
77	122
78	67
79	82
80	49
81	79
82	114
83	63
84	88
85	56
86	35
87	43
88	59

89	36
90	27
91	31
92	49
93	31
94	53
95	36
96	22
97	37
98	51
99	55
100	35

=====

F.4 Experiment 3

Exp.3; 70% alte Wörter; 30% alte Wörter; Liste 1+2

1	812
2	175
3	70
4	57
5	30
6	139
7	707
8	122
9	138
10	38
11	34
12	216
13	286
14	45
15	43
16	54
17	220
18	54
19	245
20	51
21	38
22	132

23	156
24	148
25	150
26	255
27	36
28	47
29	42
30	68
31	16
32	146
33	117
34	95
35	74
36	9
37	27
38	239
39	23
40	150
41	10
42	33
43	25
44	240
45	140
46	23
47	65
48	316
49	292
50	1992

=====

F.5 Experiment 4

Exp.4; Gruppe1 dann Gruppe2; Liste 1+2

1	422
2	55
3	32
4	27
5	16
6	36

7	303
8	108
9	75
10	30
11	18
12	100
13	467
14	30
15	121
16	27
17	135
18	56
19	398
20	120
21	13
22	71
23	234
24	129
25	657
26	362
27	45
28	27
29	50
30	20
31	38
32	290
33	73
34	89
35	14
36	17
37	151
38	350
39	58
40	96
41	11
42	115
43	23
44	437
45	86

46	4
47	64
48	139
49	288
50	513
=====	

F.6 Experiment 5

Exp.5; Studenten dann Senioren; Liste 1+2

1	881
2	210
3	110
4	131
5	108
6	159
7	671
8	257
9	232
10	121
11	41
12	159
13	311
14	81
15	128
16	52
17	134
18	67
19	338
20	129
21	19
22	90
23	141
24	142
25	328
26	595
27	237
28	138
29	183

30	127
31	295
32	413
33	234
34	217
35	121
36	84
37	144
38	201
39	93
40	118
41	90
42	132
43	64
44	246
45	108
46	40
47	89
48	131
49	151
50	229

=====

Exp.5; Studenten dann Senioren; Liste 1

1	434
2	111
3	50
4	73
5	52
6	94
7	322
8	129
9	120
10	55
11	21
12	75
13	152
14	48
15	64
16	23

17	54
18	31
19	184
20	68
21	7
22	42
23	60
24	86
25	165
26	271
27	109
28	76
29	105
30	79
31	139
32	177
33	130
34	120
35	74
36	39
37	54
38	101
39	52
40	74
41	38
42	57
43	30
44	133
45	62
46	16
47	26
48	59
49	82
50	137

=====
=====

Exp.5; Studenten dann Senioren; Liste 2

1	447
2	99
3	60

4	58
5	56
6	65
7	349
8	128
9	112
10	66
11	20
12	84
13	159
14	33
15	64
16	29
17	80
18	36
19	154
20	61
21	12
22	48
23	81
24	56
25	163
26	324
27	128
28	62
29	78
30	48
31	156
32	236
33	104
34	97
35	47
36	45
37	90
38	100
39	41
40	44
41	52
42	75

43 34

44 113

45 46

46 24

47 63

48 72

49 69

50 92

====

Ehrenwörtliche Erklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig verfasst sowie sämtliche Belege deutlich gemacht und korrekt angegeben habe.

Mannheim, den 21.04.2009

Tina Sarah Auer

Danksagung

Zum Gelingen dieser Arbeit haben viele Personen beigetragen, denen ich an dieser Stelle meinen Dank aussprechen möchte.

Zunächst möchte ich meinem Betreuer Professor Edgar Erdfelder dafür danken, dass er bereits während meines Studiums durch seine engagierte Lehre und ansteckende Forschungsbegeisterung mein Interesse an der kognitiven Psychologie verstärkt hat und mir nach meinem Studium die Promotion ermöglichte. Seine inhaltlichen Anregungen, methodischen Tipps und sein konstruktives Feedback haben mir über manchen kritischen Zeitpunkt der Promotion hinweggeholfen und ganz entscheidend zu dieser Arbeit beigetragen.

Bei Professor Rüdiger Pohl möchte ich mich dafür bedanken, dass er das Zweitgutachten für meine Dissertation übernommen hat. Auch bei ihm fand ich während der Promotionszeit immer ein offenes Ohr und er stand mir bei Problemen und Fragen jederzeit hilfsbereit zur Verfügung.

Meinen Kollegen Meike Kroneisen, Benjamin Hilbig und Morten Moshagen möchte ich dafür danken, dass sie Teile der Arbeit gelesen haben und mir wichtige Hinweise, Anregungen und konstruktive Rückmeldungen gegeben haben. Insbesondere Morten Moshagen hat es mir durch seine schnellen Rückmeldungen und seine große Lesebereitschaft ermöglicht, die Arbeit „rechtzeitig“ abzugeben.

Bei Martin Brandt und Lutz Cüpper möchte ich mich für die vielen Diskussionen, nicht nur im Rahmen unseres Kolloquiums, bedanken. Auch den Weg zum Mittagessen habe ich oft genug genutzt, um Euch nach Eurer Meinung und Erfahrung zu fragen. Danke für Eure Geduld und Unterstützung! Monika Undorf möchte ich an dieser Stelle für die gute und produktive Zusammenarbeit während des SFB-Projekts sowie den (nicht nur fachlichen) Austausch danken.

Auch allen anderen Mitarbeitern des Lehrstuhls Psychologie III, Rahel Mangold, Brigitte Krieg und unseren Hiwis möchte ich für die vielfältige Unterstützung danken. Vor allem möchte ich dabei Beatrice Kuhlmann und Sylvia Sperlich hervorheben. Beatrice Kuhlmann hat neben vielem Anderem insbesondere die Datenerhebungen koordiniert und tatkräftig unterstützt. Sylvia Sperlich hat durch ihre flinken Literaturrecherchen, ihre Geduld bei der Überarbeitung meiner Grafiken und ihre APA-Style Kenntnisse einen wichtigen Beitrag in der Endphase meiner Dissertation geleistet.

Darüber hinaus möchte ich mich auch bei den von mir betreuten Studenten bedanken, die sich im Rahmen ihres Experimentalpraktikums mit meinem Modell beschäftigen durften und dabei zum Teil ein großes Engagement an den Tag legten. Herzlichen Dank an Simon Böhm, Manuela Glasbrenner, Kim Griemsmann, Lisa Haider, Simona Maltese, Maika Nellen, Janin Rössl und Anne Schiefer. Auch Eva Link möchte ich für die Datenerhebung der älteren Probanden im Rahmen ihrer Diplomarbeit danken. Sie hat bei der aufwendigen Seniorenerhebung viel Geduld und Engagement bewiesen.

Valentin Eck danke ich für die Hilfe beim Ausgleich meines orthographischen Rekognitions- sowie Produktionsdefizits.

Natürlich sei auch allen älteren wie jüngeren Probanden gedankt, die als Versuchspersonen an meinen Experimenten teilgenommen haben.

Ein besonders herzlicher Dank geht an meine Familie und Achim, ohne deren vielfältige und unermüdliche Unterstützung diese Arbeit sicherlich nicht entstanden wäre. Insbesondere für ihr Verständnis, ihre Rücksichtnahme und Geduld möchte ich mich an dieser Stelle bedanken und ihnen diese Arbeit daher widmen.

Mein abschließender Dank geht an B.J., der das Fertigstellen dieser Promotion ganz wesentlich vorangetrieben hat und mir zugleich die notwendige Kraft und Gelassenheit gegönnt hat, um dieses Vorhaben zu beenden. Es war schön, die letzten Schritte dieses Wegs mit Dir gemeinsam zu gehen.