

Dissoziation zwischen implizitem und explizitem
Gedächtnis: Artefakt unzureichender Reliabilität
impliziter Gedächtnistests?

Inaugural-Dissertation

zur

Erlangung des Doktorgrades der
Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Christian Kellinger

aus Temeschburg

Düsseldorf, März 2014

Aus dem Institut für Experimentelle Psychologie
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. Axel Buchner

Koreferent: PD Dr. Raoul Bell

Tag der mündlichen Prüfung: 25.04.2014

Inhaltsverzeichnis

Zusammenfassung.....	6
Abstract	7
1 Allgemeine Einleitung.....	8
1.1 Implizites Gedächtnis.....	9
1.1.1 Definition und Abgrenzung	10
1.1.2 Priming und andere Formen impliziten Lernens.....	11
1.1.3 Typische implizite Gedächtnistests und deren Erfassung.....	12
1.2 Dissoziationen zwischen impliziten und expliziten Gedächtnistests	16
1.2.1 Definition einer Dissoziation	17
1.2.2 Studienlage zu den Dissoziationen.....	17
1.2.2.1 Amnesie.....	18
1.2.2.2 Geteilte Aufmerksamkeit.....	19
1.2.2.3 Interferenz	20
1.2.2.4 Verarbeitungstiefe.....	21
1.2.2.5 Depression	23
1.2.2.6 Länge des Retentionsintervalls	24
1.2.2.7 Lebensalter	25
1.2.2.8 Stimulusmodalität und Präsentationsformat.....	26
1.2.2.9 Dissoziationen zwischen impliziten Gedächtnistests.....	26
1.3 Erklärungsmodelle	28
1.3.1 Aktivierung und Elaboration	29
1.3.2 Theorien multipler Gedächtnissysteme.....	30
1.3.3 Prozesstheorien	33
1.4 Erklärungsmodelle im Licht der empirischen Befundlage.....	34
1.4.1 Aktivierungstheorien und die Befundlage	35
1.4.2 Theorien multipler Gedächtnissysteme und die Befundlage.....	37
1.4.3 Prozesstheorien und die Befundlage	39
1.5 Alternative Erklärungen für die Befunde.....	40
1.5.1 Explizite Kontaminationen impliziter Gedächtnistests.....	40
1.5.2 Vergleichbarkeit von Gedächtnistests	46
1.5.3 Reliabilität impliziter Gedächtnistests	47
1.5.3.1 Testtheoretische Definition der Reliabilität	47

1.5.3.2	Zusammenhang der Hauptgütekriterien	48
1.5.3.3	Bedeutung der Reliabilität für die Psychometrie	48
1.5.3.4	Schätzung der Reliabilität	50
1.5.3.5	Dissoziationsbefunde und Reliabilitätsdifferenzen beteiligter Tests	50
1.5.3.6	Ursachen geringer Reliabilität impliziter Gedächtnistests	53
1.5.3.7	Methoden zur Steigerung der Reliabilität impliziter Gedächtnistests.....	54
1.6	Ziele der eigenen Arbeit.....	59
2	Experiment 1	62
2.1	Einleitung	62
2.2	Methode.....	63
2.2.1	Stichprobe	63
2.2.2	Material	64
2.2.3	Prozedur	65
2.2.4	Design	67
2.3	Ergebnisse	68
2.4	Diskussion.....	74
3	Experiment 2.....	78
3.1	Einleitung	78
3.2	Methode.....	80
3.2.1	Stichprobe	80
3.2.2	Material.....	81
3.2.3	Prozedur	81
3.2.4	Design	82
3.3	Ergebnisse	83
3.4	Diskussion.....	87
4	Experiment 3.....	89
4.1	Einleitung	89
4.2	Methode.....	91
4.2.1	Stichprobe	91
4.2.2	Material.....	92
4.2.3	Prozedur	93
4.2.4	Design	96
4.3	Ergebnisse	97
4.4	Diskussion.....	103

5	Allgemeine Diskussion	104
5.1	Hintergründe und Ziele der Arbeit.....	104
5.2	Zusammenfassung der Ergebnisse.....	106
5.3	Bewertung der Befunde, und Schlussfolgerungen.....	107
5.3.1	Befunde zu Reliabilitätsschätzungen	107
5.3.2	Befunde zum Vergleich von Leistungen junger und älterer Probanden.....	111
5.4	Ausblick	112
6	Literaturverzeichnis	114
7	Anhang	127

Zusammenfassung

Dissoziationen zwischen impliziten und expliziten Gedächtnistests werden häufig als funktionale Dissoziationen im Sinne unterschiedlicher Gedächtnissysteme oder unterschiedlicher Gedächtnisprozesse angesehen (Buchner & Wippich, 2000). Dissoziationen können jedoch auch methodische Artefakte repräsentieren, die allein durch Unterschiede in der Reliabilität zweier Gedächtnistests entstehen (Buchner & Wippich, 2000). Da die Reliabilität von Gedächtnistests oft nicht berichtet wird, bleibt eine Mehrdeutigkeit bei der Interpretation dieser Befunde bestehen (Buchner & Brandt, 2003). Für eine eindeutige Interpretation von Dissoziationen ist die Kenntnis der Reliabilität der Tests unabdingbar. In der vorliegenden Arbeit werden drei Experimente vorgestellt, in denen die Annahme überprüft wurde, dass Dissoziationen Artefakte geringer Reliabilität impliziter Gedächtnistest in Verbindung mit hoher Reliabilität expliziter Gedächtnistests sein könnten. In Experiment 1 wurde der implizite Wortstammergänzungstest im Hinblick auf eine Reliabilitätssteigerung optimiert. Durch eine alternative Methode der Bestimmung von Leistungswerten von Probanden wurde eine Reliabilitätssteigerung erreicht, was die Basis für die folgenden Experimente darstellte. In Experiment 2 wurde eine einfache Dissoziation zwischen dem impliziten Wortstammergänzungstest und der expliziten Form des Wortstammergänzungstests als Funktion des Alters repliziert. Dafür wurde die gebräuchliche Methode der Leistungswertbestimmung angewandt, was zu einer geringen Reliabilität des Wortstammergänzungstests führte. Wurde die alternative Methode der Leistungswertbestimmung angewandt, so blieb die Dissoziation, trotz Reliabilitätssteigerung im impliziten Wortstammergänzungstest, bestehen. Auch in Experiment 3, in dem die Befunde aus dem zweiten Experiment auf den impliziten Bildidentifikationstest und den expliziten Bilderkennungstest generalisiert werden sollten, wurden weder durch die gebräuchliche noch durch die alternative Methode der Leistungswertbestimmung, die zu einer Erhöhung der Reliabilität des impliziten Tests führte, Unterschiede zwischen jungen und älteren Probanden im impliziten Bildidentifikationstest aufgedeckt. Die Befunde sprechen dafür, dass geringe Reliabilitätsschätzungen impliziter Gedächtnistests durch die bisher gebräuchliche Methode zur Bestimmung von Leistungswerten begründet sind. Sie sprechen zudem dafür, dass Dissoziationen zwischen impliziten und expliziten Gedächtnistests als Funktion des Alters keine Artefakte geringer Reliabilität impliziter Gedächtnistests sind.

Abstract

Dissociations between implicit and explicit memory tests have been widely interpreted as functional dissociations in the sense of different underlying memory systems or different memory processes (Buchner & Wippich, 2000). These dissociations could also represent methodical artifacts that occur solely because of reliability differences between two memory tests (Buchner & Wippich, 2000). Because reliability estimates of memory tests often are not reported, an ambiguity remains in the interpretation of these findings (Buchner & Brandt, 2003). For unambiguous interpretation of findings of dissociations, knowledge of reliability estimates of involved memory tests is indispensable. The aim of the three experiments in the present work was to test the assumption that dissociations could be artifacts of low reliability of implicit memory tests in association with high reliability of explicit memory tests. In Experiment 1, the reliability of the implicit word-stem completion task was to be enhanced. Through the application of an alternative approach of determining performance scores to the data an increase in the reliability estimate was achieved, which was the foundation for the subsequent experiments. In Experiment 2, at first a simple dissociation between the implicit word-stem completion task and the explicit word-stem cued recall task was replicated, as a function of age. This was accomplished, through applying the common approach of determining performance scores to the data, which led to a low reliability estimate for the word-stem completion task. When the alternative approach of determining performance scores was used, the dissociation pattern remained the same, in spite of an increase in the reliability estimate for the implicit word-stem completion task. Also in Experiment 3, where findings of the second experiment were to be generalized to an implicit picture identification task and an explicit picture recognition task, no differences were found between young and older participants in the implicit picture identification task. This was the case even if the alternative approach of determining performance scores was used, which led to an increase in the reliability estimate for the implicit memory test as compared to the common approach. The findings indicate that the common approach of determining performance scores could be the reason for the low reliability estimates for implicit memory tests to date. They also indicate that dissociations between implicit and explicit memory tests, as a function of age, are not artifacts of low reliability of implicit memory tests.

1 Allgemeine Einleitung

Für Zimbardo und Gerrig (2004, S. 293) ist Gedächtnis die Fähigkeit, Informationen zu enkodieren, zu speichern und zu einem späteren Zeitpunkt wieder abrufen zu können. Es entsteht aus Lernprozessen, die sowohl bewusst als auch unbewusst ablaufen können (Cowan, 1997). Bei der Klassifikation von Gedächtnissystemen unterscheidet man je nachdem, wie lange Informationen behalten werden können, zwischen dem sensorischen Gedächtnis, dem Kurzzeitgedächtnis und dem Langzeitgedächtnis. Abbildung 1 veranschaulicht die verschiedenen Formen bzw. Inhalte des Langzeitgedächtnisses.

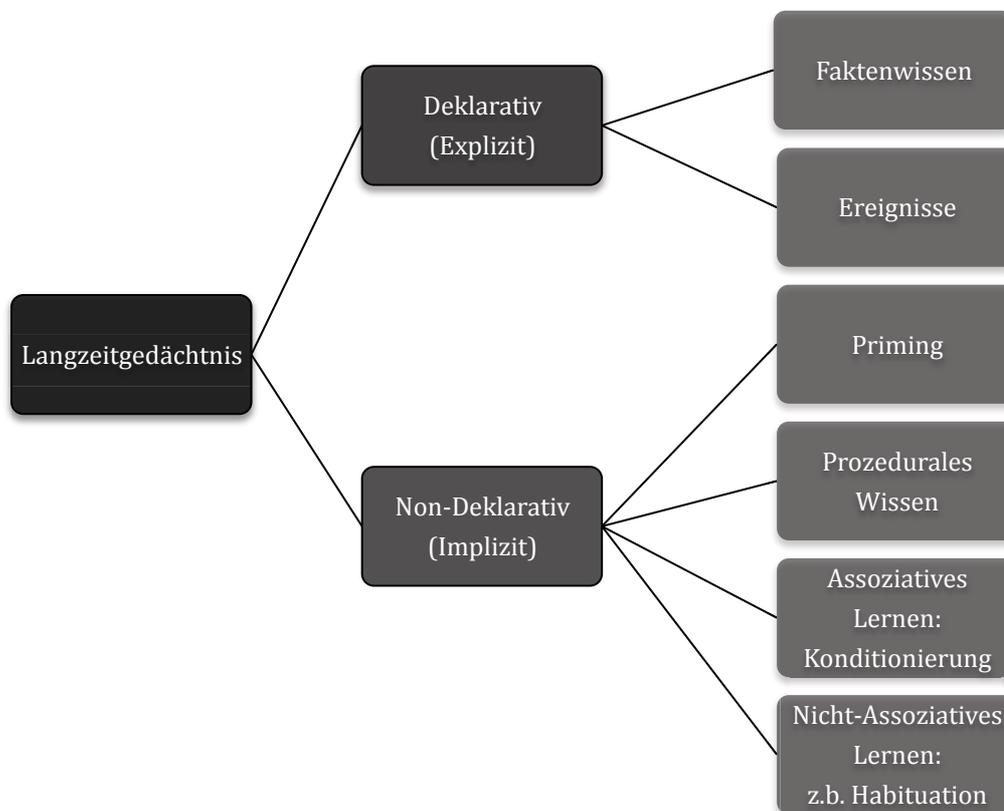


Abbildung 1: Verschiedene Formen bzw. Inhalte des Langzeitgedächtnisses. (Abbildung angepasst nach Kandel, Kupfermann & Iversen, 2000, eigene Übersetzung).

Beim Langzeitgedächtnis unterscheidet man deklaratives Gedächtnis (explizites Gedächtnis) und nondeklaratives Gedächtnis (implizites Gedächtnis). Im deklarativen Gedächtnis wird Faktenwissen (semantisches Wissen) und Wissen über autobiographische Ereignisse (episodisches Wissen) abgelegt. Dieses Wissen ist explizierbar; es

kann verbal darüber berichtet werden. Bestandteile des nondeklarativen Gedächtnisses sind Inhalte, die nicht verbalisierbar sind und automatisierte Prozesse, die unbewusst ablaufen. Dazu gehören prozedurales Wissen, assoziatives Lernen, nichtassoziatives Lernen und Priming. Als Beispiele für prozedurales Wissen seien das Fahrradfahren oder andere motorische Fertigkeiten genannt. Unter Konditionierung, die durch assoziatives Lernen entsteht, versteht man das Erlernen von Reiz-Reaktions-Mustern. Habituation oder Gewöhnung ist ein Beispiel für nichtassoziatives Lernen. Darunter versteht man eine abgeschwächte Reaktion auf einen wiederholt präsentierten Reiz, der sich als nicht bedeutend herausgestellt hat. Auf die verschiedenen Formen von Priming, von denen die Ermittlung des Repetition Priming zur Erfassung von implizitem Gedächtnis in den Experimenten dieser Arbeit diente, soll später näher eingegangen werden.

1.1 Implizites Gedächtnis

Das implizite Gedächtnis ist erst seit einigen Jahrzehnten Gegenstand systematischer wissenschaftlicher Forschung. Es handelt sich dabei um ein Konzept, welches aus der neuropsychologischen Forschung an Amnestikern hervorgegangen ist und später auch in der kognitiven Psychologie, in der Sozialpsychologie und in weiteren Bereichen der Psychologie übernommen und intensiv erforscht wurde. Gedächtnisstudien ohne bewussten Gedächtnisabruf gibt es bereits seit den Untersuchungen von Ebbinghaus (1885/1966) mit der Ersparnis-methode. Damit können Erfahrungsnachwirkungen erfasst werden, selbst wenn nach längeren zeitliche Intervallen kein bewusster Abruf gelernter Inhalte mehr möglich ist. Solche Erfahrungsnachwirkungen sind durch eine verkürzte Lerndauer nachweisbar, die bis zur korrekten Wiedergabe von wiederholt gelernten Listen von Nonsense-Silben benötigt wird, im Vergleich zu der Lernzeit die benötigt wurde, um die gleichen Listen mit Silben beim ersten Lernversuch richtig wiederzugeben. In neuerer Zeit sind vor allem die Studien an dem Amnestiker H.M. (z.B. Milner, Corkin & Teubner, 1968) und die Arbeit von Warrington und Weiskrantz (1970) als Pionierstudien im Bereich der Forschung um das implizite Gedächtnis zu nennen.

1.1.1 Definition und Abgrenzung

Nach Richardson-Klavehn und Bjork (1988) werden die Begriffe «implizites Gedächtnis» und «explizites Gedächtnis» in der Literatur nicht immer einheitlich verwendet. Einerseits sind damit Gedächtnistests gemeint, die entweder implizites Gedächtnis oder explizites Gedächtnis erfassen sollen. Andererseits sind diesen Gedächtnistests hypothetisch zugrunde liegende Gedächtnisprozesse (z.B. Roediger, 1990) oder Gedächtnissysteme (z.B. Tulving & Schacter, 1990) gemeint (Richardson-Klavehn & Bjork, 1988). Manche Autoren sehen hinter dem Begriff «implizites Gedächtnis» weniger ein distinktes Gedächtnissystem oder spezifische Gedächtnisprozesse. Für sie bezieht sich dieser Begriff lediglich auf eine bestimmte Form der Erfassung von Erfahrungsnachwirkungen, für die kein Bezug zur Lernepisode besteht (Buchner & Jansen-Osmann, 2006). Von implizitem Gedächtnis spricht man demnach – auf einer rein deskriptiven Ebene –, wenn vorhergehende Erfahrungen die Leistung in einer Aufgabe verbessern, ohne dass Bewusstsein für diese Erfahrungen bestünde (Graf & Schacter, 1985). Der fehlende Bezug zur Lernepisode ist das entscheidende Merkmal einer impliziten Art der Erfassung von Gedächtnisinhalten. Generell müssen Personen in impliziten Tests konzeptuelles, lexikalisches, perzeptuelles oder prozedurales Wissen zeigen (Richardson-Klavehn & Bjork, 1988). Nach Schacter, Chiu und Ochsner (1993) sind weitere in der Literatur gebrauchte Begriffe mit ähnlicher Bedeutung «Gedächtnis ohne Bewusstsein» (Jacoby & Witherspoon, 1982), «nondeklaratives Gedächtnis» (Squire, 1992) und «indirektes Gedächtnis» (Johnson & Hasher, 1987). Implizite Tests werden deshalb auch als indirekte Tests bezeichnet, weil Probanden Aufgaben bewältigen sollen, deren Lösungswahrscheinlichkeit von den Erfahrungsnachwirkungen vorangegangener Lernepisoden abhängt, wobei jedoch kein Bezug zu den entsprechenden Lernepisoden hergestellt wird. Explizite Tests dagegen bezeichnet man auch als direkte Tests, weil ein direkter Bezug zur Lernepisode hergestellt wird. Genauso wird implizites Gedächtnis oft auch als inzidentelles Gedächtnis und explizites Gedächtnis als intentionales Gedächtnis bezeichnet. In expliziten Tests sollen Probanden frühere Lernepisoden intentional oder willentlich abrufen, während Erfahrungsnachwirkungen in impliziten Tests bei der Bearbeitung einer nicht auf die Lernepisode bezogenen Aufgabe inzidentell oder unwillentlich auftreten. Diese Unterscheidungen beziehen sich jedoch lediglich auf die Art der Erfassung von Gedächtnis, wobei kein Bezug zu bestimmten Erklärungsmodellen hergestellt werden soll (Richardson-Klavehn & Bjork, 1988).

Als Beispiel für eine Aufgabe, die implizite Gedächtnisinhalte erfasst – oder genauer gesagt – eine Aufgabe, die Gedächtnisinhalte auf eine implizite Art erfasst, sei der Wortstammergänzungstest genannt. Vorerfahrungen mit Wörtern, die in einer Lernphase dargeboten werden, können unter perzeptuell anspruchsvollen Bedingungen (z.B. wenn nur einzelne Buchstaben eines Wortes sichtbar sind) in einer späteren Testphase – in der kein Bezug zur Lernepisode hergestellt wird – zu einer besseren Gedächtnisleistung führen (Richardson-Klavehn & Bjork, 1988). Abgegrenzt wird das implizite Gedächtnis von dem expliziten Gedächtnis, dessen Inhalte direkt gemessen werden und dessen Erfassung bis vor wenigen Jahrzehnten einen Großteil der Gedächtnisforschung ausmachte. Durch entsprechende Testinstruktionen wird bei der Erfassung des expliziten Gedächtnisses der Bezug zur Lernepisode hergestellt, was einen bewussten Abruf von gelernten Inhalten möglich macht. Als Beispiel für eine Aufgabe, die explizite Gedächtnisinhalte erfasst, sei das «Wiedererkennen» genannt. Dabei soll beurteilt werden, ob Personen, Ereignisse oder Objekte aus einer vorherigen Lernphase bekannt sind (Buchner & Jansen-Osmann, 2006). Eine weitere explizite Aufgabe ist das «freie Reproduzieren». Dabei sollen ohne weitere Hilfen Personen, Ereignisse oder Objekte abgerufen werden, die in einer vorherigen Lernphase dargeboten wurden (Buchner & Jansen-Osmann, 2006).

1.1.2 Priming und andere Formen impliziten Lernens

Eine häufig untersuchte Form impliziten Lernens ist das Repetition Priming, welches auch als perzeptuelles Priming oder als direktes Priming bezeichnet wird. Dabei handelt es sich um den Effekt einer schnelleren oder genaueren Verarbeitung von Stimuli, deren Verarbeitung durch eine vorherige Darbietung bereits gebahnt wurde. Repetition Priming basiert auf einer perzeptuellen Repräsentation von Stimuli, die durch ihre physischen Eigenschaften definiert ist. Entscheidend für die Entstehung von Repetition Priming ist die Überlappung perzeptueller Eigenschaften von Stimuli zwischen Lernphase und Testphase eines Experimentes. Das Repetition Priming – die am häufigsten untersuchte Form von Priming – kann von dem konzeptuellen Priming unterschieden werden. Beim semantischen Priming – einer Form von konzeptuellem Priming – beeinflusst die vorherige Darbietung von Stimuli die Verarbeitung von nachfolgenden Stimuli, wenn zwischen ihnen eine inhaltliche Beziehung besteht. Semantisches Priming erfordert somit Repräsentationen von Stimuli, die auf

ihrer Bedeutung beruhen.

Zur Erfassung von Repetition Priming in perzeptuellen impliziten Gedächtnistests werden in einer Lernphase beispielsweise Wörter oder Bilder von Objekten dargeboten. In einer anschließenden Testphase werden Stimuli aus der Lernphase und neue Stimuli unter erschwerten visuellen oder akustischen Wahrnehmungsbedingungen dargeboten. Dies kann durch eine kurze Darbietungszeit in der Nähe der Entdeckungsschwelle oder durch eine undeutliche bzw. unvollständige Darbietungsweise erfolgen. Dadurch, dass mit dem ersten Begriff geantwortet werden soll, welcher Probanden zu den Stimuli einfällt, fehlt der Bezug zur Lernepisode. Probanden sollen dazu typischerweise möglichst schnell Wörter ergänzen oder Objekte benennen. Von direktem Priming oder Repetition Priming wird gesprochen, wenn eine schnellere oder genauere Identifikation von Stimuli aus der Lernphase erfolgt, im Vergleich zu einer Basisratenmessung, die an Stimuli durchgeführt wird, die nicht in der Lernphase dargeboten wurden. Indirektes Priming zeigt sich durch jede Änderung in der Testleistung, die aus der Darbietung von Informationen resultiert, welche in assoziativer, semantischer, graphemischer, phonemischer oder morphologischer Weise mit den Teststimuli zusammenhängen (Richardson-Klavehn & Bjork, 1988).

Neben dem direkten Priming und dem indirekten Priming gibt es noch andere Formen des impliziten Lernens, wie z.B. den Erwerb von motorischen Fertigkeiten (prozedurales Wissen). Man denke nur an motorische Abläufe beim Spielen eines Musikinstrumentes oder beim Fahrradfahren. Während Priming typischerweise nach einem einzelnen Durchgang oder nach wenigen Durchgängen beobachtet werden kann und das Gedächtnis für bestimmte Stimuli widerspiegelt, erfordert das Erlernen von Fertigkeiten viele Durchgänge (Schacter et al., 1993). Weitere Beispiele impliziten Lernens sind perzeptuelles Lernen und assoziatives Lernen, wie z.B. die klassische Konditionierung oder die operante Konditionierung. Ihnen allen ist gemeinsam, dass sie nicht bewusstseinspflichtig sind und schwer verbalisierbar oder explizierbar sind.

1.1.3 Typische implizite Gedächtnistests und deren Erfassung

Es gibt verschiedene Möglichkeiten, implizite Gedächtnistests zu kategorisieren. Ein Unterscheidungskriterium ist die Art des verwendeten Stimulusmaterials, ein weite-

res die Sinnesmodalität, mit der Stimuli erfasst werden. Es gibt Primingstudien zum visuellen Priming von Wörtern, zum visuellen Priming von Objekten und zum auditorischen Priming von Wörtern. Die Mehrheit der Primingstudien wurde zum visuellen Priming von Wörtern durchgeführt. Eine dritte Möglichkeit der Kategorisierung impliziter Gedächtnistests ist die Unterscheidung in illusionsbasierte Tests und in leistungsorientierte Tests.

Im Gegensatz zu leistungsorientierten impliziten Tests, bei denen die Gedächtnisleistung direkt – d.h. mit Hilfe der Genauigkeit oder der Schnelligkeit von Probandenurteilen erfasst wird – verwendet man bei illusionsbasierten Tests Gedächtnis als Werkzeug, um Aufgaben zu lösen (Jacoby & Kelley, 1987). Illusionsbasiert sind beispielsweise solche impliziten Gedächtnistests, bei denen die Einschätzung der Berühmtheit von Personen als Indikator für Erfahrungsnachwirkungen verwendet wird. Sie basieren auf dem sogenannten «Effekt falscher Berühmtheit» (Jacoby, Kelley, Brown & Jasechko, 1989), wobei vorher dargebotene Namen von berühmten und nicht berühmten Personen eher als berühmt eingestuft werden als Namen von Personen, die nicht dargeboten wurden. Die Entstehung dieser Illusion läuft nach Buchner und Jansen-Osmann (2006) in zwei Schritten ab. In dem ersten Schritt werde wiederholt dargebotene Information schneller verarbeitet, was einen Vertrautheitseindruck schafft. In dem zweiten Schritt werde die schnellere Verarbeitung der Information automatisch einem Kontext zugeordnet. Normalerweise wird neben dem Vertrautheitseindruck auch der Kontext erinnert, in dem die Information vorher dargeboten wurde (Buchner & Jansen-Osmann, 2006). Wenn der richtige Kontext jedoch nicht mehr verfügbar ist, kann dieser Vertrautheitseindruck fälschlicherweise auf eine andere aktuell verfügbare Urteilsdimension attribuiert werden, z.B. auf die Berühmtheit von Personen (Buchner & Jansen-Osmann, 2006). Im Laufe der Zeit wird der ursprüngliche Lernkontext, in dem die Information dargeboten wurde, immer weniger verfügbar und der Vertrautheitseindruck wird zunehmend auf die neue Urteilsdimension fehlattribuiert (Buchner & Jansen-Osmann, 2006). Nach Buchner und Brandt (2003) können diese Urteilsdimensionen sehr vielfältig sein, d.h. Aussagen können eher für wahr (Begg, Anas & Farinacci, 1992), Dinge für schöner (Kunst-Wilson & Zajonc, 1980) oder Antworten als eher richtig (Kelley & Lindsay, 1993) angesehen werden.

Bei den zur Erfassung impliziten Gedächtnisses häufiger verwendeten leistungsori-

entierten Tests wird – je nach Aufgabentyp – die Genauigkeit oder die Geschwindigkeit gemessen, mit denen Urteile abgegeben werden. Diese Tests können weiterhin entweder als eher perzeptuell oder als eher konzeptuell eingestuft werden (Roediger, Marsh & Lee, 2002). Bei perzeptuellen impliziten Tests werden vor allem *physische* Stimuluseigenschaften verarbeitet, während bei konzeptuellen impliziten Tests die *Bedeutung* von Stimuli im Vordergrund steht. Perzeptuelle implizite Tests zeichnen sich dadurch aus, dass visuelle oder akustische Hinweisreize, die in der Testphase eines Experimentes unter erschwerten Wahrnehmungsbedingungen dargeboten werden, perzeptuelle Ähnlichkeit mit Stimuli der Lernphase haben. Typischerweise sollen mit Hilfe dieser Hinweisreize Stimuli aus der Lernphase sowie neue, nicht dargebotene Stimuli, benannt werden. Perzeptuelle Prozesse wirken dann am besten, wenn die physischen Eigenschaften von Stimuli aus der Lernphase mit den physischen Eigenschaften der entsprechenden Hinweisreize in der Testphase übereinstimmen. Wenn jedoch eine Änderung des Stimulusformats (z.B. von Wörtern in der Lernphase, zu Bildern in der Testphase) oder eine Änderung der Stimulusmodalität erfolgt (z.B. von auditiver zu visueller Darbietung), so gibt es keine Überlappung physischer Eigenschaften mehr. Perzeptuelle Prozesse können dann nicht mehr richtig wirken, das Priming in perzeptuellen Tests wird geringer oder verschwindet ganz (z.B. Jacoby & Dallas, 1981; Kirsner & Smith, 1974). Perzeptuelles Priming ist damit durch eine hohe Spezifität gekennzeichnet, da seine Existenz in hohem Umfang von Änderungen zwischen den Bedingungen bei der Enkodierung und den Bedingungen beim Abruf abhängig ist. Im Gegensatz dazu können konzeptuelle Prozesse am besten zur Geltung kommen, wenn es eine Überlappung semantischer Eigenschaften zwischen Lernphase und Testphase gibt. Hinweisreize der Testphase sind hier auf einer konzeptuellen Ebene mit den Stimuli aus der Lernphase verbunden, ohne eine perzeptuelle Ähnlichkeit zu besitzen. Die Änderung des Stimulusformats hat bei konzeptuellen Tests deshalb keine Auswirkung auf die Höhe des Primings (z.B. Mitchell & Bruss, 2003).

Häufig verwendete perzeptuelle implizite Gedächtnistests sind Wortfragmentergänzung, Wortstammergänzung, Wortidentifikation und Bildfragmentidentifikation. Bei dem Wortfragmentergänzungstest werden in einer Lernphase ganze Wörter dargeboten. In der darauf folgenden Testphase werden Wortfragmente gezeigt. Das sind Wörter, bei denen an einigen Stellen Buchstaben fehlen (z.B. Fr_g__n_ für das Wort Fragment). Die Wortfragmente sollen mit dem ersten Wort ergänzt werden, welches

Probanden dazu einfällt. Die Hälfte der Wortfragmente kann mit Wörtern ergänzt werden, die in der Lernphase dargeboten wurden. Die andere Hälfte der Wortfragmente kann mit Wörtern einer zweiten – neuen – Wortliste ergänzt werden. Priming ist vorhanden, wenn mehr Wortfragmente mit den entsprechenden Wörtern aus der Lernphase ergänzt werden können, als Wortfragmente mit Wörtern der neuen Wortliste. Bei der Wortstammergänzungsaufgabe werden – im Unterschied zur Wortfragmentergänzungsaufgabe – nur die ersten drei Buchstaben von Wörtern (z.B. Ler___ für das Wort Lerche) dargeboten. Auch hier kann die Hälfte der Wortstämme mit Wörtern aus der Lernphase ergänzt werden. Die andere Hälfte der Wortstämme gehört zu einer zweiten Wortliste und dient als Basisratenkorrektur. Wortstämme sollen mit dem ersten Wort ergänzt werden, welches Probanden dazu einfällt. Von Priming wird gesprochen, wenn mehr Wortstämme mit Wörtern aus der Lernphase ergänzt werden, als Wortstämme mit Wörtern der zweiten Liste. Bei der Wortidentifikationsaufgabe werden, nachdem in einer Lernphase Wörter dargeboten wurden, in der darauf folgenden Testphase sowohl Wörter aus der Lernphase, als auch neue Wörter gezeigt. Diese Wörter werden für einen kurzen Zeitraum – in der Nähe der Entdeckungsschwelle – eingeblendet. Probanden sollen die Wörter identifizieren, wobei der Bezug zur Lernepisode nicht hergestellt wird. Die Identifikationsleistung für Wörter, die bereits in der Lernphase gezeigt wurden, liegt dabei typischerweise über der Identifikationsleistung für Wörter, die neu sind. Kommen in der Lernphase Bilder als Stimulusmaterial zum Einsatz, werden diese Bilder – sowie die gleiche Anzahl an neuen Bildern – für kurze Zeit in der anschließenden Testphase eingeblendet und sollen von den Probanden identifiziert werden. In einer alternativen Form dieser Aufgabe werden in der Testphase sowohl Bilder aus der Lernphase, als auch neue Bilder fragmentiert dargeboten. Bei dieser sogenannten Klarifikationsprozedur werden verschiedene Masken über die Bilder gelegt. Am Anfang der Prozedur werden die Bilder vollständig verdeckt. Mit jeder Defragmentationsstufe werden die Bilder dann deutlicher, indem die verwendeten Masken einen immer größer werdenden Anteil der Bilderfläche freigeben, bis es zu einer Identifikation kommt. Die Leistung wird bei der Bildidentifikationsaufgabe über die Anzahl der identifizierten Bilder bestimmt. Bei der Klarifikationsprozedur wird die Leistung über die Anzahl der Defragmentationsschritte ermittelt, die bis zur Identifikation benötigt werden. Diese Rohwerte werden einer Basisratenkorrektur unterzogen, die an Bildern erfolgt, die nicht in der Lernphase gezeigt wurden.

Obwohl typischerweise explizite Tests konzeptueller Natur sind, gibt es auch einige konzeptuelle implizite Tests, wie z.B. den Kategorien-Assoziationstest und den Wort-Assoziationstest. Beim Kategorien-Assoziationstest bekommen Probanden Namen von Kategorien vorgelegt, von denen einige dieser Kategorien Exemplare haben, die in einer vorangegangenen Lernphase gezeigt wurden. Für andere der gezeigten Kategorien wurden keine Exemplare in der Lernphase dargeboten. Es sollen nun mehrere Exemplare für jede der präsentierten Kategorien genannt werden. Konzeptuelles implizites Priming ist dann vorhanden, wenn Exemplare für Kategorien öfter genannt werden, wenn sie bereits in der Lernphase gezeigt wurden, als dann, wenn sie nicht in der Lernphase gezeigt wurden. Beim freien Wort-Assoziationstest werden assoziativ verbundene Wortpaare, sogenannte «Stimulus-Response» Paare (z.B. Haus-Tür), in einer Lernphase dargeboten. In der darauf folgenden Testphase werden die Stimulusanteile (z.B. Haus) der Wortpaare aus der Lernphase und die Stimulusanteile neuer Wortpaare – als Hinweis für die entsprechenden Antwortanteile aus der Lernphase – gezeigt. Probanden sollen mit dem ersten Wort antworten, welches ihnen zu einem bestimmten Stimulus einfällt. Priming ist dann vorhanden, wenn auf die Stimulusanteile in der Testphase häufiger die korrekten Antwortanteile der Wortpaarungen aus der Lernphase genannt werden, als auf Stimulusanteile un-studierter Wortpaarungen korrekte Antwortanteile genannt werden.

1.2 Dissoziationen zwischen impliziten und expliziten Gedächtnistests

Das Interesse an der Erforschung des impliziten Gedächtnisses kam mit den ersten Gedächtnisstudien an Amnestikern auf. Amnestiker haben üblicherweise ein intaktes Kurzzeitgedächtnis mit normaler Gedächtnisspanne. Das Langzeitgedächtnis jedoch, vor allem das Behalten neuer Inhalte, ist bei anterograder Amnesie stark eingeschränkt. So berichten Milner et al. (1968) über den Amnestiker H.M., dem nach einer Operation im Jahre 1953 zur Linderung seiner epileptischen Anfälle bilateral der Hippocampus und angrenzende Strukturen entfernt wurden. Amnestiker wie der Patient H.M. zeigen normale Leistungen bei impliziten Aufgaben wie motorischem Lernen (z.B. Corkin, 1968) und Repetition Priming (z.B. Warrington & Weiskrantz, 1968, 1970, 1974). Sie haben jedoch kein Bewusstsein dafür, diese Aufgaben jemals in Angriff genommen zu haben. Auf der anderen Seite sind die Leistungen von Amne-

stikern im Vergleich zu gesunden Personen in expliziten Gedächtnistests gering.

1.2.1 Definition einer Dissoziation

Von einer Dissoziation wird gesprochen, wenn sich die Manipulation bestimmter experimenteller Variablen (z.B. Interferenz, Verarbeitungstiefe, Länge des Retentionsintervalls) unterschiedlich auf die Leistungen in impliziten und expliziten Gedächtnistests auswirkt. Es kann zwischen einfachen Dissoziationen und doppelten Dissoziationen unterschieden werden. Bei einer einfachen Dissoziation wirkt sich die Manipulation von experimentellen Variablen nur auf einen der beiden Gedächtnistests aus. Bei einer doppelten Dissoziation werden durch die Manipulation von experimentellen Variablen beide Gedächtnistests in entgegengesetzter Weise beeinflusst. Wirkt sich die Manipulation von experimentellen Variablen jedoch in gleicher Weise auf die Leistung in impliziten und expliziten Gedächtnistests aus, so spricht man von Assoziationen oder von parallelen Effekten.

1.2.2 Studienlage zu den Dissoziationen

Zahlreiche Studien bestätigten die mit gesunden Kontrollgruppen vergleichbar guten Leistungen von Amnestikern in impliziten Gedächtnistests (Graf, Shimamura & Squire, 1985; Jacoby & Witherspoon, 1982; Johnson, Kim & Risse, 1985; Schacter, 1985; Shimamura & Squire, 1984; Squire & Cohen, 1984; Squire & McKee, 1992), während sie in expliziten Gedächtnistests deutlich schlechter abschnitten. Dissoziationen zwischen impliziten und expliziten Gedächtnistests wurden – neben dem Vergleich von Amnestikern und gesunden Personen – für die verschiedensten Variablen gefunden, wie Alzheimer (z.B. Winograd, Goldstein, Monarch, Peluso & Goldman, 1999), geteilte Aufmerksamkeit (z.B. Jacoby, Woloshyn & Kelley, 1989; Parkin & Russo, 1990), Interferenz (z.B. Graf & Schacter, 1987; Sloman, Hayman, Ohta, Law & Tulving, 1988), Schizophrenie (z.B. Bazin & Perruchet, 1996; Marie et al., 2001; Schwartz, Rosse & Deutsch, 1993), Verarbeitungstiefe (z.B. Graf & Mandler, 1984; Jacoby, 1983; Jacoby & Dallas, 1981; Roediger, Weldon, Stadler & Riegler, 1992), Depression (z.B. Bazin, Perruchet, Debonis & Féline, 1994; Denny & Hunt, 1992; Hertel & Hardin, 1990) und der Länge des Retentionsintervalls (z.B. Tulving, Schacter & Stark, 1982; Tunney, 2003). Das Lebensalter ist eine weitere – häufig untersuchte – Variable, bei

der Dissoziationen zwischen impliziten und expliziten Gedächtnistests gefunden wurden (z.B. Fleischman, Wilson, Gabrieli, Bienas & Bennett, 2004; Halpern & O'Connor, 2000; Light, Prull, LaVoie & Healy, 2000; Light & Singh, 1987; Light, Singh & Capps, 1986; Lustig & Buckner, 2004; Mitchell & Bruss, 2003; Rabbit, 1982, 1984; Wiggs, 1993). Keine der genannten Variablen hatte Einfluss auf die Leistungen in impliziten Gedächtnistests, jedoch einen hohen Einfluss auf die Leistungen in expliziten Gedächtnistests. Die Änderung der Stimulusmodalität oder des Präsentationsformates zwischen Lernphase und Testphase von Experimenten hat dagegen Einfluss auf perzeptuelle implizite Gedächtnistests. Sie hat jedoch weder Einfluss auf konzeptuelle implizite Gedächtnistests, noch auf explizite Gedächtnistests (z.B. Berry, Banbury & Henry, 1997; Craik, Moscovitch & McDowd, 1994; Graf et al., 1985; Mitchell & Bruss, 2003; Weldon & Roediger, 1987). So lassen sich selbst implizite Gedächtnistests untereinander dissoziieren, wenn der eine Test konzeptueller Natur, der andere Test jedoch perzeptueller Natur ist (z.B. Blaxton, 1989; Mitchell & Bruss, 2003; Rappold & Hashtroudi, 1991; Roediger et al., 2002; Srinivas & Roediger, 1990; Weldon & Roediger, 1987).

Nachfolgend werden exemplarisch einzelne Experimente näher betrachtet, in denen – nach Manipulation bestimmter experimenteller Variablen – Dissoziationen zwischen impliziten und expliziten Gedächtnistests gefunden wurden. Weiterhin werden Experimente beschrieben, in denen parallele Effekte bzw. Assoziationen zwischen impliziten und expliziten Gedächtnistests gefunden wurden.

1.2.2.1 Amnesie

Warrington und Weiskrantz (1970) zeigten in ihrem zweiten Experiment, dass Amnestiker eine mit gesunden Personen vergleichbare Leistung im Repetition Priming des Wortstammergänzungstests und des Wortfragmentergänzungstests aufweisen. Die Leistungen in expliziten Gedächtnistests jedoch – wie dem freien Abruf von Wörtern und der Rekognition von Wörtern –, waren deutlich geringer als die Leistungen gesunder Personen. Graf et al. (1985) untersuchten in ihrem ersten Experiment Amnestiker, die unter dem alkoholinduzierten Korsakoffsyndrom leiden, welches zu einer Schädigung von Teilen des limbischen Systems und der Frontallappen führt. Sie verglichen die Leistungen der Korsakoff-Patienten in einem impliziten Wortstammergänzungstest und einem expliziten Gedächtnistest zum freien Abruf

von Wörtern, mit den Leistungen von Alkoholikern ohne Korsakoffsyndrom, sowie mit einer Gruppe von weiteren Patienten, die keine Alkoholiker waren. In dem expliziten Test zeigten die beiden Kontrollgruppen die bessere Leistung. In dem impliziten Test gab es keinen Unterschied zwischen den Korsakoff-Patienten und den beiden Kontrollgruppen. Squire und McKee (1992) boten in einer Lernphase Amnestikern und einer Kontrollgruppe gesunder Personen Namen von berühmten und nicht berühmten Personen zum Lesen dar. Beide Gruppen beurteilten in einer anschließenden Testphase Namen als eher berühmt, wenn diese zuvor dargeboten wurden, und unabhängig davon, ob diese Namen zu berühmten oder nicht berühmten Personen gehörten. Verglichen mit Personen aus der gesunden Kontrollgruppe waren Amnestiker jedoch deutlich schlechter in einer expliziten Rekognitionsaufgabe.

1.2.2.2 Geteilte Aufmerksamkeit

In ihrem ersten Experiment untersuchten Jacoby, Woloshyn et al. (1989) – mit Hilfe des «Effektes falscher Berühmtheit» – den Einfluss geteilter Aufmerksamkeit auf den bewussten Abruf von Namen und auf den Vertrautheitseindruck von Namen. In einer Lernphase wurden Namen berühmter und nicht berühmter Menschen entweder unter Bedingungen voller Aufmerksamkeit oder unter Bedingungen geteilter Aufmerksamkeit dargeboten. In der Bedingung mit voller Aufmerksamkeit sahen die Probanden nur Namen von Personen. In der Bedingung mit geteilter Aufmerksamkeit jedoch sahen die Probanden Namen und sollten gleichzeitig aus einer Folge von Zahlen eine Sequenz aus drei ungeraden Zahlen heraushören. Anschließend wurde ein expliziter Rekognitionstest oder ein impliziter Test zur Beurteilung der Berühmtheit von Personen vorgelegt. Unter Bedingungen geteilter Aufmerksamkeit kam es in der Testphase zu einer deutlich schlechteren Rekognitionsleistung als unter Bedingungen voller Aufmerksamkeit. Die Größe des Effektes der Vertrautheit hingegen, unterschied sich nicht zwischen der Bedingung geteilter Aufmerksamkeit und der Bedingung voller Aufmerksamkeit, gemessen an den Berühmtheitsurteilen der Probanden. In der ersten Phase des Experimentes von Parkin und Russo (1990), benannten Personen fragmentierte Bilder von alltäglichen Objekten in einer Klarifikationsprozedur. Dazu wurden Bilder von Objekten zunächst in ihrer meist fragmentierten Form gezeigt. Die Bilder wurden dann mit jedem Schritt der Defragmentationsprozedur deutlicher, bis die Objekte schließlich identifiziert werden konnten. Es gab dabei zwei Aufmerksamkeitsbedingungen. Unter voller Aufmerksamkeit war

nur die genannte Aufgabe zu bewältigen. In der Bedingung geteilter Aufmerksamkeit jedoch war gleichzeitig eine zweite Aufgabe zu bewältigen. Und zwar wurden eine Reihe von Tönen – in zufälligen Intervallen – für eine Dauer von zwei bis fünf Sekunden dargeboten. Diese Töne sollten bezüglich ihrer Tonhöhe eingestuft werden, während gleichzeitig die fragmentierten Bilder identifiziert wurden. Einen Tag später wurde der Hälfte der Personen beider Aufmerksamkeitsbedingungen ein expliziter Test vorgelegt, in dem sie die Namen der Objekte vom Vortag abrufen sollten. Der anderen Hälfte von Personen wurde ein impliziter Bildidentifikationstest vorgelegt, in dem Bilder vom Vortag sowie neue Bilder – in einer Klarifikationsprozedur – identifiziert werden sollten. Die Leistung in der impliziten Aufgabe wurde durch den Vergleich der Identifikationsleistung für Bilder vom Vortag mit der Identifikationsleistung für neue Bilder bestimmt. Die Höhe des Primings in der impliziten Aufgabe unterschied sich nicht zwischen der Bedingung voller Aufmerksamkeit und der Bedingung geteilter Aufmerksamkeit. Die Abrufleistung in der expliziten Aufgabe war jedoch in der Bedingung geteilter Aufmerksamkeit schlechter als in der Bedingung voller Aufmerksamkeit.

1.2.2.3 Interferenz

Graf und Schacter (1987) boten ihren Versuchspersonen in der Lernphase ihres ersten Experimentes nicht assoziierte Wortpaare entweder unter Interferenzbedingungen oder unter Bedingungen ohne Interferenz dar. Die Probanden sollten sich Wortpaare einer Zielliste sowie Wortpaare einer Interferenzliste merken. Unter Interferenzbedingungen entsprach das jeweils erste Wort der Wortpaare der Zielliste dem ersten Wort der Wortpaare der Interferenzliste. In der Kontrollbedingung ohne Interferenz hatten die Wortlisten keine Wörter gemeinsam. In der Bedingung von proaktiver Interferenz wurde die Zielliste nach der Interferenzliste dargeboten, während in der Bedingung von retroaktiver Interferenz zuerst die Zielliste und anschließend die Interferenzliste dargeboten wurde. Darauf folgte entweder ein impliziter Gedächtnistest oder ein expliziter Gedächtnistest. Das explizite Gedächtnis wurde erfasst, indem Probanden – mit Hilfe von Abrufhinweisen – Informationen aus der Lernphase abrufen sollten. Das implizite Gedächtnis wurde durch ein Wortstammergänzungstest erfasst. Gezeigt wurde jeweils das erste Wort eines Wortpaares, sowie die ersten drei Buchstaben des zweiten Wortes. In der expliziten Form des Tests sollten die Probanden mit Hilfe der ersten drei Buchstaben des zweiten Wortes eines Wortpaa-

res (Abrufhinweis), die entsprechenden Wörter der Wortpaare aus der Lernphase abrufen. In der impliziten Form des Tests sollte mit dem ersten Wort ergänzt werden, welches den Probanden dazu einfiel. Graf und Schacter (1987) fanden, dass Interferenz die Leistungen im expliziten Test verschlechtert, während die Leistungen im impliziten Test von Interferenz unbeeinträchtigt sind. Andere Autoren jedoch fanden, dass auch die Gedächtnisleistung in impliziten Tests durch Interferenz beeinträchtigt wird (z.B. Lustig & Hasher, 2001a, 2001b; Winocur, Moscovitch & Bruni, 1996). Auch Ikier und Hasher (2006), sowie Ikier, Yang und Hasher (2008) fanden proaktive Interferenz in impliziten Tests. Diese war bei älteren Personen stärker ausgeprägt als bei jüngeren Personen.

1.2.2.4 Verarbeitungstiefe

Von einer tiefen Verarbeitung wird gesprochen, wenn durch semantische Anweisungen in der Lernphase die Aufmerksamkeit von Probanden auf die inhaltliche Bedeutung von Stimuli gerichtet wird (Schacter et al., 1993). Dies ist beispielsweise der Fall, wenn Urteile über die Bedeutsamkeit von Stimuli abgegeben werden sollen, oder wenn eingeschätzt werden soll, wie angenehm Stimuli sind. Von einer oberflächlichen Verarbeitung jedoch wird gesprochen, wenn durch entsprechende Anweisungen in der Lernphase die Aufmerksamkeit von Probanden auf physische Eigenschaften von Stimuli gerichtet wird (Schacter et al., 1993). Dies kann erreicht werden, wenn beispielsweise beurteilt werden soll, wie häufig Vokale in Wörtern enthalten sind. Oder es soll beurteilt werden, ob Konsonanten oder Vokale häufiger in Wörtern vorkommen. Jacoby und Dallas (1981) präsentierten Personen Wörter und stellten ihnen Fragen dazu. Einige dieser Fragen bezogen sich auf physische Eigenschaften der Wörter und zielten auf eine oberflächliche Verarbeitung ab. Es sollte beispielsweise die Anzahl der Buchstaben dieser Wörter gezählt werden. Andere Fragen bezogen sich auf die phonologischen Eigenschaften der Wörter. Dazu wurde gefragt, ob sich ein Wort auf ein anderes Wort reimt. Eine dritte Art von Fragen bezog sich auf semantische Eigenschaften der Wörter und sollte für eine tiefe Verarbeitung sorgen. Dazu wurden Fragen zur inhaltlichen Bedeutung der Wörter gestellt. Danach bekamen die Personen entweder einen expliziten Rekognitionstest für Wörter aus der Lernphase präsentiert oder einen impliziten perzeptuellen Identifikationstest. Bei diesem wurden Wörter aus der Lernphase sowie neue Wörter für eine kurze Zeitspanne eingeblendet und sollten von den Personen identifiziert werden.

Die Wiedererkennungsleistung im expliziten Test war am besten, wenn Fragen zu semantischen Eigenschaften von Wörtern gestellt wurden. Darauf folgte die Wiedererkennungsleistung, wenn Fragen zu phonologischen Eigenschaften von Wörtern gestellt wurden. Am schlechtesten war die Leistung, wenn Fragen zu physischen Eigenschaften von Wörtern gestellt wurden. Die Höhe des Primings in der impliziten Aufgabe war jedoch für alle Stufen der Verarbeitungstiefe gleich. Ein seltenes Beispiel für eine doppelte Dissoziation liefert die Studie von Jacoby (1983), in der das Lesen von Wörtern höheres Priming hervorbrachte, als das selbstständige Generieren von Wörtern. Im Gegensatz dazu wurden Wörter, die in der Lernphase lediglich gelesen wurden, schlechter wiedererkannt, als Wörter, die selbstständig generiert wurden. Slamecka und Graf (1978) beschrieben den sogenannten «Generation Effekt» als die Beobachtung, dass explizite Gedächtnisleistung für Wörter, die in einer Lernphase selbstständig mit Hilfe von Hinweisen generiert werden, typischerweise besser ist als explizite Gedächtnisleistung für Wörter, die nur gelesen werden. Personen bekamen in der Lernphase des dritten Experimentes von Graf und Mandler (1984) Wortlisten entweder mit Anweisungen zur semantischen Verarbeitung oder mit Anweisungen zur perzeptuellen Verarbeitung vorgelegt. In der unmittelbar danach folgenden Testphase ergänzten sie Wortstämme entweder mit impliziten Testanweisungen oder mit expliziten Testanweisungen. Beim impliziten Wortstammergänzungstest sollten Probanden Wortstämme mit dem ersten Wort ergänzen, welches ihnen dazu einfiel. In der expliziten Form des Tests sollten die Wortstämme als Abrufhinweis für Wörter aus der Lernphase genutzt werden. Es zeigte sich eine typische Dissoziation zwischen dem impliziten und dem expliziten Gedächtnistest. Die Leistungen in dem expliziten Test profitierten von der tiefen Verarbeitung der Wörter in der Lernphase, während kein Effekt der Verarbeitungstiefe in dem impliziten Test nachzuweisen war. Roediger et al. (1992) replizierten die Befunde von Graf und Mandler (1984) und erweiterten sie auf den impliziten Wortfragmentergänzungstest. Sie variierten in einer Lernphase die Verarbeitungstiefe von Wörtern entweder über semantische Anweisungen oder über graphemische Anweisungen. Auch sie fanden einen großen Effekt der Verarbeitungstiefe auf die Leistungen von Probanden in expliziten Tests. In dem impliziten Wortstammergänzungstest und Wortfragmentergänzungstest jedoch waren keine Unterschiede zwischen semantischer und graphemischer Verarbeitung nachzuweisen. Brown und Mitchell (1994) fanden allerdings in über der Hälfte aller Studien ihrer Meta-Analyse einen bedeutsamen Effekt der Verarbeitungstiefe

für implizite Gedächtnistest. Das Priming in diesen Tests war nach semantischer Verarbeitung von Stimuli höher als es nach nonsemantischer Verarbeitung war. Der Befund war unabhängig davon, ob es sich um perzeptuelle implizite Tests oder um konzeptuelle implizite Tests handelte. Auch Meier (2001) konnte keine Dissoziation zwischen impliziten und expliziten Gedächtnistests finden. Er variierte die Verarbeitungstiefe von Stimuli in einem impliziten Bildidentifikationstest und einem expliziten Rekognitionstest. Sowohl expliziter Test als auch impliziter Test waren von der Manipulation der Verarbeitungstiefe betroffen.

1.2.2.5 Depression

Der Stimmungskongruenzeffekt besagt, dass die Enkodierung und der Abruf von Informationen leichter vonstatten geht, wenn Informationen emotional kongruent sind mit der Stimmung in der sich Personen gerade befinden, als wenn Informationen emotional inkongruent mit der Stimmung sind. Verschiedene Studien untersuchten den Einfluss von klinisch manifester Depression – sowie den Einfluss von künstlich induzierter Stimmungslage – auf die Leistung in impliziten und expliziten Gedächtnistests. Denny und Hunt (1992) beispielsweise verglichen klinisch depressive Personen mit einer nicht depressiven Kontrollgruppe, indem sie ihnen Wörter mit negativer emotionaler Bindung oder Wörter mit positiver emotionaler Bindung zeigten. Die beiden Gruppen sollten auf einer sechsstufigen Skala beurteilen, wie gut die jeweiligen Wörter ihre aktuelle Situation beschreiben. Später bekamen sie einen expliziten Gedächtnistest zum freien Abruf von Wörtern und einen impliziten Wortfragmentergänzungstest vorgelegt. Es zeigte sich, dass depressive Personen im expliziten Test Wörter mit negativer emotionaler Bindung besser erinnerten, als Wörter mit positiver emotionaler Bindung. Nicht depressive Personen erinnerten Wörter mit positiver emotionaler Bindung besser, als Wörter mit negativer emotionaler Bindung. Im impliziten Test jedoch hatte die emotionale Bindung der Wörter den gleichen Einfluss auf die Leistung von Probanden beider untersuchter Gruppen. Um die Vergleichbarkeit des Stimulusmaterials expliziter und impliziter Tests zu maximieren, verglichen Bazin et al. (1994) depressive und gesunde Personen in dem impliziten Wortstammergänzungstest und der expliziten Form des Wortstammergänzungstests miteinander. Die beiden Tests unterschieden sich lediglich bezüglich der Testanweisungen, die nur im expliziten Fall einen Bezug zur Lernepisode herstellen. Im expliziten Test zeigten klinisch depressive Personen schwächere Leistungen als gesunde

Personen. Im impliziten Test jedoch schnitten beide Gruppen gleich gut ab.

1.2.2.6 Länge des Retentionsintervalls

Tulving et al. (1982) untersuchten den Einfluss der Länge des Retentionsintervalls, d.h. dem zeitlichen Abstand zwischen Lernphase und Testphase, auf die Leistung in impliziten und expliziten Gedächtnistests. In der Lernphase ihres Experimentes boten sie dazu Personen zunächst eine Liste mit Wörtern dar. In Testphasen – die eine Stunde später sowie eine Woche danach stattfanden – bekamen die Personen einen expliziten Rekognitionstest und einen impliziten Wortfragmentergänzungstest vorgelegt. Nach einer Woche zeitlichem Abstand zwischen Lernphase und Testphase war die Rekognitionsleistung sehr schlecht, während die Primingwerte im Wortfragmentergänzungstest unverändert blieben. Tunney (2003) fand in einer Studie zum Erwerb künstlich erzeugter Grammatikregeln, dass Priming in einem impliziten Test länger bestehen bleibt, als die Gedächtnisleistung in einem expliziten Rekognitionstest. In einer typischen Aufgabe dazu wurden Personen eine Folge von Buchstaben präsentiert, die einem künstlich geschaffenen Regelwerk genügten und auf die sie mit einem entsprechenden Tastendruck reagieren sollten. Einer Kontrollgruppe wurden zufällig erzeugte Buchstabenfolgen gezeigt. Die Personen wurden sofort, nach sieben Tagen und nach zwei Wochen getestet. In der Testphase des Experimentes reagierten die Personen auf alte und neue Buchstabenfolgen in der gleichen Weise, wie sie es in der Lernphase getan hatten. Es wurden auch hier die Reaktionszeiten erfasst. Anschließend wurde ein Rekognitionstest durchgeführt, in dem die Personen gefragt wurden, ob sie die Buchstabenfolgen aus der Lernphase wiedererkannten. Das Priming des impliziten Tests wurde als Differenz der Reaktionszeiten auf die alten und neuen Buchstabenfolgen berechnet. Wenn die Testphase unmittelbar auf die Lernphase folgte, waren Priming und Rekognitionsleistung bei der Gruppe, denen nach einem Regelwerk erzeugte Buchstabenfolgen präsentiert wurden höher, als bei der Kontrollgruppe, denen zufällige Buchstabenfolgen präsentiert wurden. Sieben Tage sowie zwei Wochen später, blieb die Höhe des Primings auf gleichem Niveau. Die Rekognitionsleistung jedoch nahm ab. Einige Studien jedoch fanden keine Dissoziation zwischen impliziten und expliziten Gedächtnistests, als Funktion der Länge des Retentionsintervalls. Ein Beispiel dafür ist das Experiment von McBride und Doshier (1997). In einem Zeitraum von 90 Minuten wurden ähnliche Vergessensraten für den impliziten Wortstammergänzungstest sowie für die explizite Form des

Wortstammergänzungstests gefunden. Ein weiteres Beispiel ist das erste Experiment von McBride, Doshier und Gage (2001), in dem für Wortfragmentergänzungstests mit impliziten und expliziten Anweisungen ähnliche Vergessensraten über einen Zeitraum von 45 Minuten gefunden wurden.

1.2.2.7 Lebensalter

Light et al. (1986) verglichen junge und ältere Personen in dem impliziten Wortfragmentergänzungstest und dem expliziten Worterkennungstest miteinander. Während ältere Personen schlechter bei der Erkognition von Wörtern abschnitten, gab es in dem impliziten Wortfragmentergänzungstest keine Unterschiede zwischen jungen und älteren Personen. In einer weiteren Studie verglichen Light und Singh (1987) ältere und junge Personen in verschiedenen impliziten und expliziten Gedächtnistests miteinander. Im impliziten Testteil waren Wortstämme zu ergänzen und eine perzeptuelle Identifikationsaufgabe zu bewältigen. Im expliziten Testteil war eine Aufgabe zum freien Abruf von Wörtern, die explizite Form des Wortstammergänzungstests und ein Worterkennungstest zu bewältigen. In den expliziten Tests schnitten ältere Personen im Vergleich zu den jungen Personen schlechter ab. Bei den impliziten Tests hingegen war kein Unterschied zu finden. Java (1992) untersuchte den Einfluss des Lebensalters auf die Höhe des Primings in einer impliziten Aufgabe zur Lösung von Anagrammen. Im Vergleich zu einem expliziten Erkognitionstest, in dem ältere Personen schlechter abschnitten als junge Personen, wurde kein Altersunterschied im Priming des impliziten Tests gefunden. Mitchell und Bruss (2003) verglichen junge und ältere Personen in verschiedenen impliziten und expliziten Gedächtnistests miteinander. Die Autoren fanden keine Unterschiede zwischen den jungen und den älteren Personen im Priming der impliziten Tests. In den expliziten Tests jedoch schnitten die jungen Personen besser ab als die älteren Personen. Fleischman et al. (2004) führten an einer größeren Gruppe älterer Personen eine Längsschnittstudie über eine Dauer von vier Jahren durch. Die Autoren fanden einen Abfall expliziter Gedächtnisleistungen, während sich die Höhe des Primings impliziter Gedächtnistests nicht verringerte. Einige Studien konnten jedoch keine Dissoziation zwischen impliziten und expliziten Gedächtnistests – als Funktion des Lebensalters – nachweisen (z.B. Chiarello & Hoyer, 1988; Davis et al., 1990; Drury, Kinsella & Ong, 2000; Maki, Zonderman & Weingartner, 1999).

1.2.2.8 Stimulusmodalität und Präsentationsformat

Graf et al. (1985) fanden in ihrem ersten Experiment, dass sich das Priming in dem impliziten Wortstammergänzungstest durch Änderung der Präsentationsmodalität von Stimuli reduziert. Die Autoren boten dazu Personen in der Lernphase ihres ersten Experimentes Wörter visuell und auditorisch dar. In einer anschließenden Testphase bekamen die Personen einen impliziten Wortstammergänzungstest und einen expliziten Test zum freien Abruf von Wörtern vorgelegt. Wenn ein Modalitätswechsel erfolgte – von auditorischer Präsentation in der Lernphase zu visueller Präsentation in der Testphase –, reduzierte sich die Höhe des Primings verglichen mit dem Fall, wenn kein Modalitätswechsel stattfand. Die Leistungen im expliziten Test dagegen wurden nicht durch den Modalitätswechsel beeinflusst. Weldon und Roediger (1987) zeigten Personen in der Lernphase ihres ersten Experimentes Bilder und Wörter. In der Testphase wurde entweder ein expliziter Test zum freien Abruf der Stimuli oder ein impliziter Wortfragmentergänzungstest vorgelegt. Der explizite Test zeigte den typischen Bildüberlegenheitseffekt. Die Leistung beim freien Abruf von Bildern war besser als die Leistung beim freien Abruf von Wörtern. Im impliziten Test produzierten Wörter jedoch mehr Priming als Bilder. Die Änderung des Präsentationsformates zwischen Lernphase und Testphase, d.h. wenn bildhafte Äquivalente von Wörtern in der Lernphase präsentiert wurden, reduzierte das Priming im Wortfragmentergänzungstest. Änderte sich das Präsentationsformat in dem impliziten Wortfragmentergänzungstest jedoch nicht zwischen Lernphase und Testphase, d.h. wenn in beiden Teilen des Experimentes Wörter oder Teile von Wörtern als Stimuli dienten, war die Höhe des Primings am größten. Ein Wechsel der Stimulusmodalität zwischen Lernphase und Testphase sorgt nicht immer für das Verschwinden des Repetition Priming, wie z.B. die Studie von Light und LaVoie (1993) zeigt.

1.2.2.9 Dissoziationen zwischen impliziten Gedächtnistests

Blaxton (1989) bot Personen in ihrem ersten Experiment Wörter unter verschiedenen Bedingungen dar. Es gab eine Bedingung ohne Kontext (z.B. xxx-Köln), eine Bedingung mit Kontext (z.B. Parfüm-Köln) und eine Bedingung, in der Wörter mit Hilfe von Hinweisen generiert werden sollten (z.B. Parfüm-K___). In der Testphase sollte entweder der perzeptuelle implizite Wortfragmentergänzungstest bearbeitet werden, oder in einem konzeptuellen impliziten Test allgemeine Wissensfragen beantwortet

werden. Bei der Beantwortung allgemeiner Wissensfragen entstand das größte Priming, wenn die Wörter in der Lernphase selbst generiert wurden. Im Gegensatz dazu war das Priming in dem perzeptuellen Wortfragmentergänzungstest am größten, wenn Wörter in der Lernphase ohne Kontext dargeboten wurden. In einem zweiten Experiment fand Blaxton (1989) heraus, dass die Änderung der Präsentationsmodalität von Stimuli keinen Einfluss auf das Priming hatte, wenn der konzeptuelle implizite Test bearbeitet wurde. Im Gegensatz dazu war das Priming im perzeptuellen Wortfragmentergänzungstest größer, wenn Wörter visuell präsentiert wurden, als wenn sie auditiv präsentiert wurden. In ihrem dritten Experiment wies Blaxton (1989) einen Teil ihrer Versuchspersonen an, sich die Bedeutung von Wörtern bildlich vorzustellen, während die restlichen Personen nicht diese Anweisung erhielten. Das Priming in dem konzeptuellen impliziten Test profitierte am meisten davon, wenn Personen sich die Bedeutung von Wörtern bildlich vorstellten. In dem perzeptuellen impliziten Wortfragmentergänzungstest machte es jedoch keinen Unterschied, ob Personen sich die Bedeutung von Wörtern bildlich vorstellten oder ob sie das nicht taten. Im ersten Experiment von Srinivas und Roediger (1990) bekamen Personen entweder einzelne Wörter dargeboten oder sie sollten Wörter mit Hilfe von Hinweisen selber generieren. Die Wörter sollten anschließend laut ausgesprochen werden. Daraufhin wurden verschiedene perzeptuelle und konzeptuelle implizite Tests dargeboten. In dem perzeptuellen Wortstammergänzungstest beispielsweise war das Priming größer für Personen, die Wörter einfach nur dargeboten bekommen haben, im Vergleich zu Personen, die Wörter mit Hilfe von Hinweisen selber generiert haben. In den konzeptuellen Tests verhielt es sich entgegengesetzt dazu. In ihrem ersten Experiment variierten Mitchell und Bruss (2003) das Präsentationsformat ihrer Stimuli zwischen Lernphase und Testphase. In einer Lernphase wurde eine gemischte Liste mit Wörtern und Bildern gezeigt. Die Wörter sollten nachgesprochen werden, die Bilder sollten benannt werden. In einer späteren Testphase wurden verschiedene konzeptuelle und perzeptuelle implizite Tests vorgelegt. Bei dem Wortfragmentergänzungstest, einem perzeptuellen impliziten Test, wurde Priming nur für Wörter – nicht jedoch für Bilder – gefunden. Bei dem Bildfragmentidentifikationstest – einem weiteren perzeptuellen impliziten Test –, wurde Priming nur für Bilder, nicht jedoch für Wörter gefunden. In einer Aufgabe zur Generierung von Exemplaren für Kategorien – einem konzeptuellen impliziten Test –, wurde Priming für Bilder und Wörter gefunden. Konzeptuelle implizite Tests erzeugten Priming, unab-

hängig vom Wechsel des Stimulusformates zwischen Lernphase und Testphase. Für perzeptuelle implizite Tests jedoch – wie dem Wortfragmentergänzungstest und dem Bildfragmentidentifikationstest –, verschwand das Priming mit dem Wechsel des Stimulusformates zwischen Lernphase und Testphase.

Dissoziationen zwischen perzeptuellen impliziten Gedächtnistests sind ebenfalls bekannt. So boten Weldon und Roediger (1987) Personen in der Lernphase ihres vierten Experimentes eine Reihe von Wörtern und Bildern dar. In einer anschließenden Testphase wurden entweder Wortfragmente oder Bildfragmente gezeigt. Diese gehörten entweder zu Wörtern aus der Lernphase, Bildern aus der Lernphase oder zu Stimuli, die nicht in der Lernphase gezeigt wurden. Wörter produzierten beim Wortfragmentergänzungstest mehr Priming als die bedeutungsgleichen Bilder. Bei der Bildfragmentidentifikation produzierten Bilder mehr Priming als ihre bedeutungsäquivalenten Wörter. Die Änderung des Präsentationsformates zwischen Lernphase und Testphase wirkte sich in beiden perzeptuellen impliziten Gedächtnistests negativ auf die Testleistungen aus.

Selbst unterschiedliche Formen impliziten Lernens können untereinander dissoziiert werden. So zeigen Studien an Patienten mit der Alzheimer-Erkrankung vermindertes Priming, bei intaktem Erwerb von motorischen Fertigkeiten. Patienten mit der Huntington-Erkrankung jedoch wiesen intaktes Priming auf, bei gestörtem Erwerb von motorischen Fertigkeiten (z.B. Butters, Heindel & Salmon, 1990).

1.3 Erklärungsmodelle

In den folgenden Abschnitten werden die wichtigsten Erklärungsmodelle zum Priming impliziter Gedächtnistests und zur Dissoziation zwischen impliziten und expliziten Gedächtnistests vorgestellt. Darunter fallen die Aktivierungstheorien (z.B. Graf & Mandler, 1984), die Theorien multipler Gedächtnissysteme (z.B. Tulving & Schacter, 1990) und die Theorien unterschiedlicher Gedächtnisprozesse (z.B. Roediger, 1990). Da die experimentellen Befunde zu den genannten Phänomenen sehr vielseitig sind, überrascht es nicht, dass keine der hier vorgestellten Erklärungsmodelle allen Daten gerecht werden kann.

1.3.1 Aktivierung und Elaboration

Eine Theorie zur Erklärung von Dissoziationen zwischen impliziten und expliziten Gedächtnistests ist die Aktivierungstheorie (Graf & Mandler, 1984). Nach Graf und Mandler (1984) ist die mentale Repräsentation oder das Schema eines Ereignisses aus perzeptuellen Bestandteilen, semantischen Bestandteilen und deren Beziehung zueinander aufgebaut. Die Verarbeitung eines solchen Ereignisses – z.B. der Darbietung eines Wortes – soll dieses Schema aktivieren und stärken. Impliziten Gedächtnisleistungen liegt – nach der Aktivierungstheorie von Graf und Mandler (1984) – eine zeitlich begrenzte Aktivierung von bereits bestehenden Gedächtnisrepräsentationen zugrunde. Eine solche Aktivierung soll ihrer Natur nach automatisch erfolgen, ohne dass die Beteiligung elaborativer Prozesse erforderlich wäre. Für explizite Gedächtnisleistungen jedoch sollen elaborative Prozesse notwendig sein, um neue episodische Gedächtnisspuren auszubilden. So kommt es bei einer Schemaaktivierung – im Kontext anderer mentaler Ereignisse – zur Ausbildung neuer Beziehungen zu diesen Ereignissen und zu einer Reaktivierung bereits bestehender Beziehungen (Graf & Mandler, 1984). Durch die Beziehungen von Ereignissen, welche durch Elaborationsprozesse ausgebildet werden, werden entsprechende Schemata besser explizit abgerufen, weil neue Gedächtnisspuren für den Abruf entstehen und alte Gedächtnisspuren gestärkt werden (Graf & Mandler, 1984).

Nach Schacter (1987) wird eine in der Lernphase eines impliziten Tests aktivierte Gedächtnisrepräsentation, die inhaltlich nicht modifiziert wird, durch einen entsprechenden Hinweisreiz in der Testphase (z.B. ein Wortstamm oder ein Wortfragment) leichter verfügbar, als eine nicht aktivierte Gedächtnisrepräsentation. Sie enthält jedoch keine kontextuellen Informationen über das Auftreten des Stimulus als Teil einer vorherigen Episode und hilft somit nicht bei der expliziten Erinnerung an die Episode (Schacter, 1987). Nach der Aktivierungstheorie von Graf und Mandler (1984) entsteht das Repetition Priming impliziter Tests alleine durch eine Aktivierung von bereits im Gedächtnis bestehenden Repräsentationen bei der Enkodierung, während die Leistung in expliziten Tests zusätzlich von einer elaborativen Enkodierung kontextueller Informationen in der Lernphase von Experimenten abhängt. Wenn jedoch keine kontextuelle Information in der Lernphase eingespeichert wurde, wird die Leistung in der Testphase expliziter Tests schlechter, während das Priming impliziter Tests davon unbeeinflusst sein kann (Graf & Mandler, 1984). Die Verarbeitung eines

Wortes beispielsweise, führt nach Graf und Mandler (1984) zu einer Aktivierung und Stärkung seiner mentalen Repräsentation. Dies erhöht seine Verfügbarkeit und damit ebenfalls die Wahrscheinlichkeit, es aus einem später folgenden Hinweisreiz reproduzieren zu können (Graf & Mandler, 1984). Um jedoch Beziehungen zwischen einem Wort und dem Kontext seines Auftretens herstellen zu können, werden zusätzlich dazu elaborative Prozesse benötigt. Integrative Verarbeitung aktiviert Schemata und macht diese verfügbarer, während elaborative Verarbeitung die Verfügbarkeit und die Abrufbarkeit erhöht (Graf & Mandler, 1984).

Beim impliziten Wortstammergänzungstest beispielsweise werden die Wortanfänge von Wörtern aus der Lernphase und von neuen Wörtern präsentiert, wobei Personen mit dem ersten Wort antworten sollen, welches ihnen dazu einfällt. Die Präsentation eines Wortstammes führt zu einer Schemaaktivierung aller Wörter, die dazu passend ergänzt werden können. Weil aber das Schema des Zielwortes aus der Lernphase bereits aktiviert ist, kommt es im Vergleich zu den konkurrierenden Wortschemata – die erst jetzt aktiviert werden – zu einer schnelleren Ausbreitung der Aktivierung. Bei einem expliziten Test jedoch – wie z.B. dem Abruf von Wörtern mit Abrufhinweis – wird durch die Testanweisungen der Bezug zur Lernepisode hergestellt. Die Wahrscheinlichkeit, mit der Wörter aus der Lernphase abgerufen werden können, hängt von den verfügbaren Gedächtnisspuren zu den Zielwörtern ab. Die Ausbildung dieser Gedächtnisspuren wiederum hängt von den Elaborationsprozessen in der Lernphase ab. Dissoziationen zwischen impliziten und expliziten Gedächtnistests kommen nach Graf und Mandler (1984) deshalb zustande, weil Wörter durch die unterschiedlichen Anforderungen der Tests im impliziten Wortstammergänzungstest verfügbar sein können, ohne jedoch in der expliziten Form des Wortstammergänzungstests abrufbar sein zu müssen.

1.3.2 Theorien multipler Gedächtnissysteme

Die Erschaffung eines Klassifikationssystems – beispielsweise für verschiedene Formen von Gedächtnis – beinhaltet die Unterteilung der Mitglieder dieses Systems in verschiedene Kategorien. Eine solche Unterteilung basiert immer auf bestimmten Unterscheidungskriterien. Im Falle einer Unterscheidung in implizites und explizites Gedächtnis, könnte man eine Einteilung aufgrund neuroanatomischer Kriterien vornehmen – wie es Theorien multipler Gedächtnissysteme tun –, die in diesem Ab-

schnitt besprochen werden. Diese Theorien beziehen ihre Daten vor allem aus Studien mit Amnestikern. Eine weitere Möglichkeit ist es, eine Einteilung auf der Grundlage unterschiedlicher Gedächtnisprozesse vorzunehmen, wie es Theorien tun, die ihren Ursprung in der kognitiven Psychologie haben. Diese Theorien werden im nächsten Abschnitt besprochen.

Grundlage der Unterscheidung in ein implizites (prozedurales) und in ein explizites (deklaratives) Gedächtnissystem ist die Vorstellung neuroanatomisch distinkter Gebiete, die hypothetisch für Leistungen in impliziten oder in expliziten Gedächtnistests verantwortlich gemacht werden. Theorien multipler Gedächtnissysteme gehen davon aus, dass explizite und implizite Gedächtnisinhalte in unterschiedlichen Bereichen des Gehirns verarbeitet und gespeichert werden. Dissoziationen zwischen impliziten und expliziten Gedächtnistests entstehen aufgrund der unterschiedlichen Eigenschaften der zugrunde liegenden Gedächtnissysteme (Schacter, 1987). Mit dem Nachweis stochastischer Unabhängigkeit impliziter und expliziter Gedächtnistests schloss man auf die funktionale Unabhängigkeit postulierter zugrunde liegender Gedächtnissysteme (Jacoby & Witherspoon, 1982; Tulving et al., 1982). Aufgrund von Dissoziationen zwischen impliziten und expliziten Gedächtnistests, beim Vergleich von Amnestikern mit gesunden Personen und aufgrund von Dissoziationen, die als Funktion experimenteller Manipulationen weiterer Variablen bei gesunden Populationen gefunden worden sind, postulierten verschiedene Autoren die Existenz mehrerer Gedächtnissysteme (z.B. Schacter, 1987; Schacter & Tulving, 1994; Squire, 1992). Beispielhaft ist die Unterscheidung in deklaratives und prozedurales Gedächtnissystem (Squire, 1992; Squire, Knowlton & Musen, 1993), sowie die Unterscheidung in episodisches und semantisches Gedächtnissystem zu nennen. Nach Squire und Cohen (1984) ist bewusster Abruf verbalisierbaren Wissens in expliziten Tests auf das deklarative Gedächtnissystem zurückzuführen. Das Priming impliziter Tests, motorische Fertigkeiten und andere Formen von implizitem Gedächtnis, die ohne bewussten Abruf auskommen, beruhen nach Squire und Cohen (1984) auf dem prozeduralen Gedächtnissystem. Weil die beiden Gedächtnissysteme weitestgehend voneinander unabhängig sind, sind Dissoziationen zu erwarten (Roediger, 1990). Das episodische Gedächtnissystem wird für die explizite Erinnerung an zurückliegende Ereignisse zuständig angesehen, während das semantische Gedächtnissystem zuständig für Leistungen in impliziten Tests sein soll (Schacter, 1987).

Man vermutet bei Amnestikern eine selektive Schädigung desjenigen Gedächtnissystems, welches für den bewussten Abruf von Informationen zuständig ist, während Gedächtnissysteme für andere Arten des Lernens relativ intakt bleiben (Roediger, 1990). Amnestiker haben Schäden im Hippocampus und den umliegenden Strukturen des medialen Temporallappens, die mit der korrekten Funktion des expliziten Gedächtnisses in Verbindung gebracht werden (Squire, 1992). Schäden an diesen Strukturen werden mit Problemen bei der Speicherung von *neuen* episodischen und semantischen Gedächtnisinhalten assoziiert, während beispielsweise das Gedächtnis für *frühere* Ereignisse davon unbeeinflusst sein kann. Aus diesem Grund wird angenommen, dass der Hippocampus nur eine Zwischenstation bei der Speicherung von Inhalten des episodischen und semantischen Langzeitgedächtnisses ist, der vor allem die Funktion der «Bindung» verschiedener Aspekte von Gedächtnisspuren zukommt (Dew & Cabeza, 2011).

Eine spezifizierte Version der Theorie multipler Gedächtnissysteme (Tulving & Schacter, 1990) postuliert ein präsemantisch arbeitendes perzeptuelles Repräsentationssystem, auf dem das perzeptuelle Priming impliziter Gedächtnistests beruhen soll. Das perzeptuelle Repräsentationssystem besteht nach Tulving und Schacter (1990) aus verschiedenen interagierenden und spezifischen Subsystemen, die in bestimmten Bereichen des Gehirns lokalisiert sein sollen und Informationen über Form und Struktur – jedoch nicht über die Bedeutung von Wörtern und Objekten – verarbeiten sollen. Die Subsysteme des perzeptuellen Repräsentationssystems sollen Informationen implizit verarbeiten und somit nicht willentlich zugänglich sein. Das perzeptuelle Repräsentationssystem soll unabhängig von dem episodischen oder deklarativen Gedächtnis funktionieren, welches die Basis für explizite Inhalte darstellen soll. Tulving und Schacter (1990) unterscheiden drei Subsysteme des perzeptuellen Repräsentationssystems. Das visuelle Repräsentationssystem für Wörter soll Informationen zur visuellen Wortform speichern. Das visuelle Repräsentationssystem für Objekte soll Informationen zu perzeptuellen Eigenschaften von Objekten speichern. Das auditorische Repräsentationssystem für Wörter soll Informationen zur auditorischen Wortform speichern. Alle Subsysteme seien präsemantisch und sollen auf Gehirnstrukturen basieren, die perzeptuelle Analysen durchführen. Demnach soll die inhaltliche Bedeutung von Wörtern und Objekten bei der Verarbeitung nicht berücksichtigt werden. Nach Schacter (1992) kommen Hinweise auf Systeme, die auf einer präsemantischen Ebene arbeiten, aus Studien an Personen, die Wörter

lesen können und damit Zugang zur Repräsentation der visuellen Wortform haben, jedoch kein Verständnis für die inhaltliche Bedeutung dieser Wörter zeigen (z.B. Schwartz, Saffran & Marin, 1980). Bildgebende Verfahren zeigen, dass das System für die visuelle Wortform im extrastriären okzipitalen Kortex liegen könnte (z.B. Petersen, Fox, Posner, Mintum & Raichle, 1989). Weitere Studien zur präsemantischen Natur der perzeptuellen Repräsentationssysteme beschreiben Personen, denen es Schwierigkeiten bereitet, semantische Informationen zu visuellen Objekten abzurufen, bei denen jedoch der Zugang zu perzeptuellen Eigenschaften des gleichen Objektes ungestört ist (z.B. Sartori & Job, 1988). Nach Schacter (1992) spiegeln die präsemantischen Repräsentationssysteme den unbewussten Ausdruck des Gedächtnisses für Vorerfahrungen wider. Sie sollen perzeptuelle Priming Phänomene erklären, wie sie zum Beispiel bei dem impliziten Wortstammergänzungstest auftreten. Perzeptuelles Priming kommt nach Schacter (1992) dann zustande, wenn die perzeptuellen Eigenschaften von Stimuli der Lernphase, die in den Subsystemen des perzeptuellen Repräsentationssystems gespeichert sein sollen, mit den perzeptuellen Eigenschaften der Hinweisreize übereinstimmen, die in der Testphase präsentiert werden. Nach Tulving und Schacter (1990) jedoch soll das perzeptuelle Repräsentationssystem nicht an den Gedächtnisleistungen expliziter Tests und konzeptueller impliziter Tests beteiligt sein.

1.3.3 Prozesstheorien

Nach den Prozesstheorien (z.B. Roediger, Weldon & Challis, 1989) unterscheiden sich implizites und explizites Gedächtnis durch die ihnen zugrunde liegenden Enkodierungsprozesse und Abrufprozesse. Dissoziationen zwischen impliziten und expliziten Gedächtnistests spiegeln Unterschiede in den kognitiven Prozessen wider, die den Tests zugrunde liegen (Roediger, 1990). Eine wichtige Grundannahme dieser Theorien sei die Vorstellung, dass Gedächtnisleistungen davon abhängen, wie gut kognitive Prozesse zwischen Lernphase und Testphase übereinstimmen. Roediger (1990) nennt als Basis dafür das «Prinzip der Enkodierungsspezifität» von Tulving und Thomson (1973) und den Ansatz des «Transfer-appropriate Processing» (Morris, Bransford & Franks, 1977). Demnach seien die Leistungen in impliziten und expliziten Gedächtnistests dann am besten, wenn die kognitiven Prozesse bei der Enkodierung und beim Abruf übereinstimmen. Weiterhin nimmt Roediger (1990) an, dass

explizite und implizite Gedächtnistests typischerweise unterschiedliche Abrufprozesse erfordern und somit auch von unterschiedlichen Enkodierungsstrategien profitieren. Während die Leistung in vielen expliziten Gedächtnistests auf der Übereinstimmung von semantischen Prozessen zwischen Lernphase und Testphase beruht und von einer elaborativer Verarbeitung profitiert, beruht die Leistung in vielen impliziten Tests auf der Übereinstimmung perzeptueller Prozesse zwischen Lernphase und Testphase und profitiert eher von einer oberflächlichen Verarbeitung (Roediger, 1990). Wenn Stimuli in der Lernphase eines Experimentes perzeptuell verarbeitet werden, d.h. wenn vor allem physische Eigenschaften von Stimuli enkodiert werden, dann soll das die Leistung in *perzeptuellen* Gedächtnistests verbessern. Im Gegensatz dazu soll eine semantische Verarbeitung von Stimuli, bei der die inhaltliche Bedeutung der Stimuli enkodiert wird, die Leistung in *konzeptuellen* Gedächtnistests verbessern. Nach Roediger (1990) sind die Gedächtnisleistungen in expliziten Tests – im Gegensatz zu den Gedächtnisleistungen in (perzeptuellen) impliziten Tests – relativ beständig gegenüber der Änderung oberflächlicher Stimuluseigenschaften (wie Präsentationsformat und Stimulusmodalität) zwischen Lernphase und Testphase, reagieren jedoch empfindlich auf konzeptuelle Änderungen. Explizite Tests, bei denen die semantische Verarbeitung im Vordergrund steht, beruhen nach Roediger (1990) auf konzeptuellen Prozessen. Sie sollen den Abruf von semantischen oder kontextuellen Informationen erfordern, die mit den Stimuli in der Lernphase enkodiert wurden. Probanden müssen dazu – beispielsweise beim freien Abruf von Wörtern – die Lernepisode mental rekonstruieren (Richardson-Klavehn & Bjork, 1988). Dagegen beruhen viele implizite Tests, bei denen die Verarbeitung von oberflächlichen Stimuluseigenschaften im Vordergrund steht, auf perzeptuellen Prozessen (Roediger, 1990) und erfordern den Abruf von physischen Stimuluseigenschaften. Die meisten expliziten Gedächtnistests beruhen auf konzeptuellen Prozessen. Ein Teil impliziter Gedächtnistests beruht auf perzeptuellen Prozessen, während ein anderer Teil impliziter Gedächtnistests auf konzeptuellen Prozessen beruht.

1.4 Erklärungsmodelle im Licht der empirischen Befundlage

Die beschriebenen Theorien haben unterschiedliche Erklärungen für die Entstehung von Priming und die Entstehung von Dissoziationen zwischen impliziten und expli-

ziten Gedächtnistests. Weiterhin machen sie bestimmte Voraussagen, die empirisch überprüft werden können. Je häufiger Voraussagen empirisch bestätigt werden können, desto mehr spricht dies für die Gültigkeit der entsprechenden Theorie. Am besten kann zwischen der Gültigkeit von Theorien entschieden werden, wenn diese sich widersprechende Voraussagen machen und die empirischen Befunde dazu eindeutig und replizierbar sind.

1.4.1 Aktivierungstheorien und die Befundlage

Erklärungsmodelle, die von der Aktivierung bereits bestehender Gedächtnisrepräsentationen als Grundlage impliziten Gedächtnisses ausgehen, können Dissoziationen zwischen impliziten und expliziten Gedächtnistests gut erklären, bei denen die Manipulation experimenteller Variablen einen großen Einfluss auf explizite Tests hat, jedoch nicht auf implizite Tests (Richardson-Klavehn & Bjork, 1988). Aktivierungstheorien können Priming von bereits bestehenden Gedächtnisrepräsentationen erklären, wie z.B. Priming von linguistischen Einheiten. Bei bekannten Wörtern, eng assoziativ verknüpften Konzepten und bei bekannten Objekten wird angenommen, dass für sie bereits Gedächtnisrepräsentationen bestehen. Alles was für Priming impliziter Tests nötig ist, ist die Aktivierung einer bereits bestehenden Gedächtnisrepräsentation in der Lernphase von Experimenten (Richardson-Klavehn & Bjork, 1988). Eine Aktivierung von neuen Stimuli (z.B. von Pseudowörtern und neuen Assoziationen) ist jedoch nach den Aktivierungstheorien definitionsgemäß nicht möglich, weil für sie keine Gedächtnisrepräsentationen bestehen. Somit sind Befunde von Studien an Amnestikern und gesunden Personen, die Priming für Stimuli mit vorbestehenden Gedächtnisrepräsentationen zeigen, sowie fehlendes Priming für Pseudowörter und weitere Stimuli ohne vorbestehende Gedächtnisrepräsentation (z.B. Diamond & Rozin, 1984; Forster & Davis, 1984; Kirsner & Smith, 1974; Schacter, 1985), konsistent mit Voraussagen von Aktivierungstheorien. Befunde von Studien jedoch, bei denen Amnestiker und gesunde Personen implizites Gedächtnis für Informationen ohne vorbestehende Gedächtnisrepräsentationen zeigen (z.B. Graf & Schacter, 1985; Feustel, Shiffrin & Salasoo, 1983; Salasoo, Shiffrin & Feustel, 1985), sind mit Aktivierungstheorien nicht vereinbar. Die genannten Studien und weitere Studien, die das Priming von neuen Stimuli ohne bestehende Gedächtnisrepräsentationen untersuchten, verwendeten dafür Pseudowörter, assoziativ unverbundene

Wortpaare oder unbekannte Objekte als Material. Aufgrund der Neuartigkeit des präsentierten Materials ist anzunehmen, dass es dafür keine vorbestehenden Gedächtnisrepräsentationen gab. Priming für Pseudowörter bei klinisch gesunden Personen erklären Aktivierungstheorien mit dem Einfluss von explizitem Gedächtnis. Sie können jedoch nicht Befunde von Studien an Amnestikern erklären, bei denen implizites Gedächtnis für neue Informationen besteht, wie beim Priming von Pseudowörtern, der klassischen Konditionierung und dem Erwerb von motorischen Fertigkeiten (Richardson-Klavehn & Bjork, 1988).

Da Aktivierung und Konsolidierung von Schemata Prozesse sind, die bei jeder Stimuluspräsentation automatisch ablaufen und nach Aktivierungstheorien die einzigen an impliziten Gedächtnisleistungen beteiligten Prozesse sind, sollte die Manipulation der Verarbeitungstiefe von Stimuli keinen Einfluss auf die Leistung impliziter Gedächtnistests haben. Die Manipulation der Verarbeitungstiefe von Stimuli in der Lernphase sollte sich im Gegensatz dazu in den Leistungen expliziter Gedächtnistests bemerkbar machen, da zu den automatisch ablaufenden Prozessen elaborative Prozesse hinzukommen, die dazu dienen, Beziehungen zwischen Stimuli und dem Kontext ihres Auftretens herzustellen. Dies wird erleichtert, wenn eine tiefe Verarbeitung von Stimuli in der Lernphase zur Bildung neuer Gedächtnisspuren beitragen konnte. Befunde von Studien, bei denen die Höhe des Primings vorbestehender Gedächtnisrepräsentationen nicht von elaborativer Verarbeitung in der Lernphase beeinflusst wird (z.B. Jacoby & Dallas, 1981), sind somit gut durch Aktivierungstheorien erklärbar.

Gut erklärbar sind ebenfalls Befunde von Studien an gesunden Probanden und an Amnestikern, die eine rasche Abnahme von Priming zeigen (z.B. Graf & Mandler, 1984; Shimamura & Squire, 1984). Problematisch für Aktivierungstheorien jedoch sind Befunde von Studien, bei denen eine Aktivierung von vorbestehenden Gedächtnisrepräsentationen über einen Zeitraum von Tagen oder Wochen anhält, obwohl man bei Priming von einem zeitlich sehr befristeten Phänomen ausgeht (z.B. Jacoby & Dallas, 1981; Sloman et al., 1988; Tulving et al., 1982; Tunney, 2003). Priming bei klinisch gesunden Probanden, welches für längere Zeit aufrechterhalten wird, erklären Aktivierungstheorien mit dem Einfluss von explizitem Gedächtnis (Richardson-Klavehn & Bjork, 1988). Explizite Gedächtnisleistungen nehmen jedoch recht schnell ab und sind kaum als Erklärung für solche Befunde geeignet. Zudem

gibt es auch bei Amnestikern Beispiele von lang anhaltendem Priming (z.B. Cohen & Squire, 1980). Da man jedoch bei Amnestikern Kontaminationen durch explizite Gedächtnisstrategien ausschließen kann, können diese keine Erklärung für länger andauerndes Priming sein.

Ebenfalls problematisch für Theorien, die von der Aktivierung bereits bestehender Gedächtnisrepräsentationen ausgehen, ist der Befund, dass Priming abgeschwächt wird, wenn sich perzeptuelle Eigenschaften von Stimuli zwischen Lernphase und Testphase ändern (z.B. Jacoby & Dallas, 1981). Dies ist beispielsweise beim Wechsel der Modalität der Darbietung von Stimuli der Fall oder beim Wechsel des Präsentationsformates von Stimuli.

1.4.2 Theorien multipler Gedächtnissysteme und die Befundlage

Nach Schacter (1987) können Theorien multipler Gedächtnissysteme Befunde von Studien an Amnestikern gut erklären, die uneingeschränkten Erwerb von motorischen Fertigkeiten zeigen, bei gleichzeitigem Verlust der Fähigkeit zum bewussten Abruf von früheren Lernepisoden (z.B. Corkin, 1968). Der Erwerb von motorischen Fertigkeiten beruht auf dem prozeduralen Gedächtnissystem, welches bei Amnestikern verschont geblieben ist und das von dem geschädigten deklarativen Gedächtnissystem abgegrenzt werden kann (Schacter, 1987). Der Befund jedoch, dass Patienten mit der Alzheimer Erkrankung abgeschwächtes Priming, jedoch intakten Erwerb von motorischen Fertigkeiten zeigen, während Patienten mit der Huntington Erkrankung intaktes Priming, jedoch gestörten Erwerb von motorischen Fertigkeiten zeigen (z.B. Butters et al., 1990), steht im Widerspruch zu Theorien multipler Gedächtnissysteme (Schacter, 1987). Da Priming und der Erwerb von motorischen Fertigkeiten als Teil des gleichen – prozeduralen – Gedächtnissystems angesehen werden, dürften sie nicht dissoziiert sein, ohne eine weitere Unterteilung des prozeduralen Gedächtnissystems in weitere unabhängige Subsysteme vorzunehmen. Eine Schwäche dieser Theorien ist deshalb, dass für jede neue Dissoziation weitere Gedächtnissysteme postuliert werden müssen. Die gleichen Probleme ergeben sich bei der Interpretation von Befunden aus Studien, die Dissoziationen zwischen impliziten Gedächtnistests aufzeigen (z.B. Witherspoon & Moscovitch, 1989). Dissoziationen zwischen konzeptuellen und perzeptuellen impliziten Gedächtnistests – wie sie beispielsweise Srinivas und Roediger (1990) oder Blaxton (1989) gefunden haben – brin-

gen die Theorie multipler Gedächtnissysteme in Erklärungsnot, weil sie mit dem prozeduralen und dem deklarativen Gedächtnis lediglich zwei separate Gedächtnissysteme postuliert hat. Um jedoch Dissoziationen zwischen konzeptuellen impliziten Gedächtnistests und perzeptuellen impliziten Gedächtnistests zu erklären, müsste das implizite Gedächtnissystem weiter in Subsysteme unterteilt werden.

Tulving und Schacter (1990) postulierten ein präsemantisch arbeitendes perzeptuelles Repräsentationssystem, welches für Priming in perzeptuellen impliziten Gedächtnistests zuständig sein soll. Nach Schacter (1992) ist Priming für Wörter und Wortpaare bei Amnestikern (z.B. Shimamura & Squire, 1984) konsistent mit der Voraussetzung, dass visuelles Priming von Wörtern durch ein Subsystem des perzeptuellen Repräsentationssystems mediiert wird, welches in posterioren kortikalen Regionen liegt. Schädigungen bei Amnestikern können typischerweise im Hippocampus und angrenzenden Strukturen beobachtet werden, während posteriore Anteile des Kortex intakt bleiben (Schacter, 1992). Befunde von Studien, bei denen Amnestiker fehlendes Priming für Pseudowörter zeigen, sind nicht konsistent mit Theorien multipler Gedächtnissysteme. Priming sollte gleichermaßen für bereits bestehende Gedächtnisrepräsentationen sowie für neue Information auftreten (Schacter, 1987). Falls visuelles Priming von verbalen Stimuli bei Amnestikern durch ein funktionierendes Wortform System – einem Subsystem des perzeptuellen Repräsentationssystems – mediiert wäre, sollten Amnestiker Priming nicht nur für bekannte Wörter, sondern auch für Pseudowörter zeigen (Schacter, 1992). Das Wortform System würde in diesem Fall für Pseudowörter neue perzeptuelle Repräsentationen schaffen (Schacter, 1992). Für klinisch gesunde Personen konnte neben Priming für bekannte Wörter auch Priming für Pseudowörter gefunden werden (z.B. Feustel et al., 1983), wie es eine Theorie fordert, die präsemantische Repräsentationssysteme annimmt.

Ein weiteres Indiz dafür, dass perzeptuelles Priming von Wörtern auf einem präsemantisch arbeitenden Wortform System beruht, liefern nach Schacter (1992) Studien, in denen die Manipulation der Verarbeitungstiefe von Stimuli in der Lernphase die Leistung in perzeptuellen impliziten Gedächtnistests unbeeinflusst lässt, die Leistung in expliziten Gedächtnistests und in konzeptuellen impliziten Gedächtnistests jedoch beeinflusst (z.B. Graf & Mandler, 1984; Jacoby & Dallas, 1981). Weil das perzeptuelle Repräsentationssystem auf einer präsemantischen Ebene tätig ist, werden perzeptuelle implizite Tests nicht von semantischen Enkodierungsbedingungen be-

einflusst. Das perzeptuelle Repräsentationssystem ist jedoch nicht an den Gedächtnisleistungen in expliziten Tests und an den Gedächtnisleistungen in konzeptuellen impliziten Tests beteiligt (Tulving & Schacter, 1990).

Die Annahme stochastischer Unabhängigkeit impliziter und expliziter Gedächtnistests und damit die funktionelle Unabhängigkeit ihrer zugrunde liegenden Gedächtnissysteme, wird von einigen Kritikern stark angezweifelt (z.B. Shimamura, 1985). Weiterhin wird argumentiert, dass stochastische Unabhängigkeit zwischen impliziten und expliziten Gedächtnistests – sollte sie tatsächlich vorhanden sein – nicht notwendigerweise ein Beweis für die funktionelle Unabhängigkeit der hypothetisch zugrunde liegenden Gedächtnissysteme sein muss.

1.4.3 Prozesstheorien und die Befundlage

Prozesstheorien sind bezüglich ihren Stärken und Schwächen ein Spiegelbild zu Aktivierungstheorien (Schacter, 1987). Sie können Dissoziationen zwischen expliziten Gedächtnistests und Dissoziationen zwischen impliziten Gedächtnistests erklären, wenn ein erster Test konzeptuell und der zweite Test perzeptuell ist (z.B. Roediger, Srinivas & Weldon, 1989). Blaxton (1989) zeigte in ihrem ersten Experiment, dass nicht die Unterscheidung in deklaratives und prozedurales Gedächtnissystem Prädiktor für Dissoziationen zwischen impliziten und expliziten Gedächtnistests ist. Entscheidend ist es, ob Informationen konzeptuell oder perzeptuell verarbeitet werden. Neben einem typischen konzeptuellen expliziten Test und einem typischen perzeptuellen impliziten Test, nahm Blaxton (1989) ebenfalls einen perzeptuellen expliziten Test und einen konzeptuellen impliziten Test in die Versuchsanordnung mit auf. Dissoziationen traten zwischen dem konzeptuellen impliziten Test und dem perzeptuellen impliziten Test auf, sowie zwischen dem konzeptuellen expliziten Test und dem perzeptuellen expliziten Test. Das ist konsistent mit Prozesstheorien, widerspricht jedoch Theorien multipler Gedächtnissysteme.

Perzeptuelles Priming bei Amnestikern erklären Prozesstheorien durch verschonte perzeptuelle Prozesse. Konzeptuelle Prozesse, die für explizite Gedächtnisleistungen verantwortlich gemacht werden, sollen dagegen nicht mehr richtig funktionieren. Problematisch für diese Interpretation sind jedoch Befunde von Studien, in denen Amnestiker Priming für konzeptuelle implizite Tests zeigen (z.B. Shimamura, 1986).

1.5 Alternative Erklärungen für die Befunde

Dissoziationen zwischen impliziten und expliziten Gedächtnistests in der großen Variationsbreite ihres Auftretens angemessen zu interpretieren, ist eine schwierige Aufgabe. Keine der vorgestellten Theorien ist in der Lage, der Gesamtheit der Datenergebnisse gerecht zu werden. Es gibt jedoch auch einige methodische Ansätze zur Interpretation der Befunde. Diese Ansätze sind nicht als Alternative oder als Konkurrenz zu den bereits bestehenden Theorien zu verstehen. Sie weisen eher auf methodische Probleme hin, die von der Testkonstruktion bis zur Testdurchführung reichen. Erst wenn diese Probleme im Vorfeld gelöst wurden ist es sinnvoll, Erklärungsmodelle an den Befunden zu evaluieren. Da die vorliegende Arbeit methodisch orientiert ist und die Beteiligung problematischer methodischer Faktoren an Dissoziationsbefunden angenommen wird, ist eine ausführliche Diskussion dieser alternativen Erklärungsansätze wichtig. An geeigneter Stelle werden Vorschläge gemacht, um den Einfluss dieser methodischen Faktoren auf die Befunde zu verringern. Zunächst soll die Problematik der Prozessreinheit von impliziten und expliziten Gedächtnistests angesprochen werden. Explizite Kontaminationen impliziter Gedächtnistests beispielsweise können dazu führen, dass Befunde von Assoziationen nicht eindeutig interpretierbar sind. Ein weiterer wichtiger Aspekt dieser Arbeit ist die Vergleichbarkeit von impliziten und expliziten Gedächtnistests. Nur wenn Vergleichbarkeit gewährleistet ist, können Befunde zweifelsfrei auf die Manipulation einer experimentellen Variablen zurückgeführt werden. Ist diese nicht gegeben, können Befunde auch auf Unterschiede in den verwendeten Tests zurückzuführen sein (Merikle & Reingold, 1991). Als letzter Aspekt können einfache Dissoziationen zwischen Tests auch das Resultat von Artefakten sein, wie sie beispielsweise durch geringe Reliabilität impliziter Gedächtnistests in Verbindung mit hoher Reliabilität expliziter Gedächtnistests entstehen können (Buchner & Wippich, 2000). Auf den Reliabilitätsaspekt soll besonders ausführlich eingegangen werden, da er zentraler Bestandteil der vorliegenden Arbeit war.

1.5.1 Explizite Kontaminationen impliziter Gedächtnistests

Ein Problem, das bei der Erfassung impliziter Gedächtnisinhalte auftreten kann, ist die Mitwirkung expliziter Gedächtnisinhalte an der Testleistung in impliziten Ge-

dächntistests. In der Literatur werden häufig Bedenken darüber geäußert, in die Erfassung impliziter Gedächtnisinhalte könnten bewusste Abrufstrategien mit einfließen (z.B. Kinoshita, 2001; McKone & Slee, 1997; Weldon & Jackson-Barrett, 1993). Versuchsteilnehmer könnten Bewusstsein über den Zusammenhang zwischen Lernphase und Testphase erlangen und dieses Wissen dazu nutzen, explizite Gedächtnisstrategien zum Abruf von Stimuli aus der Lernphase anzuwenden. Die Beteiligung dieser sogenannten «prozessunreinen» Mechanismen hätte eine Verzerrung impliziter Testleistung zur Folge, da implizite Gedächtnisinhalte mit expliziten Gedächtnisinhalten konfundiert wären. So argumentieren Mitchell und Bruss (2003), dass perzeptuelle implizite Gedächtnistests – aber auch konzeptuelle implizite Gedächtnistests – umso anfälliger für explizite Kontaminationen seien, je mehr Antwortzeit den Probanden in der Testphase von Experimenten gelassen werde. Eine Kontamination impliziter Gedächtnistests durch explizite Abrufstrategien, weil Probanden der Bezug zur Lernepisode bewusst wird, kann sich beispielsweise durch parallele Effekte bzw. Assoziationen zeigen. Von parallelen Effekten spricht man in diesem Zusammenhang, wenn die Manipulation einer experimentellen Variablen (z.B. Verarbeitungstiefe) den gleichen Effekt auf implizite und auf explizite Gedächtnistests hat. Wenn man den Vergleich junger und älterer Personen heranzieht, dann könnte die Erfassung impliziter Gedächtnisleistungen junger Personen stärker von expliziten Kontaminationen betroffen sein, als dies bei älteren Personen der Fall wäre, bei denen die Nutzung expliziter Abrufstrategien bereits eingeschränkt ist. Die Leistung junger Personen in impliziten Gedächtnistests wäre durch den Einfluss bewusster Abrufstrategien erhöht. Eine tatsächlich bestehende Dissoziation könnte in diesem Fall verborgen bleiben. Wenn Datensätze von Probanden entfernt werden, die sich der Kontingenz zwischen Lernphase und Testphase bewusst sind, werden altersbedingte Primingunterschiede oft nivelliert (Mitchell & Bruss, 2003). Es ist jedoch zu bedenken, dass die Anwendung expliziter Abrufstrategien in impliziten Gedächtnistests nicht unbedingt damit verbunden sein muss, dass Personen Bewusstsein über den Zusammenhang zwischen Lernphase und Testphase erlangen. Aus diesem Grund ist die Beseitigung entsprechender Datensätze nur *ein* Schritt zur Lösung des Problems. Befunde von Studien, in denen parallele Effekte zwischen impliziten und expliziten Gedächtnistests gefunden wurden (z.B. Chiarello & Hoyer, 1988; Drury et al., 2000), werden so lange schwierig zu deuten sein, bis mehr über die Beteiligung expliziter Abrufstrategien an der Leistung in impliziten Tests bekannt sein wird.

Obwohl es sicher nicht gerechtfertigt wäre, zu behaupten, dass explizite Kontaminationen in allen Fällen Ursache für Primingwertunterschiede zwischen Gruppen sind, sollten dennoch Möglichkeiten genutzt werden, den Einfluss expliziter Abrufstrategien im Vorfeld eines Experimentes zu reduzieren. Im Nachhinein könnte man bestimmen, ob und in welchem Umfang sie zu impliziten Gedächtnisleistungen beigetragen haben. Maßnahmen zu treffen, die den Zusammenhang zwischen Lernphase und Testphase verbergen und Experimente nicht als Gedächtnistests zu präsentieren, sind Wege, Einflüsse expliziter Gedächtnisstrategien zu reduzieren. Implizite Gedächtnistests mit nicht zu hohen Itemschwierigkeiten und die Einschränkung von Antwortzeiten in der Testphase von Experimenten, reduzieren die Wahrscheinlichkeit der Anwendung expliziter Problemlösestrategien.

Eine Möglichkeit der Prüfung auf Prozessreinheit – wenn parallele Effekte gefunden wurden – ist es, die beteiligten Gedächtnistests auf stochastische Abhängigkeit zu untersuchen. Wenn parallele Effekte von stochastischer Abhängigkeit der Tests begleitet wird, ist zu vermuten, dass eine Form von Gedächtnis zu der Leistung in beiden Gedächtnistests beigetragen hat (Richardson-Klavehn & Bjork, 1988). Eine weitere Möglichkeit, Konfundierungen impliziter und expliziter Gedächtnisinhalte aufzudecken und ihren relativen Anteil zu bestimmen, ist die Anwendung der Prozessdissoziationsprozedur von Jacoby (1991). Bei der Prozessdissoziationsprozedur handelt es sich um eine Technik, um zwischen automatischen und kontrollierten Gedächtnisprozessen zu unterscheiden, von denen angenommen wird, dass sie in impliziten und expliziten Gedächtnistests selten unabhängig voneinander auftreten. Man hat mit der Prozessdissoziationsprozedur die Möglichkeit, den relativen Anteil automatischer und kontrollierter Gedächtnisprozesse an einer Aufgabe zu ermitteln. Implizite Gedächtnistests sollen typischerweise automatische kognitive Prozesse erfassen. Das sind unkontrollierte Prozesse, die schnell und unabhängig vom willentlichen Abruf durch Personen ablaufen. Leider können auch kontrollierte kognitive Prozesse, die langsamer und willentlich ablaufen, mit in die Leistung impliziter Gedächtnistests einfließen. Um den relativen Anteil dieser Prozesse in einem Wortstammergänzungstest zu ermitteln, ließen Jacoby, Toth und Yonelinas (1993) in einer sogenannten «Inklusionsbedingung» Personen Wortstämme mit Wörtern aus der Lernphase ergänzen. Falls entsprechende Wörter aus der Lernphase nicht erinnert werden konnten, sollte mit dem ersten verfügbaren Wort ergänzt werden. Die Inklusionsbedingung ist nach Jacoby et al. (1993) mit der expliziten Form des Wortstamm-

ergänzungstests vergleichbar, wobei kontrollierte und automatische Gedächtnisprozesse *zusammenarbeiten*. Nach Debner und Jacoby (1994) werden Rohwerte in expliziten Gedächtnistests üblicherweise auf diese Art bestimmt und durch eine Basisrate relativiert. Die Differenz beider Werte wird als Schätzung der expliziten Gedächtnisleistung verwendet (Debner & Jacoby, 1994). Diese Werte sind jedoch nach Debner und Jacoby (1994) durch automatische Abrufprozesse kontaminiert, die miterfasst werden. Um den Anteil automatischer Prozesse zu schätzen, verwendeten Jacoby et al. (1993) die sogenannte «Exklusionsbedingung», in der kontrollierte und automatische Prozesse *gegeneinander* arbeiten. Probanden wurden dazu angewiesen, Wortstämme zwar als Abrufhinweis für Wörter aus der Lernphase zu nutzen, sie jedoch mit Wörtern zu ergänzen, die nicht aus der Lernphase stammten. In der Exklusionsbedingung vermindern kontrollierte Gedächtnisprozesse die Wahrscheinlichkeit der Ergänzung von Wortstämmen mit Wörtern aus der Lernphase, während automatische Gedächtnisprozesse die Wahrscheinlichkeit der Ergänzungen von Wortstämmen mit Wörtern aus der Lernphase erhöhen (Jacoby et al., 1993). Über eine Rechenregel ist es schließlich möglich, den Anteil automatischer und kontrollierter Prozesse an dem Wortstammergänzungstest zu ermitteln. Abbildung 2 zeigt schematisch die Logik hinter der Bestimmung kontrollierter Gedächtnisprozesse. Der Anteil kontrollierter Gedächtnisprozesse wird durch die Differenz der Ergänzungen der Inklusionsbedingung und der Exklusionsbedingung geschätzt (z.B. Jacoby et al., 1993).

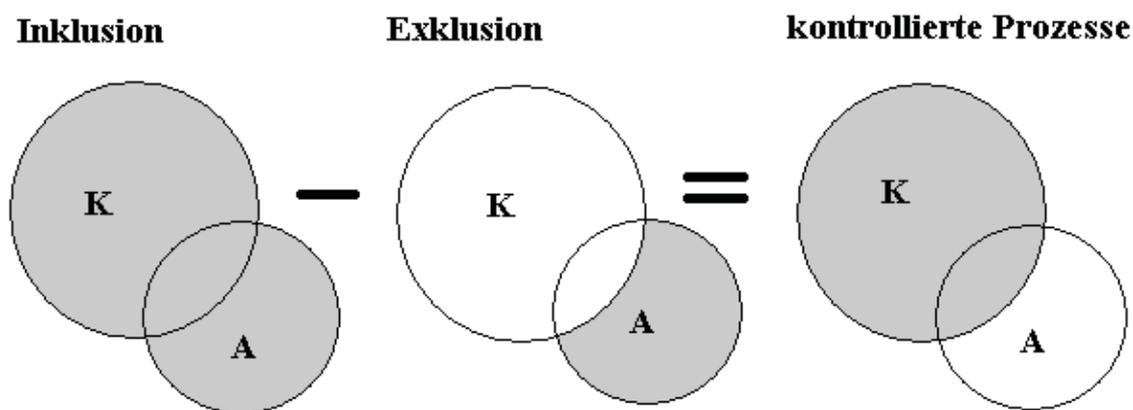


Abbildung 2: Bestimmung des Anteils kontrollierter Gedächtnisprozesse bei der Prozessdissoziationsprozedur (z.B. Jacoby et al., 1993). Der Buchstabe K steht für den Anteil kontrollierter Prozesse, der Buchstabe A steht für den Anteil automatischer Prozesse. (Abbildung nach Process Dissociation Tutorial, eigene Übersetzung).

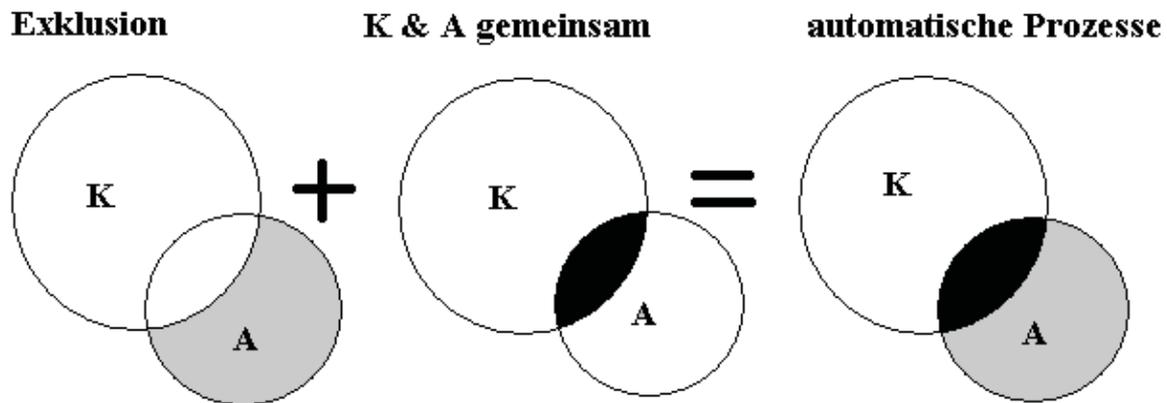


Abbildung 3: Bestimmung des Anteils automatischer Gedächtnisprozesse bei der Prozessdissoziationsprozedur (z.B. Jacoby et al., 1993). Der Buchstabe K steht für den Anteil kontrollierter Prozesse, der Buchstabe A steht für den Anteil automatischer Prozesse. (Abbildung nach Process Dissociation Tutorial, eigene Übersetzung).

Abbildung 3 veranschaulicht die Bestimmung automatischer Gedächtnisprozesse. Der Anteil automatischer Gedächtnisprozesse wird durch $A = E + K * A$ bzw. durch $A = E / (1 - K)$ geschätzt (z.B. Jacoby et al., 1993). Der Buchstabe A steht dabei für den Anteil automatischer Prozesse, der Buchstabe E steht für die Ergänzungen aus der Exklusionsbedingung und der Buchstabe K steht für den Anteil kontrollierter Prozesse. Die Prozessdissoziationsprozedur setzt Unabhängigkeit zwischen den beteiligten Prozessen voraus. Nur unter dieser Bedingung ist die bedingte Wahrscheinlichkeit des gemeinsamen Auftretens kontrollierter und automatischer Prozesse über ihr mathematisches Produkt definiert.

Die Prozessdissoziationsprozedur hat jedoch auch ihre Kritiker. So stellen einige Autoren die Annahme der Unabhängigkeit automatischer und kontrollierter Gedächtnisprozesse in Frage, auf der die Anwendung der Prozedur beruht (z.B. Cowan & Stadler, 1996; Curran & Hintzman, 1995, 1997; Graf & Komatsu, 1994; Roediger & McDermott, 1994). Curran und Hintzman (1995) zeigten beispielsweise, dass die Verletzung der Unabhängigkeitsannahme kontrollierter und automatischer Prozesse zu artifiziellen Prozessdissoziationen führen kann. Buchner, Erdfelder und Vaterrodt-Pluennecke (1995) zeigten weiterhin, dass das originale Messmodell von Jacoby (1991) nicht zwangsläufig prozessreine Messungen bewusster und unbewusster Gedächtnisprozesse liefert. Die Messungen können auch durch Antworttendenzen und Ratetendenzen kontaminiert sein. Die Autoren schlugen aus diesem Grund ein erweitertes Messmodell vor, welches auf einem multinomialen Modell basiert. Es soll

Messungen von Antworttendenzen entzerren und damit genauere Schätzungen bewusster und unbewusster Gedächtnisprozesse ermöglichen.

Verschiedene Studien sprechen jedoch dafür, dass explizite Abrufstrategien nicht grundsätzlich an der Leistung in impliziten Tests beteiligt sein müssen, vor allem dann nicht, wenn implizite Testanweisungen erfolgreich umgesetzt werden. Graf und Mandler (1984) beispielweise ließen Personen unter oberflächlichen oder unter semantischen Enkodierungsbedingungen Wörter lernen. In einer anschließenden Testphase sollten Wortstämme von Wörtern aus der Lernphase und Wortstämme neuer Wörter entweder mit impliziten oder mit expliziten Testinstruktionen ergänzt werden. Die Daten zeigten einen typischen Effekt der Verarbeitungstiefe. Wurden Wortstämme mit expliziten Testinstruktionen ergänzt, so profitierte die Testleistung von einer tiefen Verarbeitung der Wörter in der Lernphase. Wurden Wortstämme jedoch mit impliziten Testinstruktionen ergänzt, so hatte die Verarbeitungstiefe von Wörtern in der Lernphase keinen Einfluss auf die Testleistung. Wenn Personen tatsächlich explizite Abrufstrategien nutzen würden, um ihre impliziten Testleistungen zu verbessern, dann sollte der Effekt der Verarbeitungstiefe ebenfalls im impliziten Wortstammernergänzungstest messbar sein, da sie von der tiefen Verarbeitung von Wörtern in der Lernphase profitieren würden. Ein ähnliches Argument, das gegen die Nutzung expliziter Gedächtnisstrategien in impliziten Tests spricht, kommt aus der Erforschung von altersbedingten Dissoziationen zwischen impliziten und expliziten Gedächtnistests. Typischerweise findet man bei älteren Personen eine Abnahme der Gedächtnisleistung in expliziten Tests, während sich die Gedächtnisleistung in impliziten Tests nicht von der Gedächtnisleistung junger Personen unterscheidet. Wenn die Leistung in impliziten Gedächtnistests auch auf expliziten Abrufstrategien beruhen würde, so wäre die Leistung junger Personen besser als die Leistung älterer Personen, da junge Personen explizite Abrufstrategien besser nutzen können. Es käme zu den weiter oben besprochenen parallelen Effekten zwischen den beiden Gedächtnistests. Weitere Evidenz dafür, dass explizite Kontaminationen nicht unbedingt ein Problem darstellen müssen, liefern Studien, in denen doppelte Dissoziationen gefunden wurden (z.B. Jacoby, 1983). Das Priming in impliziten Tests kann nicht auf explizite Kontaminationen zurückzuführen sein, wenn ein expliziter Test sich gegensätzlich verhält.

1.5.2 Vergleichbarkeit von Gedächtnistests

Die Interpretierbarkeit von Dissoziationen zwischen impliziten und expliziten Gedächtnistests kann eingeschränkt sein, wenn Gedächtnistests miteinander verglichen werden, die sich in verschiedenen Aspekten voneinander unterscheiden. Auftretende Dissoziationen könnten in diesem Fall lediglich Unterschiede in den Gedächtnistests widerspiegeln (Merikle & Reingold, 1991). Reingold und Merikle (1988) stellten methodische Kriterien für die Vergleichbarkeit von impliziten und expliziten Gedächtnistests zusammen. Schacter, Bowers und Booker (1989) benannten dieses Set von Kriterien als «Retrieval Intentionality Criterion». Danach sind Dissoziationen zwischen zwei verschiedenen Maßen im allgemeinen umso aussagekräftiger, je ähnlicher sich die Tests zu deren Erfassung sind (Merikle & Reingold, 1991). Wenn explizite und implizite Gedächtnistests miteinander verglichen werden sollen, so ist es sinnvoll, die Aufgaben in allen ihren Eigenschaften gleich zu gestalten und den expliziten oder impliziten Testcharakter nur durch Testinstruktionen zu variieren, indem entweder ein Bezug zur Lernepisode hergestellt wird oder der Bezug zur Lernepisode weggelassen wird. Um maximale Vergleichbarkeit zu gewährleisten, sollten identische Stimuli in der Lernphase und ein identisches Antwortformat in der Testphase verwendet werden.

Nach Merikle und Reingold (1991) benutzen implizite und explizite Gedächtnistests oft unterschiedliche Hinweisreize bzw. Abrufhinweise in der Testphase. Während beispielsweise im impliziten Wortstammerngänzungstest die ersten drei Buchstaben von Wörtern präsentiert werden, werden im expliziten Rekognitionstest ganze Wörter präsentiert. In Studien, in denen Hinweisreize bzw. Abrufhinweise in der Testphase nicht die gleichen physischen Eigenschaften für alle beteiligten Tests besitzen, können Dissoziationen zwischen den Test auch durch die unterschiedlichen Hinweisreize bzw. Abrufhinweise entstanden sein (Merikle & Reingold, 1991). Der implizite Wortstammerngänzungstest und die explizite Form des Wortstammerngänzungstests beispielsweise, unterscheiden sich nur durch die Instruktionen in der Testphase. So wird in der expliziten Testform ein Bezug zur Lernepisode hergestellt, indem Wortstämme als Abrufhinweis für Wörtern aus der Lernphase genutzt werden sollen. Bei dem impliziten Wortstammerngänzungstest fehlt der Bezug zur Lernepisode, denn es soll mit dem ersten Wort ergänzt werden, welches Probanden zu den Wortstämmen einfällt. Stimuli in der Lernphase und Hinweisreize bzw. Abruf-

hinweise in der Testphase jedoch unterscheiden sich nicht zwischen der impliziten und der expliziten Form des Wortstammerngänzungstests.

1.5.3 Reliabilität impliziter Gedächtnistests

Dissoziationen entstehen, wenn die Manipulation experimenteller Variablen (z.B. Länge des Retentionsintervalls, Interferenz, Verarbeitungstiefe) sich unterschiedlich oder sogar gegensätzlich auf implizite und explizite Gedächtnistests auswirkt. Meistens werden große Gruppenunterschiede in Leistungen expliziter Gedächtnistests gefunden, während in impliziten Gedächtnistests oft keine oder nur geringe Gruppenunterschiede gefunden werden (Buchner & Brandt, 2003). Eine Möglichkeit, Dissoziationen zwischen impliziten und expliziten Gedächtnistests zu erklären, ist es, unterschiedliche Gedächtnissysteme (z.B. Tulving & Schacter, 1990) oder unterschiedliche Gedächtnisprozesse (z.B. Roediger, 1990) zu postulieren und dafür verantwortlich zu machen. Eine solche Interpretation muss jedoch nicht unbedingt angemessen sein, weil der gleiche Befund auch als methodisches Artefakt geringer Reliabilität impliziter Gedächtnistests in Verbindung mit hoher Reliabilität expliziter Gedächtnistests entstehen kann (Buchner & Wippich, 2000). Explizite Gedächtnistests sind nach Buchner und Wippich (2000) üblicherweise recht reliabel, während implizite Tests oftmals unreliabel sind. Die folgenden Abschnitte sollen ein Grundverständnis des Reliabilitätskonzepts in der Psychometrie vermitteln. Dies ist wichtig, um ein Verständnis für die experimentelle Vorgehensweise dieser Arbeit aufzubauen.

1.5.3.1 Testtheoretische Definition der Reliabilität

Nach Bühner (2006, S. 27) besteht ein Messwert in der klassischen Testtheorie aus einem wahren Wert und einem Messfehler. Der wahre Wert ist definiert als der Mittelwert über unendlich viele Messwerte einer Person, die unter gleichen Bedingungen erfasst werden. Der Mittelwert des Messfehlers über unendlich viele Messungen an einer Person ist null. Der Messfehler enthält alle unsystematischen Einflüsse einer Messung. Anzumerken ist dabei, dass der Messfehler in der klassischen Testtheorie nur unsystematische Einflüsse berücksichtigt und systematische Einflüsse dem wahren Wert zugerechnet werden (Bühner, 2006, S. 26). Die Reliabilität eines Tests oder seine Messgenauigkeit ist als der Anteil der Varianz der wahren Werte an der Ge-

samtvarianz der beobachtbaren Messwerte definiert, wobei die Gesamtvarianz der beobachtbaren Messwerte aus der Summe der Varianz der wahren Werte und der Messfehlervarianz besteht ($r_{tt} = S_T^2 / (S_T^2 + S_E^2)$). Mit zunehmend kleinerem Anteil der Messfehlervarianz an der Gesamtvarianz wird der Nenner immer kleiner, die Reliabilität steigt. Im Extremfall fehlender Messfehlervarianz bestehen Zähler und Nenner nur noch aus der Varianz der wahren Werte, was einer maximal möglichen Reliabilität entspricht. Dieser theoretische Fall einer perfekten Reliabilität von eins wird jedoch in der Praxis schwer zu realisieren sein, da Messungen stets mit Messfehlern behaftet sind.

1.5.3.2 Zusammenhang der Hauptgütekriterien

Die Reliabilität oder Zuverlässigkeit ist die Genauigkeit, mit der ein Test ein Merkmal misst, unabhängig davon, ob der Test dieses Merkmal auch messen soll. Eine hohe Reliabilität von Tests ist wünschenswert, weil solche Tests eine hohe Messpräzision besitzen. Die Validität oder Gültigkeit eines Tests ist die Genauigkeit, mit der ein Test genau das Merkmal misst, das er auch messen soll. Ein hochreliabler Test macht folglich nur dann Sinn, wenn er auch eine hohe Validität besitzt und das Merkmal erfasst, welches er erfassen soll. Ein wenig reliabler Test, der das zu erfassende Merkmal misst, wird wahrscheinlich einem hochreliablen Test vorgezogen, der ein Merkmal misst, für das kein Interesse besteht (Krauth, 1995, S. 78). Über die Minderungsformel kann berechnet werden, wie die Validität eines Tests durch den Messfehler gemindert wird (Krauth, 1995, S. 264). Die Validität – das wahrscheinlich wichtigste Gütekriterium eines Tests – kann in ihrem Betrag niemals größer werden, als die Wurzel aus der Reliabilität. Schon aus diesem Grund ist es immer sinnvoll, einen reliabilitätsoptimalen Test zu konstruieren. Eine hohe Objektivität, dem Ausmaß, inwieweit Ergebnisse unabhängig vom Untersucher sind, ist Voraussetzung für eine hohe Testreliabilität.

1.5.3.3 Bedeutung der Reliabilität für die Psychometrie

Die Reliabilität eines psychometrischen Tests ist vergleichbar mit der Messgenauigkeit eines physikalischen Messinstrumentes. Eine Laborwaage beispielsweise, die von äußeren Störeinflüssen abgeschirmt ist und eine präzise Mechanik hat, liefert aufgrund von geringeren Messfehlern genauere Messwerte ab, als eine konventionel-

le Küchenwaage. Ist die Messpräzision hoch, dann ist ein psychometrischer Test in der Lage, wahre Werte genauer durch beobachtbare Werte zu schätzen. Ist ein psychometrischer Test jedoch messfehlerbehaftet, sinkt seine Reliabilität oder Messpräzision ab. Wahre Werte werden in diesem Fall durch das Rauschen des Messfehlers überlagert und können somit nur ungenau durch beobachtbare Werte geschätzt werden. Psychometrische Tests geringer Reliabilität führen in der Konsequenz dazu, dass bedeutsame Gruppenunterschiede erst bei deutlich größeren Effekten durch statistische Tests gegen Zufallseinfluss abgesichert werden können, als dies bei psychometrischen Tests hoher Reliabilität der Fall ist. Sutcliffe (1958) fand, dass die Teststärke zum Nachweis von Gruppenunterschieden in Varianzanalysen mit steigendem Messfehler geringer wurde. Ausgehend von festen Effektstärken führen Tests höherer Reliabilität zu einer größeren Teststärke und damit zu besseren Chancen, signifikante Gruppenunterschiede zu finden, als Tests geringerer Reliabilität (Hallahan & Rosenthal, 1996). Rogers und Hopkins (1988) geben Formeln an, um die Teststärke – in Abhängigkeit von der Reliabilität von Tests – zu berechnen. Kopriva und Shaw (1991) stellen Tafeln zur Schätzung der Teststärke in Varianzanalysen bereit, in Abhängigkeit von Reliabilität, Effektstärke und Stichprobenumfang. An dieser Stelle gewinnt die Inferenzstatistik an Bedeutung, deren Existenzberechtigung an das Vorhandensein des Messfehlers geknüpft ist. Hätte man es mit perfekt reliablen Messungen zu tun, die mit keinerlei Messfehler beaufschlagt wären, so könnte man Rohwerte direkt miteinander vergleichen und entscheiden, ob Unterschiede zwischen Stichproben bedeutend sind. Da man es jedoch nicht mit perfekt reliablen Messungen zu tun hat, werden statistische Tests benötigt, um entscheiden zu können, ob Stichproben sich tatsächlich oder nur aufgrund von Messfehlern unterscheiden.

Zuletzt soll nicht unerwähnt bleiben, dass Reliabilität nicht nur eine Eigenschaft von psychometrischen Tests ist, sondern auch eine Eigenschaft der Population (Krauth, 1995, S. 250). Die Reliabilität geht gegen null, wenn sich Mitglieder einer Population nicht in ihren wahren Werten voneinander unterscheiden. Aus diesem Grund können heterogene Stichproben zu höheren Reliabilitätschätzungen derselben Tests führen als homogene Stichproben. Dies kann damit begründet werden, dass Reliabilitätskoeffizienten durch hohe Testwertestreuungen heterogener Populationen ansteigen. Entsprechend können Stichproben die Reliabilität von Populationen erheblich unterschätzen oder überschätzen, je nachdem ob Stichproben geringere oder

größere Testwertestreuungen aufweisen, als Populationen aus denen sie entnommen wurden (Lienert, 1989, S. 238).

1.5.3.4 Schätzung der Reliabilität

Die Höhe der Reliabilität kann nur *geschätzt* werden, da die Varianz der wahren Werte nicht bekannt ist und auch nicht berechnet werden kann. Die Güte dieser Schätzung hängt von der Äquivalenz der Durchführungsbedingungen von Messungen ab, deren Messwerte untereinander korreliert werden sollen (Bühner, 2006, S.124). Messungen können an einzelnen Items, Tests oder Teilen von Tests durchgeführt werden. Dabei müssen wahre Werte und Messfehler dieser Messungen – je nach gestelltem Kriterium – gewisse Äquivalenzbedingungen erfüllen. Bei streng parallelen Messungen beispielsweise dürfen sich wahre Werte und Messfehlervarianz nicht voneinander unterscheiden. Je besser Messungen miteinander korrelieren, desto höher fällt die Reliabilitätsschätzung eines Tests aus und desto geringer ist der Einfluss des Messfehlers. In der Praxis wird die Reliabilität durch die Methode der Testhalbierung, der Ermittlung der inneren Konsistenz, der Retestmethode oder der Paralleltestmethode bestimmt. Zur Bestimmung der Testhalbierungsreliabilität werden Tests nach einer bestimmten Methode in zwei Testhälften unterteilt, deren wahre Werte und Messfehler gleich sein sollten. Die Korrelation der Rohwertpaare beider Testhälften wird als Schätzung für die Reliabilität herangezogen. Anschließend wird die Reliabilitätsschätzung über eine Korrekturformel auf die gesamte Testlänge hochgerechnet (Bühner, 2006, S. 127). Die Methode der inneren Konsistenz unterscheidet sich von der Methode der Testhalbierung nur dadurch, dass Tests in genau so viele Untertests zerlegt werden, wie sie Items enthalten (Bühner, 2006, S. 127). Bei der Retestmethode werden in einem ausreichend bemessenen Zeitabstand dieselben Tests an denselben Stichproben erhoben (Bühner, 2006, S. 128). Bei der Paralleltestmethode werden zwei Parallelformen eines Tests an denselben Stichproben erhoben (Bühner, 2006, S. 129). Jede der hier vorgestellten Methoden der Reliabilitätsbestimmung hat ihre Vor- und Nachteile und damit ihr bevorzugtes Anwendungsgebiet.

1.5.3.5 Dissoziationsbefunde und Reliabilitätsdifferenzen beteiligter Tests

Der Reliabilität von Gedächtnistests – als wichtigem Aspekt psychometrischer Datenerfassung – wurde in bisherigen Gedächtnisstudien wenig Beachtung geschenkt.

Reliabilitätsschätzungen von $r_{tt} < .80$, gelten als gering. Mittlere Reliabilitätsschätzungen liegen im Bereich zwischen $r_{tt} = .80$ und $r_{tt} = .90$, und Reliabilitätsschätzungen von $r_{tt} > .90$ gelten als hoch (Bühner, 2006, S. 140). Die Replizierbarkeit von postulierten Effekten wird oft als Beweis für ausreichende Reliabilität von Tests angesehen (Meier & Perrig, 2000). Einige Untersuchungen fanden jedoch deutlich geringere Reliabilitätsschätzungen für implizite Gedächtnistests, im Vergleich zu expliziten Gedächtnistests (z.B. Buchner & Brandt, 2003; Buchner & Wippich, 2000; Meier & Perrig, 2000). Mit Hilfe einer Korrelationsanalyse fanden Perruchet und Baveux (1989) geringere Interkorrelationen für implizite Gedächtnistests als für explizite Gedächtnistests. Meier und Perrig (2000) konnten diese Befunde in ihrer eigenen Studie bestätigen. Die Höhe von Interkorrelationen ist begrenzt durch die Höhe der Reliabilitätsschätzungen der daran beteiligten Tests. Die geringen Interkorrelationen impliziter Tests in den genannten Studien könnten durch ihre geringe Reliabilität bedingt sein.

Unbekannte oder geringe Reliabilität von Tests führt zu Mehrdeutigkeit in der Interpretation von Dissoziationen zwischen impliziten und expliziten Gedächtnistests (Buchner & Brandt, 2003). Werden Daten mit typischerweise sehr reliablen expliziten Gedächtnistests erhoben, so sind häufig große Gruppenunterschiede vorhanden (Buchner & Brandt, 2003). Kommen jedoch deutlich unrelablere implizite Gedächtnistests zur Anwendung, so werden keine oder nur geringe Gruppenunterschiede gefunden (Buchner & Brandt, 2003). Solche und ähnliche Befunde von klassischen einfachen Dissoziationen zwischen impliziten und expliziten Gedächtnistests, legten bisher den Schluss unterschiedlicher zugrunde liegender Gedächtnissysteme (z.B. Tulving & Schacter, 1990) oder Gedächtnisprozesse (z.B. Roediger, 1990) nahe. Leider ist die Interpretation dieser Befunde aufgrund geringer Reliabilität vieler impliziter Tests nicht eindeutig. Nach Buchner und Wippich (2000) können artifizielle Dissoziationen zwischen zwei Tests, die genau das gleiche Konstrukt messen, auftreten, wenn sich die Tests ausschließlich bezüglich ihrer Reliabilität hinreichend voneinander unterscheiden. Je höher die Reliabilität einer der beiden beteiligten Tests ist, desto höher ist nach Buchner und Wippich (2000) die Wahrscheinlichkeit, statistisch signifikante Gruppenunterschiede zu finden. Je geringer die Reliabilität des zweiten Tests ist, desto geringer ist die Wahrscheinlichkeit, signifikante Gruppenunterschiede zu finden (Buchner & Wippich, 2000). Viele der bisherigen Befunde von Dissoziationen könnten somit auch Artefakte geringer Reliabilität impliziter Gedächtnistests

sein. Zusammen mit hoher Reliabilität expliziter Gedächtnistests könnten Muster einfacher Dissoziationen auch dann zu Stande kommen, wenn Gruppenunterschiede in impliziten Tests tatsächlich vorhanden sind, diese jedoch aufgrund geringer Messpräzision impliziter Tests unentdeckt bleiben.

Geringe Reliabilität impliziter Gedächtnistests hat zudem negative Auswirkungen auf die Replikationswahrscheinlichkeit von Befunden. LeBel und Paunonen (2011) führten dazu eine Monte Carlo Simulation durch, um den Einfluss der Reliabilität eines hypothetischen impliziten Tests auf die Replizierbarkeit von Befunden eines einfachen zwei-Gruppen Versuchsplans mit unabhängigen Daten zu ermitteln. Sie generierten dafür Datensätze verschiedener Reliabilitätsstufen. Weiterhin wurde der Einfluss fünf verschiedener Stichprobenumfänge ($N = 10$ bis $N = 50$) sowie der Einfluss der Größe des Populationseffektes auf die Replizierbarkeit von Befunden untersucht. Die drei Stufen der untersuchten Populationseffektgröße waren nach Cohen (1977) kleine, mittlere und große Effekte ($d = .2$, $d = .5$ und $d = .8$). Für kleine und für mittlere Effektgrößen wurde ein linearer Zusammenhang zwischen der Höhe der Reliabilität des Tests und der Replizierbarkeit von Befunden festgestellt. Für große Effekte und für Stichprobenumfänge ab $N = 20$ wird der zunächst steil lineare Zusammenhang zwischen Reliabilität und Replizierbarkeit immer flacher. Zusätzlich hängt die Replizierbarkeit stark mit der Größe der Effekte zusammen. Große Effekte haben eine deutlich höhere Wahrscheinlichkeit repliziert zu werden, als kleine Effekte. Mit zunehmendem Stichprobenumfang steigt die Teststärke und damit auch die Replizierbarkeit von Befunden. Bei der höchstmöglichen Reliabilität von $r_{tt} = 1$ wird die maximale Replikationswahrscheinlichkeit und die damit theoretisch mögliche Teststärke des jeweiligen Setups erreicht. So berichteten LeBel und Paunonen (2011) eine Wahrscheinlichkeit, kleine Populationseffekte zu replizieren, von lediglich $p = .17$, bei einem Stichprobenumfang von $N = 50$ und einer recht hohen Reliabilität von $r_{tt} = .90$. Eine Reliabilitätsschätzung dieser Größenordnung ist typischerweise bei expliziten Tests zu finden. Bei kleinen Effekten werden somit sehr hohe Stichprobenumfänge benötigt, um selbst mit sehr reliablen Tests Befunde replizieren zu können. Der Einfluss der Reliabilität auf die Replizierbarkeit von Befunden ist besonders ab mittleren Effektstärken deutlich ausgeprägt und wächst mit dem Stichprobenumfang. Bei kleinen Stichprobenumfängen und der damit verbundenen geringen Teststärke spielt Reliabilität noch eine untergeordnete Rolle. Dies ändert sich jedoch bei zunehmendem Stichprobenumfang. LeBel und Paunonen (2011) berichteten den

Vergleich eines typischen impliziten Tests ($r_{tt} = .30$) mit einem typischen expliziten Test ($r_{tt} = .80$). Bei einem Stichprobenumfang von $N = 40$ Personen und einer mittleren Effektstärke ist die Wahrscheinlichkeit der Replikation von Befunden $p = .25$ für den impliziten Test und $p = .50$ für den expliziten Test. In diesem praxisnahen Vergleich, bei dem ein unreliabler impliziter Test und ein reliabler expliziter Test simuliert werden, ist die Wahrscheinlichkeit, Befunde zu replizieren, doppelt so hoch für den expliziten Test im Vergleich zu dem impliziten Test. Dies könnte eine Erklärung für Replikationsprobleme von Dissoziationen zwischen impliziten und expliziten Gedächtnistests sein. Explizite Gedächtnistests, die aufgrund hoher Reliabilität zuverlässig Befunde replizieren, in Verbindung mit impliziten Gedächtnistests, die aufgrund ihrer geringen Reliabilität Befunde nur unzuverlässig replizieren, führen zu den gemischten Befunden, die oftmals in der Literatur zu finden sind. Buchner und Wippich (2000) zeigten, dass Reliabilitätsdifferenzen zwischen impliziten und expliziten Gedächtnistests für das in der Literatur oft anzutreffende Befundmuster häufiger Dissoziationen und weniger Assoziationen verantwortlich sein können. Die in der Literatur häufig berichteten Dissoziationen zwischen impliziten und expliziten Gedächtnistests können somit erst dann eindeutig interpretiert werden, wenn die Reliabilität der verwendeten Tests berichtet wird. Solange dies nicht der Fall ist, kann nicht darüber entschieden werden, ob es sich um tatsächlich bestehende Dissoziationen oder um Artefakte von Reliabilitätsdifferenzen der verwendeten Tests handelt (Buchner & Brandt, 2003). Es ist leider nicht bekannt, wie groß der Anteil von Artefakten an bisherigen Befunden ist, weil psychometrische Eigenschaften von Gedächtnistests üblicherweise nicht berichtet werden (Buchner & Brandt, 2003). Doppelte Dissoziationen, in denen experimentelle Manipulation einer unabhängigen Variablen explizite und implizite Gedächtnistests in umgekehrter Weise beeinflusst, sind von dem Reliabilitätsproblem nicht betroffen (Buchner & Wippich, 2000).

1.5.3.6 Ursachen geringer Reliabilität impliziter Gedächtnistests

Reliabilitätsdifferenzen zwischen impliziten und expliziten Gedächtnistests können nach Buchner und Wippich (2000) auf die im Folgenden beschriebenen Ursachen zurückgeführt werden. In expliziten Gedächtnistests sind Testinstruktionen typischerweise eindeutig formuliert (Buchner & Wippich, 2000). Im expliziten Rekognitions-test beispielsweise gibt es nur zwei mögliche Antwortalternativen. Alte Items müssen von neuen Items unterschieden werden und die richtige Antwort im Sinne

der Testinstruktionen ist auch die objektiv korrekte Antwort (Buchner & Wippich, 2000). Eine genaue Definition des Leistungsziels, wenige und vorher genau festgelegte Antwortalternativen sowie die Erzeugung einer Atmosphäre, in der Probanden an der oberen Grenze ihrer Leistungsfähigkeit antworten müssen, führen zu geringer Variabilität kognitiver Prozesse (Buchner & Wippich, 2000). Eine geringe Variabilität kognitiver Prozesse wiederum führe zu hoher Reliabilität. In impliziten Gedächtnistests jedoch – wie beispielsweise dem häufig angewandten Wortstammernergänzungstest – sind Testinstruktionen oft weniger genau als in expliziten Gedächtnistests (Buchner & Wippich, 2000). So sollen Wortstämme mit dem ersten Wort ergänzt werden, welches Probanden einfällt. Nach Buchner und Wippich (2000) erlauben ungenau formulierte Testinstruktionen Probanden viel Spielraum, richtige Antworten im Sinne der Testinstruktionen zu geben. Das Problem dabei ist jedoch, dass nicht jede richtige Antwort im Sinne der Testinstruktionen eine korrekte Antwort im Sinne des impliziten Tests ist, weil nur in der Lernphase präsentierte Wörter als korrekte Antworten gewertet werden (Buchner & Wippich, 2000). Nach Buchner und Wippich (2000) könnten Probanden Wortstämme beispielsweise auch aufgrund ihrer phonologischen oder graphemischen Eigenschaften ergänzen oder sie könnten weiterhin versuchen, mit besonders ungewöhnlichen Wörtern zu antworten. Oder sie versuchen, Wortstämme mit Wörtern aus der Lernphase zu ergänzen, weil dies als besonders einfacher Weg erscheint, Aufgaben zu lösen (Buchner & Wippich, 2000). Je nach gewählter Strategie, würden sich Itembeantwortungen systematisch voneinander unterscheiden (Buchner & Wippich, 2000). Insgesamt bringt großer Spielraum bei der Wahl von Antwortstrategien, verursacht durch unpräzise Testinstruktionen, hohe Variabilität kognitiver Prozesse mit sich. Hohe Variabilität kognitiver Prozesse verursacht eine Verringerung der Reliabilität von Messungen.

1.5.3.7 Methoden zur Steigerung der Reliabilität impliziter Gedächtnistests

Nach Buchner und Brandt (2003) werden in der Literatur selten Angaben zu Gütekriterien von Gedächtnistests gemacht. Damit ist auch die Höhe der Reliabilität vieler impliziter Gedächtnistests unbekannt. Wenn die Reliabilität impliziter Gedächtnistests jedoch berichtet wird, dann ist sie meistens weitaus geringer, als die Reliabilität vergleichbarer expliziter Gedächtnistests (Buchner & Wippich, 2000). Die geringe Reliabilität impliziter Gedächtnistests erschwert es, im Gegensatz zu der meist hohen Reliabilität expliziter Gedächtnistests, kleinere Effekte nachzuweisen und replizieren

zu können. Um die Vergleichbarkeit impliziter und expliziter Gedächtnistests in Dissoziationsstudien zu gewährleisten, ist es unter anderem wichtig, implizite Tests zu verwenden, deren Reliabilität sich auf dem Niveau expliziter Tests befindet.

Um eine Erhöhung der Reliabilität von Messungen zu erreichen, ist es notwendig, den Anteil von Fehlervarianz an der Gesamtvarianz von Daten zu reduzieren. Es sollten zu diesem Zweck psychometrische Tests zur Anwendung kommen, die beispielsweise mit Methoden der klassischen Testtheorie reliabilitätsoptimiert wurden. Nach Krauth (1995, S. 20) besteht der erste Schritt einer Testkonstruktion in der Anlage eines Itempools, einer Zusammenstellung von manifesten Variablen (Items), von denen man annimmt, dass sie mit der zu messenden latenten Variablen (Konstrukt) zusammenhängen. Die Zusammenstellung von Items wird anschließend einer möglichst repräsentativen Stichprobe von Probanden vorgelegt. Alle Aussagen zu Gütekriterien von Tests sind nur gültig für die Population, für die eine Eichstichprobe repräsentativ ist (Krauth, 1995, S. 20). Die an einer Eichstichprobe erhobenen Daten bilden die Basis für eine anschließende Itemanalyse. In der Itemanalyse werden Itemparameter, wie beispielsweise Itemschwierigkeiten, Itemreliabilitäten oder Itemtrennschärfen geschätzt. Man versucht nun mit Hilfe der Schätzungen von Itemparametern Items aus dem Pool so auszuwählen, dass der Test im Hinblick auf Reliabilität und Validität optimal wird (Krauth, 1995, S. 20). Es schließt sich eine Verteilungsanalyse und die Bestimmung von Normwerten an, mit denen man Personen aufgrund ihres Testwertes in der Eichpopulation einordnen kann (Krauth, 1995, S. 21).

Zur Reliabilitätsoptimierung von Tests müssen alle Items ausgeschlossen werden, die zu einer Verringerung der Reliabilitätsschätzung beitragen. Eine Zusammenstellung von Items, die ausschließlich auf hohen Itemreliabilitäten basiert, wäre jedoch ein Fehler, denn die Höhe der Testreliabilität hängt ebenfalls von Itemvarianzen und Itemkorrelationen ab (Krauth, 1995, S. 273). Ein Test, der eine Zusammenstellung von hochreliablen Items beinhaltet, muss deshalb nicht notwendigerweise der Test mit der höchsten Reliabilität sein. Bei der Methode der Homogenisierung werden Items geringer Eigentrennschärfen aus dem Itempool entfernt. Im Hinblick auf eine Reliabilitätsoptimierung sind homogen mittelschwere Items aufgrund ihrer höheren Eigentrennschärfen heterogenen Items vorzuziehen (Bühner, 2006, S. 99). Unter der Eigentrennschärfe eines Items versteht man seine Korrelation mit der gesamten Ska-

la. Inhaltlich beschreibt die Trennschärfe, wie gut ein Item alle anderen Items der Skala widerspiegelt (Bühner, 2006, S. 95). Der Zusammenhang zwischen Itemschwierigkeit und Trennschärfe ist umgekehrt U-förmig. Mittelschwere Items sind am trennschärfsten, während leichte und schwere Items weniger trennscharf sind (Lienert, 1989, S. 40). Die Verbesserung der Testreliabilität erfolgt aus dem Grund, dass Itembeantwortungen von Items mittlerer Schwierigkeit ähnlich verteilt sind und nur gleich verteilte Itembeantwortungen maximal miteinander korrelieren können. Außerdem erhöhen mittlere Schwierigkeitsindizes die Wahrscheinlichkeit hoher Streuungen und damit hoher Trennschärfen (Bühner, 2006, S. 99). Hohe Trennschärfen wiederum gewährleisten eine maximale Differenzierung zwischen Personen (Bühner, 2006, S. 98). Eine ausreichende Merkmalsstreuung, wenn es sich dabei um systematische Varianz handelt, ist eine notwendige Voraussetzung für hohe Korrelationen zwischen den Items und damit einer hohen Testreliabilität (Bühner, 2006, S.105). Die α -Maximierung ist ein Verfahren zur Erhöhung der Testreliabilität, bei dem sukzessive die Items mit den geringsten Trennschärfen aus dem Itempool entfernt werden (Bühner, 2006, S. 146). Es werden dazu Items ausgesondert, die nur schlecht mit der Skala korrelieren. Das Cronbach- α als Maß der inneren Konsistenz von Tests steigt an und damit auch die Reliabilität. Die Homogenisierung von Tests durch α -Maximierung kann zu einer Einengung von Konstrukten führen, weil durch die Ähnlichkeit von Items kleinere Verhaltensausschnitte erfasst werden (Bühner, 2006, S. 138). Dies ist kritischer bei Leistungstests, die auch in äußeren Merkmalsbereichen differenzieren müssen. Bei solchen Tests sollte man einige leichte und schwere Items im Itempool belassen. Leistungstests bestehen aus einem Itempool, der das gesamte Merkmalspektrum erfasst. Items werden dabei nach steigender Schwierigkeit angeordnet (Krauth, 1995, S. 182), wobei meist keine Antwortzeitbegrenzung oder eine sehr großzügig bemessene Antwortzeitbegrenzung vorgegeben wird. Der Hintergrund ist der, dass die schwierigsten Items nur von den fähigsten Probanden lösbar sein sollen. Auf der anderen Seite des Merkmalspektrums befinden sich sehr leichte Items. Insgesamt erreicht man durch sinnvoll zusammengestellte Leistungstests eine gute Differenzierung zwischen Probanden aller Fähigkeitsausprägungen. Schnelligkeitstests bestehen typischerweise aus einem homogenen Pool mittelschwererer Items und haben deshalb meist schon an sich eine höhere Reliabilität. Die Bearbeitungszeit der Items wird dabei so bemessen, dass keiner der Probanden alle Items lösen kann. Die Differenzierung zwischen Probanden erfolgt bei Schnelligkeitstests

durch die unterschiedliche Bearbeitungsgeschwindigkeit der Items.

In psychometrischen Tests mit einem dichotomen Antwortformat werden für die Trennschärfeanalyse punktbiseriale Korrelationen zwischen Items und Skala ermittelt, was eine Berechnung auf Itemebene voraussetzt und damit nur mit Rohwerten vorgenommen werden kann. Priming impliziter Tests jedoch wird nicht auf Itemebene erfasst. Zur Bestimmung des Primingwertes werden Rohwerte einzelner Items aufsummiert und einer Basisratenkorrektur unterzogen. Priming stellt damit ein Maß dar, welches über mehrere Items ermittelt wird. Die Steigerung der Reliabilität auf der Ebene von Items stellt sich deshalb als schweres Unterfangen heraus (Meier & Perrig, 2000). Würde man Trennschärfeanalysen mit Rohwerten impliziter Tests durchführen, so wären die Daten durch verschiedene Antworttendenzen kontaminiert (Buchner & Brandt, 2003). Aus diesen Gründen wurde in den Experimenten der vorliegenden Arbeit auf Trennschärfeanalysen und α -Maximierung zur Reliabilitätsoptimierung von impliziten Gedächtnistests verzichtet.

Eine weitere Möglichkeit der Reliabilitätssteigerung ist die Testverlängerung mit homogenen Items (Bühner, 2006, S. 138). Durch Ausmittlung des unsystematischen Messfehlers steigt die Höhe der Reliabilität eines psychometrischen Tests bei einer Testverlängerung an. Eine Verdoppelung der Testlänge führt zu einer Vervierfachung der Varianz der wahren Werte, jedoch nur zu einer Verdoppelung der Fehlervarianz. Der Anteil der Varianz der wahren Werte an der beobachtbaren Varianz wird zunehmend größer, was einer höheren Reliabilität entspricht. Je mehr homogene Items einer Skala hinzugefügt werden, desto höher wird das Cronbach- α der inneren Konsistenz.

Fehlervarianz lässt sich jedoch auch bei der Datenerhebung verringern. Durch standardisierte Darbietung von Testinstruktionen kommt es zu höherer Durchführungsobjektivität. Zeitbegrenzungen und Hilfestellungen müssen für alle Probanden gleich sein. Durch allgemein verständliche Instruktionen sind Rückfragen und damit verbundene Unterschiede in Hilfestellungen vermeidbar. Übungsbeispiele können hilfreich sein, Missverständnissen vorzubeugen (Bühner, 2006, S. 138). Automatische Answererfassung ist manueller Answererfassung durch den Versuchsleiter vorzuziehen, da die Objektivität unter manueller Answererfassung leiden kann. Zwischen Objektivität und Reliabilität besteht eine lineare Abhängigkeit. Je größer die Objektivität

viät von Messungen, desto höher ist im allgemeinen auch ihre Reliabilität (Lienert, 1989, S.40).

Implizite Gedächtnistests sind nicht an sich unreliabel sondern vermutlich deshalb, weil sie eine große Spannbreite an Möglichkeiten zulassen, Aufgaben zu lösen (Buchner & Brandt, 2003). Deshalb müssen Maßnahmen angewandt werden, um die Variabilität kognitiver und nicht kognitiver Prozesse einzuschränken und nur diejenigen Gedächtnisprozesse zu erfassen, die erfasst werden sollen. Eine Maßnahme ist die Einschränkung der Anzahl möglicher Antwortalternativen in impliziten Gedächtnistests. Die explizite Form des Wortstammergänzungstests beispielsweise verlangt von Probanden, Wortstämme als Abrufhinweis für Wörter aus der Lernphase zu verwenden und mit diesen zu ergänzen. Hier ist das Leistungsziel klar definiert und es gibt nur eine korrekte Antwort. In Verbindung mit einer Atmosphäre, in der Probanden an ihrer oberen Leistungsgrenze antworten müssen, führt dies – durch geringe Variabilität in den Antwortstrategien – zu einer hohen Reliabilität des expliziten Tests. Der implizite Wortstammergänzungstests jedoch erfordert, dass Wortstämme mit dem ersten Wort ergänzt werden, welches Probanden einfällt. Je mehr mögliche Ergänzungen für Wortstämme existieren, desto mehr gültige Antworten gibt es im Sinne der Testinstruktionen. Es gibt jedoch nur eine korrekte Antwort im Sinne von implizitem Gedächtnis. Viele mögliche Antwortalternativen haben eine erhöhte Variabilität kognitiver Prozesse zur Folge und damit eine Verringerung der Reliabilität. Leistungsatmosphäre kann möglicherweise dadurch geschaffen werden, dass Probanden angewiesen werden, zügig zu antworten und indem ihnen mitgeteilt wird, dass ihre Antwortzeiten erfasst werden. Zusätzlich kann Probanden regelmäßiges Feedback über ihre Leistungen gegeben werden.

Eine Beschränkung des verfügbaren Zeitrahmens, der Probanden für Antworten auf Hinweisreize in der Testphase impliziter Gedächtnistests zur Verfügung gestellt wird, ist eine weitere mögliche Maßnahme, um der Variabilität von Gedächtnisprozessen entgegenzuwirken. Der Vorteil einer Antwortzeitbegrenzung gegenüber der alleinigen Anweisung zügig zu antworten ist es, dass Probanden nicht die Möglichkeit einer Interpretation des Begriffes «zügig» gegeben wird. Probanden könnten eine solche Anweisung sonst eventuell unterschiedlich deuten. Aber auch ein und derselbe Proband könnte sich für die verschiedenen Itembeantwortungen unterschiedlich viel Zeit lassen. In dem Wortstammergänzungstest beispielsweise, könnte

der zeitliche Rahmen für die Ergänzung von Wortstämmen auf einige Sekunden begrenzt werden.

Eine Möglichkeit der Einschränkung der Variabilität von Gedächtnisprozessen in impliziten Gedächtnistests mit Wörtern als Stimuli ist es, Probanden in der Lernphase von Experimenten dargebotene Wörter aussprechen zu lassen oder Probanden anzuweisen, sich die Begriffe zu den dargebotenen Wörtern bildlich vorzustellen. Diese Maßnahmen sollten dadurch, dass alle Probanden gleiche Enkodierungsstrategien verwenden, zu einer Einengung des Suchraums beim Abruf führen.

Um die Wahrnehmung eines Zusammenhangs zwischen Lernphase und Testphase von Experimenten zu verhindern, sollte eine gewisse Zeitspanne zwischen Lernphase und Testphase eingeplant werden, in der beispielsweise Ablenkungsaufgaben bearbeitet werden. Zuletzt können Personen aus der Untersuchung ausgeschlossen werden, die in einer anschließenden Befragung einen Zusammenhang zwischen Lernphase und Testphase wahrgenommen haben. Diese Maßnahmen sollten den Einfluss expliziter Abrufstrategien auf implizite Gedächtnisleistungen verringern.

1.6 Ziele der eigenen Arbeit

Das Lebensalter von Personen ist eine Variable, für die häufig Dissoziationen zwischen impliziten und expliziten Gedächtnistests gefunden wurden. In drei Experimenten beispielsweise verglichen Light und Singh (1987) junge und ältere Personen in verschiedenen impliziten und expliziten Gedächtnistests miteinander. In einem ersten Experiment wurden Probanden im impliziten Wortstammergänzungstest, dem freien Abruf von Wörtern und der Rekognition von Wörtern miteinander verglichen. In einem zweiten Experiment wurden Probanden im impliziten Wortstammergänzungstest und der expliziten Form des Wortstammergänzungstests miteinander verglichen. Diese beiden Tests sind – bis auf die Testinstruktionen – identisch aufgebaut und bieten damit eine optimale Vergleichbarkeit. In einem dritten Experiment wurden Probanden – zur Generalisierung der Befunde – in zwei perzeptuellen impliziten Identifikationstests und zwei expliziten Gedächtnistests miteinander verglichen. Dissoziationen zwischen impliziten und expliziten Gedächtnistests wurden in allen drei Experimenten gefunden. Während sich die Leistungen in den expliziten Gedächtnistests zwischen jungen und älteren Probanden unterschieden,

wurden in den impliziten Gedächtnistests lediglich unbedeutende Gruppenunterschiede gefunden. Light und Singh (1987) merkten an, dass die geringen Gruppenunterschiede in den impliziten Gedächtnistests aufgrund zu geringer Teststärke nicht verlässlich abgesichert werden konnten. Diese und ähnliche Befunde von bedeutenden Unterschieden in expliziten Gedächtnistests sowie fehlenden oder unbedeutenden Unterschieden in impliziten Gedächtnistests, sind häufig in der Literatur zu finden. Da Light und Singh (1987) die psychometrischen Eigenschaften der von ihnen verwendeten Tests nicht berichteten, können die Befunde, neben der Interpretation funktioneller Dissoziationen im Sinne unterschiedlicher Gedächtnissysteme (z.B. Tulving & Schacter, 1990) oder Gedächtnisprozesse (z.B. Roediger, 1990), auch auf mangelnde Reliabilität impliziter Gedächtnistests zurückzuführen sein. Zusammen mit der hohen Reliabilität expliziter Gedächtnistests, könnte dies zu Artefakten führen, die sich als Dissoziationen zwischen impliziten und expliziten Gedächtnistests äußern. Da Reliabilitätsschätzungen von Tests üblicherweise nicht berichtet werden, bleibt immer eine Mehrdeutigkeit bei der Interpretation der Befunde (Buchner & Brandt, 2003). Dies ist besonders deshalb der Fall, weil einige Untersuchungen große Reliabilitätsdifferenzen zwischen impliziten und expliziten Gedächtnistests gefunden haben (z.B. Buchner & Brandt, 2003; Buchner & Wippich, 2000; Meier & Perrig, 2000). Aus diesem Grund wäre es interessant herauszufinden, ob Befunde von Dissoziationen – wie sie beispielsweise von Light und Singh (1987) gefunden wurden – nur aufgrund unzureichender Reliabilität impliziter Gedächtnistests zustande kommen oder ob sie auch dann bestehen bleiben, wenn reliable implizite Tests zur Anwendung kommen.

In dem ersten Experiment der vorliegenden Arbeit wurde der implizite Wortstamm-ergänzungstest oder eine weitere Form des impliziten Wortstamm-ergänzungstests, bei dem Maßnahmen zur Optimierung der Reliabilität durchgeführt wurden, Probanden vorgelegt. Zur Beurteilung des Erfolgs der Maßnahmen wurden anschließend die Reliabilitätsschätzungen beider Tests miteinander verglichen. Die Optimierung erfolgte mit Hilfe von Maßnahmen, die zum Ziel hatten, die Variabilität von kognitiven Prozessen im impliziten Wortstamm-ergänzungstest einzuschränken. In dem zweiten Experiment sollte eine klassische Dissoziation zwischen einem impliziten und einem expliziten Gedächtnistest repliziert werden. Dazu wurde jungen und älteren Probanden ein unreliabler impliziter Wortstamm-ergänzungstest oder eine explizite Form des Wortstamm-ergänzungstests vorgelegt. Um die Vergleichbarkeit

zwischen beiden Gedächtnistests zu maximieren, unterschieden sich beide Tests ausschließlich in den Testinstruktionen voneinander, die nur im expliziten Fall einen Bezug zur Lernepisode herstellten. Anschließend sollte festgestellt werden, ob durch Maßnahmen, die zu einer Erhöhung der Reliabilität des impliziten Wortstammergänzungstests im ersten Experiment führten, Unterschiede in den Gruppenmittelwerten junger und älterer Probanden aufgedeckt werden können. In dem dritten Experiment sollten die Befunde aus dem zweiten Experiment auf einen weiteren impliziten und expliziten Gedächtnistest generalisiert werden. Der perzeptuelle Identifikationstest von Bildern in einer Klarifikationsprozedur ist ein impliziter Gedächtnistest mit vergleichsweise restriktiven Testbedingungen. Probanden wird wenig Spielraum zur Anwendung verschiedener Antwortstrategien gelassen, was sich in einer höheren Reliabilität dieses Gedächtnistests widerspiegeln sollte (Buchner & Wippich, 2000). Als expliziter Gedächtnistest wurde der Bilderkennungstest gegenübergestellt.

2 Experiment 1

Dissoziationen zwischen impliziten und expliziten Gedächtnistests werden häufig mit unterschiedlichen Gedächtnissystemen (z.B. Tulving & Schacter, 1990) oder Gedächtnisprozessen (z.B. Roediger, 1990) assoziiert. Dissoziationen können jedoch auch als Artefakte geringer Reliabilität impliziter Gedächtnistests auftreten (Buchner & Wippich, 2000). Bevor über Erklärungsmodelle für die Entstehung von Dissoziationen diskutiert wird, sollten die psychometrischen Gütekriterien dissozierbarer impliziter und expliziter Tests bekannt sein. Die Reliabilität dieser Tests sollte in einer Größenordnung liegen, die sinnvolles Experimentieren ermöglicht. Kann diese Grundannahme nicht gemacht werden, so führt dies dazu, dass eventuell vorhandene Gruppenunterschiede nicht gefunden werden können und Befunde nicht repliziert werden können. Falls es jedoch gelingen sollte, reliable implizite Tests reliablen expliziten Tests gegenüberzustellen, dann sollten Befunde aus Dissoziationsstudien in Zukunft eindeutiger zu interpretieren sein.

2.1 Einleitung

Im ersten Experiment wurde ein unbearbeiteter impliziter Wortstammergänzungstest oder eine im Hinblick auf Reliabilitätssteigerung bearbeitete Form des impliziten Wortstammergänzungstests Probanden vorgelegt. Wie nachfolgend genauer beschrieben, wurden dabei die Enkodierungsbedingungen in der Lernphase und die zur Verfügung stehende Antwortzeit in der Testphase manipuliert. Dazu wurde Probanden in einer Lernphase eine Abfolge von Wörtern dargeboten. Probanden, die dem unbearbeiteten Wortstammergänzungstest zugewiesen worden sind, wurden angewiesen, die Häufigkeit präsentierter Wörter im deutschen Sprachgebrauch zu beurteilen. Probanden, die der im Hinblick auf Reliabilitätssteigerung bearbeiteten Form des Wortstammergänzungstests zugewiesen worden sind, wurden zusätzlich zu der Häufigkeitsbeurteilung angewiesen, die präsentierten Wörter zuerst nachzusprechen. Die sprachliche Verarbeitung der Stimuli diente dazu, die Breite des Spektrums möglicher Enkodierungsprozesse durch Probanden einzuschränken. Dadurch sollte eine Reliabilitätssteigerung des Tests bewirkt werden. Nach einer Ablenkungsaufgabe, die eine Wahrnehmung des Zusammenhangs zwischen Lernphase und

Testphase verhindern sollte, folgte die Testphase. In der Testphase des Experimentes wurden sowohl Wortstämme präsentiert, die mit Wörtern aus der Lernphase ergänzt werden konnten, als auch die gleiche Anzahl von Wortstämmen, die nur mit neuen Wörtern ergänzt werden konnten. Probanden sollten Wortstämme mit dem ersten Wort ergänzen, welches ihnen dazu einfiel. Probanden, die dem unbearbeiteten Wortstammerngänzungstest zugewiesen worden sind, hatten keine Antwortzeitbegrenzung, während Probanden, die dem im Hinblick auf Reliabilitätssteigerung bearbeiteten Wortstammerngänzungstest zugewiesen worden sind, für jede Wortstammerngänzung fünf Sekunden Antwortzeit zur Verfügung hatten. Der Zweck der Antwortzeitbegrenzung war es, durch Einschränkung der Breite des Spektrums möglicher Antwortstrategien eine Reliabilitätssteigerung zu bewirken.

Falls die Maßnahmen zur Optimierung der Reliabilität des impliziten Wortstammerngänzungstests in einer Einschränkung der Variabilität kognitiver Prozesse resultieren, so sollte dies eine Erhöhung der Reliabilität zur Folge haben. Die im Hinblick auf eine Reliabilitätssteigerung überarbeitete Form des impliziten Wortstammerngänzungstests sollte eine höhere Reliabilität aufweisen als die unbearbeitete Form des Tests.

2.2 Methode

In den kommenden Abschnitten wird der Aufbau des ersten Experimentes näher beschrieben. Im Einzelnen wird auf die Stichprobe (2.2.1), das Material (2.2.2), die Prozedur (2.2.3) und auf das Design (2.2.4) des Versuchsaufbaus näher eingegangen.

2.2.1 Stichprobe

An Experiment 1 nahmen 288 Probanden teil, wovon die meisten Studierende der Heinrich-Heine-Universität waren. Von den Probanden waren 214 Personen weiblich und 74 Personen männlich. Die Altersspanne der Probanden reichte von 18 bis 76 Jahren ($M = 26.98$, $SD = 11.71$). Von den Probanden hatten 2 Personen den Hauptschul- oder Volksschulabschluss, eine Person den Realschulabschluss und 285 Personen das Abitur. Gute Deutschkenntnisse wurden für die Teilnahme an dem Experiment vorausgesetzt. Die Probanden wurden einzeln getestet und randomisiert dem

unbearbeiteten Wortstammergänzungstest oder dem im Hinblick auf Reliabilitätssteigerung bearbeiteten Wortstammergänzungstest zugewiesen. Ebenfalls randomisiert war die Zuweisung der Probanden zu Wortliste 1 oder zu Wortliste 2. Probanden, die Wortliste 1 zugewiesen worden sind, bekamen in der Lernphase Wörter der Wortliste 1 präsentiert, während Probanden, die Wortliste 2 zugewiesen worden sind, Wörter der Wortliste 2 präsentiert bekamen. Die Variable «Wortliste» diente zur Bestimmung der Basisrate und ging nicht in das eigentliche Design des Experimentes mit ein. Insgesamt gab es damit folgende vier Bedingungen mit gleicher Probandenanzahl: *unbearbeiteter Wortstammergänzungstest, Wortliste 1 (n = 72)*; *unbearbeiteter Wortstammergänzungstest, Wortliste 2 (n = 72)*; *im Hinblick auf Reliabilitätssteigerung bearbeiteter Wortstammergänzungstest, Wortliste 1 (n = 72)*; und *im Hinblick auf Reliabilitätssteigerung bearbeiteter Wortstammergänzungstest, Wortliste 2 (n = 72)*. Die Probanden wurden am Ende des Experimentes durch Bezahlung entlohnt.

2.2.2 Material

Als Stimulusmaterial wurden 184 Hauptwörter verwendet. Die Wörter wurden so ausgewählt, dass ihre Wortstämme, die aus den ersten drei Buchstaben der jeweiligen Wörter bestanden, nicht mehrfach vorkamen. Weiterhin wurde darauf geachtet, dass Wortstämme mit mindestens fünf voneinander verschiedenen Hauptwörtern ergänzt werden konnten. Es gab zwei Wortlisten, wobei die Zuweisung der Wörter zu den beiden Wortlisten randomisiert erfolgte. Dazu wurden zunächst alle Wörter in einer alphabetischen Reihenfolge angeordnet. Anschließend wurden immer jeweils zwei aufeinanderfolgende Wörter der Anordnung randomisiert, entweder der ersten Wortliste oder der zweiten Wortliste zugeordnet. In der Lernphase wurden Wörter einer der beiden Wortlisten präsentiert, je nachdem welcher Wortliste Probanden zugeordnet worden sind. Zusätzlich zu diesen Wörtern gab es in der Lernphase acht Übungswörter am Anfang jeder Wortliste. Zwischen Lernphase und Testphase kam zur Ablenkung der Probanden eine Aufmerksamkeitssaufgabe zum Einsatz, die in keinem Bezug zu den hier untersuchten Gedächtnistests stand. In der Testphase wurden Wortstämme beider Wortlisten präsentiert. Die Wortstämme der Übungswörter wurden der Präsentation der übrigen Wortstämme vorangestellt. Alle Inhalte des Experimentes sowie die Instruktionen, wurden auf einem Apple iMac Computer mit einer Bildschirmdiagonalen von 20 Zoll präsentiert. Die Probanden

trugen Kopfhörer als Schallschutz vor ablenkenden Nebengeräuschen. Das Experiment fand unter lärmberuhigten Bedingungen statt.

2.2.3 Prozedur

Nach Eingabe der demographischen Daten wurde die Studie den Probanden als eine Normierungsstudie zur subjektiven Einschätzung von Worthäufigkeiten im deutschen Sprachgebrauch vorgestellt. Die Probanden wurden einzeln getestet. Die Lernphase des Experimentes bestand für alle Probanden aus der Präsentation von Wörtern aus einer der beiden Wortlisten, je nachdem zu welcher Wortliste Probanden zugeordnet worden sind. Die Abfolge der Wörter in der Lernphase sowie die Abfolge der Wortstämme in der Testphase waren für jede Person randomisiert. Wörter in der Lernphase wurden für eine Dauer von jeweils vier Sekunden präsentiert, gefolgt von einem leeren Bildschirm, der 500 ms andauerte. Nach Ablauf dieser Zeit wurde das nächste Wort präsentiert. Am Anfang einer Wortliste wurde eine Übungsliste mit acht Wörtern präsentiert. Dies diente dazu, dass Probanden sich mit der Übung vertraut machen konnten. Unter jedem Wort war das Symbol eines Mikrofons eingeblendet, um die Probanden auf ihre Rückmeldung aufmerksam zu machen. Die Probanden beider Versuchsbedingungen wurden angewiesen, verbal zu beurteilen, ob das jeweils präsentierte Wort «selten» oder «häufig» im deutschen Sprachgebrauch vorkommt. Diese Anweisung diente dazu, die Aufmerksamkeit der Probanden aufrecht zu erhalten, um eine vollständige Enkodierung des Stimulussets zu gewährleisten. Probanden, die dem unbearbeiteten Wortstammergänzungstest zugewiesen worden sind, erhielten lediglich die Anweisung, die Worthäufigkeit von Wörtern im deutschen Sprachgebrauch zu beurteilen. Probanden, die dem im Hinblick auf Reliabilitätssteigerung bearbeiteten Wortstammergänzungstest zugewiesen worden sind, erhielten zusätzlich dazu die Anweisung, die Wörter zuerst laut auszusprechen. Den Probanden des im Hinblick auf Reliabilitätssteigerung bearbeiteten Wortstammergänzungstests wurde mitgeteilt, dass ihre Antwortzeiten erfasst werden und sie möglichst zügig und spontan antworten sollen. In dieser Phase des Experimentes gab es jedoch keine Sprachaufzeichnung. Diese Anweisung diente lediglich dazu, eine Leistungsatmosphäre zu erzeugen. Die Lernphase erstreckte sich über eine Dauer von etwa sechs Minuten.

Unmittelbar nach der Lernphase folgte eine Ablenkungsaufgabe, in der Probanden

auf ein Fixationskreuz schauen und möglichst schnell auf visuelle Reize in der Peripherie reagieren sollten. Sie wurde so ausgewählt, dass sie nicht mit Wörtern aus der Lernphase interferieren konnte. Nach Beendigung der Ablenkungsaufgabe, die eine Dauer von etwa 20 Minuten hatte, folgte die Testphase.

In der Testphase des Experimentes wurden den Probanden sowohl die Wortstämme von Wörtern der Wortliste aus der Lernphase als auch die Wortstämme von Wörtern der neuen Wortliste präsentiert. Je nachdem, welcher Wortliste Probanden in der Lernphase zugewiesen worden sind, waren die neuen Wortstämme aus der jeweils anderen Wortliste. Die Probanden beider Versuchsbedingungen wurden angewiesen, die Wortstämme mit dem jeweils ersten Wort zu ergänzen, welches ihnen dazu einfällt. Die Probanden des im Hinblick auf Reliabilitätssteigerung bearbeiteten Wortstammerngänzungstests wurden zusätzlich dazu aufgefordert, die Wortstämme spontan und zügig zu ergänzen. Ihnen wurde mitgeteilt, dass ihre Antwortzeiten erfasst werden. Mit den Wortstämmen erschien ebenfalls das Symbol eines Mikrofons im unteren Teil des Bildschirms. Die Antworten der Probanden wurden aufgezeichnet. Die Sprachaufzeichnung der Wortstammerngänzungen diente der späteren Auswertung des Experimentes. Wortstämme sollten nur mit deutschen Hauptwörtern und mit mindestens zwei weiteren Buchstaben ergänzt werden. Namen und Fremdwörter sollten nicht genannt werden. Am Anfang der Darbietung erschienen die Wortstämme der acht Übungswörter, gefolgt von den Wortstämmen der Wörter beider Wortlisten. Probanden, die dem unbearbeiteten Wortstammerngänzungstest zugewiesen worden sind, wurde keine Zeitbegrenzung für die Ergänzung der Wortstämme eingeräumt. Sie wurden angewiesen, nach der Ergänzung eines Wortstamms den nächsten Durchgang durch Tastendruck auf die Leertaste selbstständig einzuleiten. Es folgte darauf ein leerer Bildschirm für die Dauer von 500 ms, bevor der nächste Durchgang begann. Probanden, die dem im Hinblick auf Reliabilitätssteigerung bearbeiteten Wortstammerngänzungstest zugeordnet worden sind, wurde die Antwortzeit zur Ergänzung von Wortstämmen auf fünf Sekunden begrenzt. Nach Ablauf dieser Zeit wurde die Aufzeichnung beendet und nach einem Intervall von 500 ms, in dem ein leerer Bildschirm eingeblendet wurde, der nächste Durchgang automatisch eingeleitet. Die Testphase nahm etwa 30 Minuten in Anspruch.

Insgesamt dauerte das Experiment etwa eine Stunde. Nach Beendigung der Testphase wurden die Probanden über die Hintergründe des Experimentes informiert und

darum gebeten, für die Dauer von einem Monat nicht mit anderen Studierenden darüber zu sprechen.

2.2.4 Design

Dem ersten Experiment lag ein Versuchsplan mit unabhängigen Zufallsgruppen zugrunde. Die unabhängige Variable war die Testform (*unbearbeitete Form des Wortstammergänzungstests* vs. *im Hinblick auf Reliabilitätssteigerung bearbeitete Form des Wortstammergänzungstests*). Die abhängige Variable war das Primingmaß der beiden impliziten Tests.¹ Der Vergleich der Reliabilität beider Formen des Wortstammerngänzungstests war jedoch eigentlicher Gegenstand der Untersuchung. Als Schätzungen der Reliabilität wurden Testhalbierungskoeffizienten ermittelt.

Cohen's q stellt eine Effektgröße für Korrelationsunterschiede dar, die über die Differenz zweier Fisher Z -transformierter Korrelationen bestimmt wird. Cohen (1977) unterschied dabei kleine, mittlere und große Effekte ($q = 0.10$, $q = 0.30$ und $q = 0.50$). Wenn «mittelgroße» Unterschiede zwischen Korrelationen in der Population aufgedeckt werden sollen ($q = 0.40$, nach Cohen, 1977), der Fehler 1. Art auf $\alpha = .05$ und der Fehler 2. Art auf $\beta = .05$ festgelegt wird, dann werden bei dem einseitigen Testproblem nach einer a priori Teststärkeanalyse mindestens $n = 139$ Probanden in jeder der beiden Testbedingungen benötigt. Tatsächlich konnten Daten von $n = 144$ Probanden pro Bedingung bzw. von $N = 288$ Probanden insgesamt erhoben werden. Damit konnte eine Teststärke von $1 - \beta = .96$ erreicht werden. Alle Teststärkeberechnungen in den vorliegenden Experimenten wurden mit dem Programm G*Power durchgeführt (Faul, Erdfelder, Buchner & Lang, 2009; Faul, Erdfelder, Lang & Buchner, 2007). Das Signifikanzniveau wurde für alle Analysen der vorliegenden Arbeit auf $\alpha = .05$ festgelegt. Der Vergleich von Korrelationskoeffizienten in den Experimenten der vorliegenden Arbeit beruht auf einseitigen Testproblemen. Alle übrigen statistischen Tests wurden zweiseitig durchgeführt.

¹ Die Definition des Primingmaßes, welches zur Erfassung der Gedächtnisleistungen im impliziten Wortstammerngänzungstest Verwendung fand, folgt bei der Darstellung der Ergebnisse.

2.3 Ergebnisse

Zur Leistungserfassung wurde für beide implizite Wortstammergänzungstests ein Primingmaß verwendet. Probanden wurde in der Lernphase, je nachdem welcher Wortliste sie zugewiesen worden sind, eine von zwei Wortlisten präsentiert. In der Testphase wurden sowohl die Wortstämme von Wörtern der «alten» Wortliste aus der Lernphase als auch die Wortstämme von Wörtern einer nicht in der Lernphase präsentierten «neuen» Wortliste zur Ergänzung vorgelegt. Die Bearbeitung der Wortstämme von Wörtern der «alten» Wortliste diente zur Erfassung der Rohwerte der impliziten Gedächtnisleistung. Dazu wurde der Anteil derjenigen Wortstämme bestimmt, die mit Wörtern der «alten» Wortliste aus der Lernphase ergänzt worden sind, an allen Wortstämmen von Wörtern der «alten» Wortliste. Die Bearbeitung der Wortstämme von Wörtern der «neuen» Wortliste diente zur Ermittlung der Basisrate, welche zur Bestimmung des Primingwertes vom Rohwert abgezogen wurde. Eine Korrektur der Basisrate ist deshalb notwendig, weil Wortstämme von Wörtern der «alten» Wortliste mit einer bestimmten Wahrscheinlichkeit auch spontan mit «alten» Wörtern ergänzt werden, selbst wenn keine Erfahrungsnachwirkungen vorhanden sind. Zur Schätzung der Höhe dieser Wahrscheinlichkeit (der Basisrate) bedient man sich der Rate spontaner Ergänzungen von Wortstämmen der «neuen» Wortliste mit Wörtern der «neuen» Wortliste. Dazu wurde der Anteil derjenigen Wortstämme bestimmt, die mit Wörtern der «neuen» Wortliste ergänzt worden sind, an allen Wortstämmen von Wörtern der «neuen» Wortliste. Über die Differenz von Rohwert und Basisrate wurde für jeden Probanden ein Primingwert ermittelt. Der Primingwert der Skala wurde als Mittelwert der Primingwerte aller Probanden bestimmt.

Neben der bereits beschriebenen Methode zur Bestimmung der Leistungswerte² – wie sie in der Literatur gebräuchlich ist – wurde hier auch eine alternative Methode eingeführt. Die gebräuchliche Methode der Bestimmung der Leistungswerte wird üblicherweise aus dem Grund verwendet, weil sie eine parallele Methode zur Bestimmung des Sensitivitätsmaßes P_r des Zwei-Hochschwellen-Modells (Snodgrass & Corwin, 1988) darstellt. Das Zwei-Hochschwellen-Modell basiert auf der Vorstellung einer Entdeckungsschwelle für «alte» Items und einer Entdeckungsschwelle für

² Wenn von der Methode der Bestimmung von Leistungswerten gesprochen wird, dann ist damit sowohl die Bestimmung der Primingwerte impliziter Tests gemeint, als auch die Bestimmung von Leistungswerten expliziter Tests nachfolgender Experimente.

«neue» Items, welche zur Rekognition der Items überschritten werden muss. Analog zum Sensitivitätsmaß P_r , welches häufig als Leistungsmaß für Rekognitionstests herangezogen wird und dessen Sensitivitätswerte durch die Differenz der Trefferrate und der Rate falscher Alarme geschätzt werden, werden auch die Primingwerte des Primingmaßes ermittelt. Die Ermittlung der Basisrate bei der bereits beschriebenen Methode erfolgte an denselben Probanden, an denen auch die Rohwerte ermittelt wurden.³ Somit erfolgte die Bestimmung von Rohwert und Basisrate an Wortstämmen *unterschiedlicher* Wortlisten. Wie im nächsten Abschnitt näher erläutert, erschien es sinnvoll eine alternative Methode der Leistungswertbestimmung einzuführen, bei der die Basisraten an Wortstämmen der gleichen Wortliste ermittelt werden, an denen auch Rohwerte ermittelt werden. Der Hintergrund dieses Gedankens war die Vermutung, dass durch den Gebrauch unterschiedlicher Wortlisten zur Ermittlung von Primingwerten bei der gebräuchlichen Methode weitere unsystematische Varianz in die Rohwertdaten mit eingebracht werden könnte, weil unterschiedliche Wortlisten nicht unbedingt in allen Eigenschaften übereinstimmen müssen. Ein höherer Anteil an Fehlervarianz führt zu einem geringeren Anteil der Varianz der wahren Werte an der Gesamtvarianz und damit zu einer geringeren Reliabilität des Wortstammerngänzungstests. Die beiden Methoden unterscheiden sich in der Ermittlung der Basisrate voneinander. Bei der alternativen Methode wurde die Basisrate an Probanden ermittelt, die der jeweils anderen Wortliste zugewiesen worden sind. Zur Primingwertbestimmung wurden Probanden beider Wortlisten so gepaart, dass beispielsweise die Basisrate des ersten Probanden einer Wortliste von dem Rohwert des ersten Probanden der jeweils anderen Wortliste abgezogen wurde. Die gleiche Prozedur wurde für alle aufeinander folgenden Paare von Probanden beider Wortlisten durchgeführt. Somit erfolgte die Bestimmung von Rohwert und Basisrate an Wortstämmen der *gleichen* Wortliste. Dies könnte möglicherweise – bedingt durch einen geringeren Fehlervarianzanteil der Daten – zu einer höheren Reliabilitätsschätzung des impliziten Wortstammerngänzungstests führen. Der Primingwert der Skala wurde auch hier als Mittelwert der Primingwerte aller Probanden bestimmt.

Anschließend wurden die Tests in zwei Hälften geteilt. Dazu wurden Items mit gerader Nummerierung einer Testhälfte zugewiesen, während Items mit ungerader

³ Die Ermittlung der Rate falscher Alarme zur Bestimmung des Sensitivitätsmaßes P_r des Zwei-Hochschwellen-Modells erfolgt im Übrigen auch an denselben Personen, an denen die Trefferrate ermittelt wird.

Nummerierung einer zweiten Testhälfte zugewiesen wurden. Für jeden Probanden wurden somit zwei Primingwerte ermittelt, jeweils einer pro Testhälfte. Die Testhalbierungskoeffizienten wurden durch die Korrelationen der Primingwerte beider Testhälften geschätzt.

Tabelle 1

Gruppenmittelwerte des Primingmaßes, für beide Testformen und für beide Methoden der Leistungswertbestimmung, in Experiment 1 (Die Werte in Klammern repräsentieren die Standardfehler der Mittelwerte).

	Wortliste 1 Probanden	Wortliste 2 Probanden	Alle Probanden
unbearbeiteter Wortstammergeänzungstest Primingwert (gebräuchliche Methode)	.22 (.01)	.08 (.01)	.15 (.01)
im Hinblick auf Reliabilitätssteigerung bearbeiteter Wortstammergeänzungstest Primingwert (gebräuchliche Methode)	.24 (.01)	.10 (.01)	.17 (.01)
unbearbeiteter Wortstammergeänzungstest Primingwert (alternative Methode)	.16 (.01)	.14 (.01)	.15 (.01)
im Hinblick auf Reliabilitätssteigerung bearbeiteter Wortstammergeänzungstest Primingwert (alternative Methode)	.18 (.01)	.16 (.01)	.17 (.01)

Die Primingwerte beider Testformen sind in Tabelle 1 aufgeführt.⁴ Neben den Primingwerten, die mit der gebräuchlichen Methode bestimmt wurden, sind auch die Primingwerte aufgeführt, die mit der alternativen Methode bestimmt wurden. Einstichproben-t-Tests ergaben, dass die Primingwerte von Probanden beider Wortlisten, für beide Testformen und für beide Methoden der Leistungswertbestimmung, statistisch signifikant von null verschieden waren, $ts(71) > 9.10$, $ps < .001$, $\eta_p^2s > .54$. Die über Probanden beider Wortlisten gemittelten Primingwerte waren ebenfalls statistisch signifikant von null verschieden, $ts(143) > 16.02$, $ps < .001$, $\eta_p^2s > .64$. Die Primingwerte unterschieden sich zwischen Wortliste 1 und Wortliste 2 Probanden statistisch signifikant voneinander, wenn die gebräuchliche Methode der Leistungswertbestimmung angewandt wurde. Dies traf sowohl für den unbearbeiteten Wort-

⁴ Die Variablen, die den Primingwerten der entsprechenden Tests zugrunde liegen, sind der Tabelle A im Anhang zu entnehmen.

stammergänzungstest zu, $t(142) = 12.35, p < .001, \eta^2 = .52$ ($M_s = .22$ vs. $.08, SE_s = .01$ vs. $.01$), als auch für den im Hinblick auf Reliabilitätssteigerung bearbeiteten Wortstammergänzungstest, $t(142) = 11.18, p < .001, \eta^2 = .47$ ($M_s = .24$ vs. $.10, SE_s = .01$ vs. $.01$). Wurde die alternative Methode der Leistungswertbestimmung angewandt, so waren keine statistisch signifikanten Unterschiede zwischen den Primingwerten von Probanden beider Wortlisten vorhanden. Dies traf sowohl für den unbearbeiteten Wortstammergänzungstest zu, $t(142) = 1.23, p = .220, \eta^2 = .01$ ($M_s = .16$ vs. $.14, SE_s = .01$ vs. $.01$), als auch für den im Hinblick auf Reliabilitätssteigerung bearbeiteten Wortstammergänzungstest, $t(142) = 1.07, p = .285, \eta^2 < .01$ ($M_s = .18$ vs. $.16, SE_s = .01$ vs. $.01$).

Tabelle 2

Testhalbierungskoeffizienten des Primingmaßes, für beide Testformen und für beide Methoden der Leistungswertbestimmung, in Experiment 1 (Die Werte in Klammern repräsentieren die Spearman-Brown korrigierten Koeffizienten für die volle Testlänge).

	Wortliste 1 Probanden	Wortliste 2 Probanden	Alle Probanden
unbearbeiteter Wortstammergänzungstest (r_H) (gebräuchliche Methode)	.09	.21	.15 (.26)
im Hinblick auf Reliabilitätssteigerung bearbeiteter Wortstammergänzungstest (r_H) (gebräuchliche Methode)	.24	.24	.24 (.39)
unbearbeiteter Wortstammergänzungstest (r_H) (alternative Methode)	.52	.61	.57 (.73)
im Hinblick auf Reliabilitätssteigerung bearbeiteter Wortstammergänzungstest (r_H) (alternative Methode)	.55	.65	.60 (.75)

In Tabelle 2 sind die Testhalbierungskoeffizienten beider Testbedingungen aufgeführt.⁵ Es sind sowohl Testhalbierungskoeffizienten aufgeführt, die aus Primingwerten bestimmt wurden, die mit der gebräuchlichen Methode ermittelt wurden, als auch Testhalbierungskoeffizienten, die aus Primingwerten bestimmt wurden, die mit der alternativen Methode ermittelt wurden. Die Testhalbierungskoeffizienten von

⁵ Die statistischen Kennwerte, die den Testhalbierungskoeffizienten von Experiment 1 zugrunde liegen, sind Tabelle 3 zu entnehmen.

Probanden beider Wortlisten wurden gemittelt, um die Teststärke für den entscheidenden Vergleich zwischen den Testbedingungen zu erhöhen. Da Korrelationen nicht intervallskaliert sind, wurde zunächst eine Fisher Z-Transformation durchgeführt. Nach arithmetischer Mittelung der Fisher Z-Werte, erfolgte die Rücktransformation zu Korrelationskoeffizienten. In der letzten Spalte von Tabelle 2 sind sowohl die gemittelten Testhalbierungskoeffizienten als auch die Spearman-Brown korrigierten Reliabilitätsschätzungen für den gesamten Test aufgeführt. Für den Vergleich der Reliabilitätsschätzungen zwischen den Testbedingungen wurden die Testhalbierungskoeffizienten herangezogen. Wenn die gebräuchliche Methode der Leistungswertbestimmung angewandt wurde, war der Testhalbierungskoeffizient des im Hinblick auf Reliabilitätssteigerung bearbeiteten Wortstammergänzungstests (.24) nicht statistisch signifikant größer als der Testhalbierungskoeffizient des unbearbeiteten Wortstammergänzungstests (.15), $z = 0.79$, $p > .21$, $q = 0.09$. Auch wenn die alternative Methode der Leistungswertbestimmung angewandt wurde, war der Testhalbierungskoeffizient des im Hinblick auf Reliabilitätssteigerung bearbeiteten Wortstammergänzungstests (.60) nicht statistisch signifikant größer als der Testhalbierungskoeffizient des unbearbeiteten Wortstammergänzungstests (.57), $z = 0.38$, $p > .35$, $q = 0.05$. Wenn die Leistungswerte mit der alternativen Methode bestimmt wurden, war der Testhalbierungskoeffizient des unbearbeiteten Wortstammergänzungstests (.57) statistisch signifikant größer, als wenn die Leistungswerte mit der gebräuchlichen Methode bestimmt wurden (.15), $z = 4.17$, $p < .001$, $q = 0.50$. Wenn die Leistungswerte mit der alternativen Methode bestimmt wurden, war der Testhalbierungskoeffizient des im Hinblick auf Reliabilitätssteigerung bearbeiteten Wortstammergänzungstests (.60) statistisch signifikant größer, als wenn die Leistungswerte mit der gebräuchlichen Methode bestimmt wurden (.24), $z = 3.76$, $p < .001$, $q = 0.45$.

In Tabelle 3 sind die Kovarianzen zwischen den Testhälften und die Standardabweichungen der Testhälften aufgeführt, die den Testhalbierungskoeffizienten des ersten Experimentes zugrunde liegen. Man kann der Tabelle deskriptiv entnehmen, wie sich die Kovarianzen zwischen der gebräuchlichen Methode der Leistungswertbestimmung und der alternativen Methode der Leistungswertbestimmung unterscheiden. Wurde die alternative Methode der Leistungswertbestimmung angewandt, so waren die Kovarianzen für Probanden beider Wortlisten in beiden Testformen höher, als wenn die gebräuchliche Methode der Leistungswertbestimmung angewandt wurde. Zu erkennen ist ebenfalls ein leichter Anstieg der Standardabweichungen der

Testhälften für Probanden beider Wortlisten in beiden Testformen, von der gebräuchlichen Methode der Leistungswertbestimmung zur alternativen Methode der Leistungswertbestimmung.

Tabelle 3

Kovarianzen zwischen den Testhälften und die Standardabweichungen der Testhälften, wie sie für die Schätzung der Reliabilität in Experiment 1 verwendet wurden. Aufgeführt sind die Werte für beide Testformen und für beide Methoden der Leistungswertbestimmung.

	Wortliste 1 Probanden	Wortliste 2 Probanden
unbearbeiteter Wortstammerngänzungstest (gebräuchliche Methode)		
$COV_{Th1,Th2}$	0.001	0.002
SD_{Th1}	0.09	0.09
SD_{Th2}	0.09	0.09
im Hinblick auf Reliabilitätssteigerung bearbeiteter Wortstammerngänzungstest (gebräuchliche Methode)		
$COV_{Th1,Th2}$	0.002	0.003
SD_{Th1}	0.09	0.11
SD_{Th2}	0.09	0.10
unbearbeiteter Wortstammerngänzungstest (alternative Methode)		
$COV_{Th1,Th2}$	0.008	0.010
SD_{Th1}	0.13	0.12
SD_{Th2}	0.11	0.14
im Hinblick auf Reliabilitätssteigerung bearbeiteter Wortstammerngänzungstest (alternative Methode)		
$COV_{Th1,Th2}$	0.008	0.010
SD_{Th1}	0.13	0.12
SD_{Th2}	0.12	0.13

2.4 Diskussion

Ziel des ersten Experimentes war der Vergleich der Reliabilität eines impliziten Wortstammerngänzungstests, der im Hinblick auf eine Reliabilitätssteigerung bearbeitet worden ist, mit der Reliabilität eines unbearbeiteten impliziten Wortstammerngänzungstests. Eine Maßnahme zur Reliabilitätssteigerung war die Einschränkung der Variabilität der Enkodierungsprozesse in der Lernphase, indem Probanden angewiesen wurden, die visuell präsentierten Wörter laut auszusprechen. Eine andere Maßnahme war die Einschränkung der Variabilität der Antwortstrategien durch Begrenzung der Zeit, die für die Ergänzung von Wortstämmen in der Testphase zur Verfügung stand. Beide Maßnahmen zur Einschränkung der Variabilität von Gedächtnisprozessen beim impliziten Wortstammerngänzungstest führten nicht zu der vorausgesagten Reliabilitätssteigerung, unabhängig davon, mit welcher Methode die Leistungswerte bestimmt wurden. Inferenzstatistisch konnte kein statistisch signifikanter Unterschied zwischen den Reliabilitätsschätzungen des unbearbeiteten Wortstammerngänzungstests und des im Hinblick auf Reliabilitätssteigerung bearbeiteten Wortstammerngänzungstests festgestellt werden. Allein auf deskriptiver Ebene war eine Reliabilitätssteigerung zu verzeichnen. Mit einer Effektstärke von $q = 0.09$, wenn Leistungswerte mit der gebräuchlichen Methode bestimmt wurden, und einer Effektstärke von $q = 0.05$, wenn Leistungswerte mit der alternativen Methode bestimmt wurden, handelte es sich dabei jedoch um vernachlässigbar kleine Effekte. Dies sind Effekte kleiner als $q = 0.10$, nach Cohen (1977).

Bezüglich der Primingwerte wurden keine bestimmten Hypothesen aufgestellt. Auffallend jedoch war, dass sich die Primingwerte trotz vollständiger Randomisierung bei der Zuweisung von Items zu Wortlisten und bei der Zuweisung von Probanden zu Versuchsbedingungen zwischen Wortliste 1 und Wortliste 2 Probanden statistisch signifikant voneinander unterschieden, wenn die gebräuchliche Methode der Leistungswertbestimmung angewandt wurde. Die Primingwerte für Wortliste 2 Probanden waren statistisch signifikant geringer, als die Primingwerte für Wortliste 1 Probanden. Werden die Primingwerte des Wortstammerngänzungstests in ihre Bestandteile zerlegt, so fällt auf, dass in den Rohwerten und in den Basisraten Unterschiede zwischen Wortliste 1 und Wortliste 2 Probanden bestehen. In Tabelle A im Anhang sind die Rohwerte und die Basisraten aufgeführt, wie sie für die Bestimmung der Primingwerte in Tabelle 1 verwendet wurden. Liste 1 Wortstämme wur-

den eher mit den entsprechenden Wörtern aus der Lernphase ergänzt als Liste 2 Wortstämme. Liste 1 Wortstämme hatten jedoch ebenfalls eine höhere Basisrate als Liste 2 Wortstämme. Die hohen Differenzen der Primingwerte zwischen den Probanden beider Wortlisten können bei der in der Literatur gebräuchlichen Methode der Leistungswertbestimmung entstehen, bei der die Basisrate an Itembeantwortungen der jeweils anderen Wortliste ermittelt wird. Bei der Bestimmung von Primingwerten einzelner Probanden werden an Liste 2 (Liste 1) Wortstämmen ermittelte Basisraten von Rohwerten abgezogen, die an Liste 1 (Liste 2) Wortstämmen ermittelt werden. Unterscheiden sich beide Wortlisten jedoch insofern voneinander, dass Liste 1 Wortstämme mit höherer Wahrscheinlichkeit mit Wörtern aus der Lernphase ergänzt werden als Liste 2 Wortstämme und Liste 2 Wortstämme eine geringere Basisrate als Liste 1 Wortstämme haben, so kommt es zu den genannten Unterschieden zwischen den Primingwerten von Probanden beider Wortlisten. Buchner und Wipich (2000) ermittelten in ihrem Experiment 1a die Testhalbierungskoeffizienten des impliziten Wortstammerngänzungstests in einer ähnlichen Untersuchung wie der hier vorliegenden. Sie verwendeten dafür die gleichen Wortlisten und fanden ähnlich hohe Primingwertunterschiede zwischen den Probanden beider Wortlisten.

Die beschriebenen Unterschiede zwischen den Primingwerten von Probanden beider Wortlisten führten zu der Suche nach einer alternativen Methode der Leistungswertbestimmung. Die grundsätzliche Überlegung dabei war, dass es sinnvoll sein könnte, Basisraten an Wortstämmen der gleichen Wortliste zu ermitteln, an denen auch Rohwerte ermittelt werden. Die Berechnung von Produkt-Moment-Korrelationen, zur Schätzung der Testhalbierungsreliabilität, erfordert die Kenntnis der Kovarianz zwischen beiden Testhälften und die Kenntnis der Standardabweichungen beider Testhälften. Die Kovarianz ist im Zähler des Bruchs zur Berechnung der Korrelation vertreten, während das Produkt der Standardabweichungen beider Testhälften im Nenner steht ($r = cov_{Th1,Th2} / S_{Th1} * S_{Th2}$). Je höher das Ausmaß der gemeinsamen Variation der beiden Testhälften, desto größer wird der Zähler und desto höher wird die Reliabilitätsschätzung. Je größer die Varianzen innerhalb der Testhälften, desto größer wird der Nenner und desto geringer wird die Reliabilitätsschätzung. Die Verwendung unterschiedlicher Wortlisten zur Ermittlung von Primingwerten, wie dies bei der gebräuchlichen Methode geschieht, könnte die Rohwertdaten mit zusätzlicher unsystematischer Varianz beaufschlagen, was eine geringere Reliabilität des Wortstammerngänzungstests zur Folge hätte. Unsystematische Varianz in den Daten

führt zu geringerer Kovarianz zwischen den Testhälften und zu einer Erhöhung der Varianz innerhalb der Testhälften. Da beide Wortlisten – trotz vollständiger Randomisierung – nicht in allen ihren Eigenschaften vergleichbar sind, kann es bei der gebräuchlichen Methode der Bestimmung der Leistungswerte zu geringen Reliabilitätsschätzungen kommen. Unterschiede in den Eigenschaften der Wortlisten spiegeln sich in unterschiedlichen Rohwerten und Basisraten zwischen Probanden beider Wortlisten wider. Basisraten sollten deshalb an Wortstämmen der gleichen Wortliste ermittelt werden, an denen auch die Rohwerte ermittelt wurden, was mit der alternativen Methode der Leistungswertbestimmung umgesetzt werden kann.

Wenn die alternative Methode der Leistungswertbestimmung angewandt wurde, konnten keine statistisch signifikanten Unterschiede zwischen den Primingwerten von Probanden beider Wortlisten aufgedeckt werden. Wie bereits beschrieben, wurden Liste 1 Wortstämmen mit höherer Wahrscheinlichkeit mit Wörtern aus der Lernphase ergänzt als Liste 2 Wortstämmen. Liste 1 Wortstämmen hatten jedoch auch eine höhere Basisrate als Liste 2 Wortstämmen. Werden Rohwerte und Basisraten an Wortstämmen der gleichen Wortlisten ermittelt, wie dies bei der alternativen Methode der Leistungswertbestimmung geschieht, so gleichen sich diese Unterschiede gegenseitig aus (siehe Tabelle A im Anhang). Tatsächlich konnte eine statistisch signifikante Steigerung der Reliabilität des Wortstammerngänzungstests ermittelt werden, wenn die alternative Methode der Leistungswertbestimmung angewandt wurde, im Vergleich zu der Anwendung der gebräuchlichen Methode der Leistungswertbestimmung. Die Kovarianzen zwischen den Testhälften beider Wortlisten und die Standardabweichungen der Testhälften beider Wortlisten, wie sie zur Schätzung der Testhalbierungskoeffizienten in Tabelle 2 herangezogen wurden, sind in Tabelle 3 aufgeführt. Man kann der Tabelle entnehmen, dass die zu korrelierenden Testhälften deutlich besser kovariieren, wenn zur Primingwertbestimmung Rohwerte und Basisraten an Wörtern derselben Wortliste ermittelt wurden, wie dies bei der alternativen Methode der Bestimmung der Leistungswerte der Fall ist. Wurde jedoch die gebräuchliche Methode der Bestimmung der Leistungswerte angewandt, bei der Rohwerte und Basisraten an Wörtern unterschiedlicher Wortlisten ermittelt werden, dann waren die Kovarianzen zwischen den Testhälften deutlich geringer, als wenn die alternative Methode der Leistungswertbestimmung angewandt wurde. Dies lässt darauf schließen, dass durch die Basisratenkorrektur der alternativen Methode der Leistungswertbestimmung, bei der Rohwerte und Basisraten an Wortstämmen der

gleichen Wortlisten ermittelt werden, Rohwertdaten weniger durch Fehlervarianz beaufschlagt werden. Da die Höhe der Kovarianz in direktem Zusammenhang zu der Höhe der Korrelation steht, führt dies schließlich zu einer höheren Schätzung der Testreliabilität. Man kann der Tabelle 3 ebenfalls entnehmen, dass die Standardabweichungen von der gebräuchlichen Methode zur alternativen Methode der Leistungswertbestimmung leicht ansteigen. Wegen dem vergleichsweise hohen Anstieg der Kovarianzen von der gebräuchlichen Methode zur alternativen Methode der Leistungswertbestimmung, wirkt sich dies jedoch nicht negativ auf die entsprechenden Korrelationskoeffizienten aus.

Der Abzug von Basisratengruppenmittelwerten von Rohwerten einzelner Probanden als weitere Variante der Leistungswertbestimmung ist als nicht sinnvoll einzustufen. Die Reliabilitätsschätzungen wären die gleichen hohen Schätzungen wie bei der Schätzung der Reliabilität direkt aus den Rohwerten, da sich durch Abzug einer Konstanten die Kovarianz und die Varianzen der zu korrelierenden Rohwertpaare nicht ändern. Wie bereits an anderer Stelle beschrieben, sind Rohwerte, sowie Rohwerte abzüglich einer Konstanten, wahrscheinlich mit Antworttendenzen behaftet. Und deren hohe Reliabilitätsschätzungen, wie sie auch in dem vorliegenden Experiment gefunden wurden, sind sehr wahrscheinlich Ergebnis dieser Antworttendenzen (Buchner & Brandt, 2003).

3 Experiment 2

In Experiment 1 gelang es nicht, durch Maßnahmen zur Einschränkung der Variabilität von Gedächtnisprozessen die Reliabilität des impliziten Wortstammergänzungstests zu verbessern. Eine alternative Methode der Leistungswertbestimmung jedoch, bei der Basisrate und Rohwerte an Wortstämmen der gleichen Wortliste ermittelt werden, brachte eine statistisch signifikante Steigerung der Reliabilität im Vergleich zu der gebräuchlichen Methode, bei der Basisrate und Rohwerte an Wortstämmen unterschiedlicher Wortlisten ermittelt werden. In Experiment 2 wurde geprüft, ob die Dissoziation zwischen einem impliziten und einem expliziten Gedächtnistest verschwindet, wenn die alternative Methode der Leistungswertbestimmung angewandt wird, durch die eine Reliabilitätssteigerung im impliziten Test zu erwarten war. Dazu wurde zunächst die Dissoziation zwischen dem impliziten Wortstammergänzungstest und der expliziten Form des Wortstammergänzungstests – wie sie beispielsweise von Light und Singh (1987) gefunden wurde – repliziert, indem die Leistungswerte mit Hilfe der gebräuchlichen Methode bestimmt wurden. Anschließend wurden die Leistungswerte mit Hilfe der alternativen Methode bestimmt und es wurde geprüft, ob dadurch Gruppenunterschiede zwischen jungen und älteren Probanden im impliziten Wortstammergänzungstest aufgedeckt werden können. Zur Erinnerung sei erwähnt, dass Dissoziationen zwischen impliziten und expliziten Gedächtnistests Artefakte geringer Reliabilität impliziter Tests in Verbindung mit einer hohen Reliabilität expliziter Tests sein könnten (Buchner & Wippich, 2000).

3.1 Einleitung

In dem zweiten Experiment wurde der implizite Wortstammergänzungstest der expliziten Form des Wortstammergänzungstests gegenübergestellt. Beide Tests unterscheiden sich nur im Hinblick auf die Instruktionen voneinander, was ein hohes Maß an Vergleichbarkeit mit sich bringt. Unterscheiden sich Tests in weiteren Eigenschaften als den Instruktionen, können Befunde auch auf Unterschiede in den Tests zurückzuführen sein. Während in der expliziten Form des Wortstammergänzungstests der Bezug zur Lernepisode durch Instruktionen hergestellt wird, ist dieser Bezug zur Lernepisode in dem impliziten Wortstammergänzungstest nicht vorhanden.

In einer Lernphase wurden jungen und älteren Probanden Wörter auf einem Bildschirm präsentiert. Die Wörter sollten laut ausgesprochen werden und es sollte die Häufigkeit der Wörter im deutschen Sprachgebrauch beurteilt werden. Nach einer Ablenkungsaufgabe, die eine Wahrnehmung des Zusammenhangs zwischen Lernphase und Testphase verhindern sollte, folgte die Testphase. In der Testphase wurden Wortstämme präsentiert, von denen die Hälfte zu Wörtern der Wortliste gehörte, die bereits in der Lernphase präsentiert wurde. Die andere Hälfte der Wortstämme gehörte zu Wörtern einer neuen Wortliste. Probanden des impliziten Wortstammergänzungstests wurden angewiesen, die Wortstämme mit dem ersten Wort zu ergänzen, welches Ihnen dazu einfiel. Probanden der expliziten Form des Wortstammergänzungstests wurde mitgeteilt, dass einige der Wortstämme mit Wörtern ergänzt werden können, die bereits in dem ersten Teil des Experimentes gezeigt worden sind, während andere Wortstämme nur mit neuen Wörtern ergänzt werden können. Sie sollten die Wortstämme als Hinweis für Wörter aus dem ersten Teil des Experimentes nehmen und mit diesen ergänzen.

Gruppenunterschiede zwischen jungen und älteren Probanden wurden bei der expliziten Form des Wortstammergänzungstests erwartet, jedoch nicht bei dem impliziten Wortstammergänzungstest, wenn die gebräuchliche Methode zur Bestimmung der Leistungswerte angewandt wird. Die Reliabilität des impliziten Wortstammergänzungstests sollte geringer sein als die Reliabilität der expliziten Form des Wortstammergänzungstests, wenn die gebräuchliche Methode der Bestimmung der Leistungswerte angewandt wird. Die Anwendung der alternativen Methode der Bestimmung der Leistungswerte sollte zu einer Erhöhung der Reliabilität des impliziten Wortstammergänzungstests führen und damit zu einer Aufdeckung von Gruppenunterschieden zwischen jungen und älteren Probanden. Führt die Anwendung der alternativen Methode der Bestimmung der Leistungswerte zu einer Aufdeckung von Gruppenunterschieden zwischen jungen und älteren Probanden, so spricht das dafür, dass Dissoziationen Artefakte geringer Reliabilität impliziter Gedächtnistests sind. Werden auch nach einer Reliabilitätssteigerung im impliziten Wortstammergänzungstest keine Unterschiede zwischen jungen und älteren Probanden aufgedeckt, so spricht dies entweder dafür, dass keine Gruppenunterschiede vorhanden sind und die Dissoziation zwischen beiden Gedächtnistests Bestand hat oder aber dafür, dass es sich bei der Reliabilitätssteigerung um ein Artefakt handelt.

3.2 Methode

In den kommenden Abschnitten wird der Aufbau des zweiten Experimentes näher beschrieben. Im Einzelnen wird auf die Stichprobe (3.2.1), das Material (3.2.2), die Prozedur (3.2.3) und auf das Design (3.2.4) des Versuchsaufbaus näher eingegangen. Es werden hier nur die methodischen Unterschiede zu Experiment 1 beschrieben. Die Details, die hier nicht beschrieben sind, sind somit dem Methodenteil des ersten Experimentes zu entnehmen.

3.2.1 Stichprobe

An Experiment 2 nahmen 336 Probanden teil. Die eine Hälfte der Probanden waren junge Erwachsene, die andere Hälfte der Probanden waren ältere Erwachsene. Die jungen Probanden waren zum überwiegenden Teil Studierende der Heinrich-Heine-Universität. Die Altersspanne der jungen Probanden reichte von 17 bis 52 Jahren ($M = 23.52$, $SD = 4.85$). Von den jungen Probanden waren 129 Personen weiblich und 39 Personen männlich. Die älteren Probanden, die ein Mindestalter von 60 Jahren aufweisen mussten, wurden durch Inserate in einer Düsseldorfer Tageszeitung angeworben. Die Altersspanne der älteren Probanden reichte von 60 bis 94 Jahren ($M = 71.13$, $SD = 5.92$). Von den älteren Probanden waren 105 Personen weiblich und 63 Personen männlich. Voraussetzungen für die Teilnahme der älteren Probanden waren ein altersgemäßes (korrigiertes) Seh- und Hörvermögen, sowie unbeeinträchtigte kognitive Funktionen. Die Voraussetzungen wurden in einem telefonischen Interview vor der eigentlichen Teilnahme an dem Experiment ermittelt, wobei bestimmte Vorerkrankungen und kognitive Beeinträchtigungen als Ausschlusskriterium galten. Weiterhin wurde im Anschluss an das Experiment ein Demenztest (DemTect) durchgeführt, aufgrund dessen Ergebnisse keiner der älteren Probanden ausgeschlossen werden musste. Die jungen Erwachsenen unterschieden sich – auf deskriptiver Ebene betrachtet – in ihrer Schulbildung von den älteren Erwachsenen. Von den jungen Probanden hatten 2 Personen den Realschulabschluss und 166 Personen das Abitur. Von den älteren Probanden hatten 65 Personen den Hauptschul- oder Volksschulabschluss, 71 Personen den Realschulabschluss und 32 Personen das Abitur. Gute Deutschkenntnisse wurden für die Teilnahme an dem Experiment vorausgesetzt. Personen, die an Experiment 1 teilgenommen haben, wurden nicht für die

Teilnahme an Experiment 2 zugelassen. Junge und ältere Probanden wurden einzeln getestet und randomisiert entweder dem impliziten Wortstammergänzungstest oder der expliziten Form des Wortstammergänzungstests zugewiesen. Ebenfalls randomisiert war die Zuweisung der Probanden zu Wortliste 1 oder zu Wortliste 2. Die Variable «Wortliste» diente zur Bestimmung der Basisrate und ging nicht in das eigentliche Design des Experimentes mit ein. Insgesamt gab es die folgenden acht Bedingungen mit gleicher Probandenanzahl: *Wortstammergänzungstest, jung, Wortliste 1* ($n = 42$); *Wortstammergänzungstest, jung, Wortliste 2* ($n = 42$); *Wortstammergänzungstest, alt, Wortliste 1* ($n = 42$); *Wortstammergänzungstest, alt, Wortliste 2* ($n = 42$); *explizite Form des Wortstammergänzungstests, jung, Wortliste 1* ($n = 42$); *explizite Form des Wortstammernergänzungstests, jung, Wortliste 2* ($n = 42$); *explizite Form des Wortstammernergänzungstests, alt, Wortliste 1* ($n = 42$); und *explizite Form des Wortstammernergänzungstests, alt, Wortliste 2* ($n = 42$). Die Probanden wurden am Ende des Experimentes durch Bezahlung entlohnt.

3.2.2 Material

Die verwendeten Materialien waren die gleichen, wie sie auch im ersten Experiment zum Einsatz kamen. Ein Demenzttest (DemTect) kam zur Erkennung von kognitiven Beeinträchtigungen im Alter zum Einsatz.

3.2.3 Prozedur

Die Prozedur für Probanden, die den impliziten Wortstammernergänzungstest vorgelegt bekommen haben, war die gleiche, wie bei dem im Hinblick auf Reliabilitätssteigerung bearbeiteten Wortstammernergänzungstest im ersten Experiment. Die Prozedur für Probanden, welche die explizite Form des Wortstammernergänzungstests vorgelegt bekommen haben, unterschied sich nur durch die Instruktionen, durch die ein Bezug zur Lernphase hergestellt wurde, von der Prozedur für den impliziten Wortstammernergänzungstest. In der Testphase der expliziten Form des Wortstammernergänzungstests sollten Wortstämme als Abrufhinweis für Wörter aus der Lernphase verwendet werden. Die Wortstämme sollten nur dann ergänzt werden, wenn die entsprechenden Wörter aus der Lernphase abgerufen werden konnten. Im Anschluss an das Experiment wurde an den älteren Probanden ein Demenzttest (DemTect) durchgeführt.

3.2.4 Design

Dem zweiten Experiment lag ein 2 x 2-faktorieller Versuchsplan ohne Messwiederholung zugrunde mit dem quasi-experimentellen Klassifikationsfaktor «Altersgruppe» (*jung* vs. *alt*) und dem experimentellen Treatmentfaktor «Testform» (*impliziter Wortstammergänzungstest* vs. *explizite Form des Wortstammergänzungstests*). Dies führte zu den vier Testbedingungen (*impliziter Wortstammergänzungstest, jung* vs. *impliziter Wortstammergänzungstest, alt* vs. *explizite Form des Wortstammergänzungstests, jung* vs. *explizite Form des Wortstammergänzungstests, alt*) des zugrundeliegenden 2 x 2-faktoriellen Versuchsplans. Hauptsächlicher Untersuchungsgegenstand war der Vergleich junger und älterer Probanden in beiden Testformen. Die abhängigen Variablen waren das Primingmaß des impliziten Wortstammergänzungstests sowie das entsprechende Leistungsmaß der expliziten Form des Wortstammergänzungstests.⁶ Weiterhin war der Vergleich der Reliabilität beider Gedächtnistests Gegenstand des zweiten Experimentes. Als Schätzungen der Reliabilität wurden Testhalbierungskoeffizienten verwendet.

Für die Stichprobenumfangsplanung wurde die Interaktion der beiden Faktoren zugrunde gelegt. Wenn Interaktionseffekte der Größe $f = 0.20$ in der Population aufgedeckt werden sollen («kleine» bis «mittelgroße» Effekte, nach Cohen, 1977), der Fehler 1. Art auf $\alpha = .05$ und der Fehler 2. Art auf $\beta = .05$ festgelegt wird, werden nach einer a priori Teststärkeanalyse mindestens $n = 82$ Probanden in jeder der vier Bedingungen benötigt. Tatsächlich konnten Daten von $n = 84$ Probanden pro Bedingung, bzw. von $N = 336$ Probanden insgesamt, erhoben werden. Damit ergibt sich für die Interaktion der beiden Faktoren eine Teststärke von $1 - \beta = .96$.

Wenn beim Vergleich der Reliabilität beider Gedächtnistests «mittelgroße» Unterschiede zwischen Korrelationen in der Population ($q = 0.40$, nach Cohen, 1977) aufgedeckt werden sollen, dann ergibt sich bei $\alpha = .05$, $n = 168$ Probanden für jede der beiden Testformen und bei einem einseitigen Testproblem, eine Teststärke von $1 - \beta = .98$ für den Korrelationsvergleich.

⁶ Anmerkungen zur Definition der im zweiten Experiment verwendeten Gedächtnismaße folgen bei der Darstellung der Ergebnisse.

3.3 Ergebnisse

Die gebräuchliche und die alternative Methode zur Bestimmung der Leistungswerte des impliziten Wortstammerngänzungstests wurden bereits im Ergebnisteil des ersten Experimentes näher beschrieben. Zur Bestimmung der Leistungswerte der expliziten Form des Wortstammerngänzungstests wurden die gleichen Methoden angewandt.

Die Primingwerte des impliziten Wortstammerngänzungstests und die Leistungswerte der expliziten Form des Wortstammerngänzungstests wurden zweifaktoriellen univariaten Varianzanalysen ohne Messwiederholung unterzogen. Eine Varianzanalyse wurde mit Daten durchgeführt, die mit der gebräuchlichen Methode ermittelt worden sind und eine weitere Varianzanalyse mit Daten, die mit der alternativen Methode ermittelt worden sind. Die Reliabilitätsanalysen wurden auf die gleiche Weise durchgeführt, wie dies bereits in Experiment 1 beschrieben worden ist.

Tabelle 4

Gruppenmittelwerte der Leistungsmaße, für beide Altersgruppen, beide Testformen und für beide Methoden der Bestimmung der Leistungswerte, in Experiment 2 (Die Werte in Klammern repräsentieren die Standardfehler der Mittelwerte).

	junge Probanden	ältere Probanden
impliziter Wortstammerngänzungstest Primingwert (gebräuchliche Methode)	.17 (.01)	.13 (.01)
explizite Form des Wortstammerngänzungstests Leistungswert (gebräuchliche Methode)	.46 (.02)	.35 (.02)
impliziter Wortstammerngänzungstest Primingwert (alternative Methode)	.17 (.01)	.13 (.02)
explizite Form des Wortstammerngänzungstests Leistungswert (alternative Methode)	.46 (.02)	.35 (.02)

In Tabelle 4 sind die Primingwerte des impliziten Wortstammerngänzungstests und die Leistungswerte der expliziten Form des Wortstammerngänzungstests – für beide Altersgruppen und für beide Methoden der Leistungswertbestimmung – aufgeführt.⁷

⁷ Die Variablen, die den Primingwerten und den Leistungswerten der entsprechenden Tests zugrunde liegen, sind der Tabelle B im Anhang zu entnehmen.

Ein-Stichproben-t-Tests ergaben, dass die Leistungswerte von Probanden beider Wortlisten, für beide Testformen und für beide Methoden der Leistungswertbestimmung statistisch signifikant von null verschieden waren, $ts(41) > 4.32$, $ps < .001$, $\eta_p^2s > .31$. Die über Probanden beider Wortlisten gemittelten Leistungswerte (siehe Tabelle 4) waren ebenfalls statistisch signifikant von null verschieden, $ts(83) > 8.04$, $ps < .001$, $\eta_p^2s > .44$.

Wurde die gebräuchliche Methode der Bestimmung der Leistungswerte angewandt, so ergab eine 2 x 2-faktorielle univariate Varianzanalyse ohne Messwiederholung mit den Faktoren «Altersgruppe» (*jung vs. alt*) und «Testform» (*impliziter Wortstammerngänzungstest vs. explizite Form des Wortstammerngänzungstests*) statistisch signifikante Haupteffekte der Faktoren «Altersgruppe», $F(1,332) = 25.52$, $p < .001$, $\eta_p^2 = .07$, und «Testform», $F(1,332) = 293.98$, $p < .001$, $\eta_p^2 = .47$. Die eigentlich interessante Interaktion zwischen den beiden Faktoren war ebenfalls statistisch signifikant, $F(1,332) = 6.39$, $p = .01$, $\eta_p^2 = .02$. Eine simple Effects Analyse wurde daraufhin durchgeführt, um die Interaktion genauer zu untersuchen. Post-hoc-Tests zeigten dabei, dass sich die Primingwerte des impliziten Wortstammerngänzungstests nicht statistisch signifikant zwischen jungen und älteren Probanden unterschieden ($Ms = .17$ vs. $.13$, $SEs = .01$ vs. $.01$, $p = .075$, $\eta_p^2 = .01$). Die Leistungswerte der expliziten Form des Wortstammerngänzungstests unterschieden sich jedoch statistisch signifikant zwischen jungen und älteren Probanden ($Ms = .46$ vs. $.35$, $SEs = .02$ vs. $.02$, $p < .001$, $\eta_p^2 = .08$). Wurde die alternative Methode der Bestimmung der Leistungswerte angewandt, so ergab eine 2 x 2-faktorielle univariate Varianzanalyse ohne Messwiederholung mit den Faktoren «Altersgruppe» (*jung vs. alt*) und «Testform» (*impliziter Wortstammerngänzungstest vs. explizite Form des Wortstammerngänzungstests*) statistisch signifikante Haupteffekte der Faktoren «Altersgruppe», $F(1,332) = 20.11$, $p < .001$, $\eta_p^2 = .06$, und «Testform», $F(1,332) = 231.64$, $p < .001$, $\eta_p^2 = .41$. Die eigentlich interessante Interaktion zwischen den beiden Faktoren war ebenfalls statistisch signifikant, $F(1,332) = 5.04$, $p = .025$, $\eta_p^2 = .02$. Eine simple Effects Analyse wurde daraufhin durchgeführt, um die Interaktion genauer zu untersuchen. Post-hoc-Tests zeigten dabei, dass sich die Primingwerte des impliziten Wortstammerngänzungstests nicht statistisch signifikant zwischen jungen und älteren Probanden unterschieden ($Ms = .17$ vs. $.13$, $SEs = .01$ vs. $.02$, $p = .114$, $\eta_p^2 = .01$). Die Leistungswerte der expliziten Form des Wortstammerngänzungstests unterschieden sich jedoch statistisch signifikant zwischen jungen und älteren Pro-

banden ($M_s = .46$ vs. $.35$, $SE_s = .02$ vs. $.02$, $p < .001$, $\eta_p^2 = .06$).

In Tabelle 5 werden die Testhalbierungskoeffizienten der beiden Testformen präsentiert.⁸ Es sind sowohl Testhalbierungskoeffizienten aufgeführt, die aus Leistungswerten geschätzt wurden, welche mit der gebräuchlichen Methode ermittelt wurden, als auch Testhalbierungskoeffizienten, die aus Leistungswerten geschätzt wurden, welche mit der alternativen Methode ermittelt wurden.

Tabelle 5

Testhalbierungskoeffizienten der Leistungsmaße, für beide Testformen und für beide Methoden der Bestimmung der Leistungswerte, in Experiment 2 (Die Werte in Klammern repräsentieren die Spearman-Brown korrigierten Koeffizienten für die volle Testlänge).

	impliziter Wortstammergänzungstest	explizite Form des Wortstammergänzungstests
gebräuchliche Methode der Leistungswertbestimmung (r_{ii})	.22 (.36)	.81 (.90)
alternative Methode der Leistungswertbestimmung (r_{ii})	.71 (.83)	.86 (.93)

Die Testhalbierungskoeffizienten von Probanden beider Wortlisten und Probanden beider Altersgruppen wurden gemittelt, um die Teststärke für den entscheidenden Vergleich zwischen den Testformen zu erhöhen. Wenn die gebräuchliche Methode der Bestimmung der Leistungswerte angewandt wurde, war der Testhalbierungskoeffizient der expliziten Form des Wortstammergänzungstests (.81) statistisch signifikant größer, als der Testhalbierungskoeffizient des impliziten Wortstammergänzungstests (.22), $z = 8.21$, $p < .001$, $q = 0.90$. Wenn die alternative Methode der Bestimmung der Leistungswerte angewandt wurde, war der Testhalbierungskoeffizient der expliziten Form des Wortstammergänzungstests (.86) statistisch signifikant größer, als der Testhalbierungskoeffizient des impliziten Wortstammergänzungstests (.71), $z = 3.69$, $p < .001$, $q = 0.41$. Wenn die Leistungswerte mit der alternativen Methode bestimmt wurden, war der Testhalbierungskoeffizient des impliziten Wortstammergänzungstests (.71) statistisch signifikant größer, als wenn die Leistungswerte mit der gebräuchlichen Methode bestimmt wurden (.22), $z = 6.03$, $p < .001$, $q = 0.66$.

⁸ Die statistischen Kennwerte, die den Testhalbierungskoeffizienten von Experiment 2 zugrunde liegen, sind Tabelle 6 zu entnehmen.

In Tabelle 6 sind die Kovarianzen zwischen den Testhälften und die Standardabweichungen der Testhälften aufgeführt, die den Testhalbierungskoeffizienten des zweiten Experimentes zugrunde liegen. Wurde die alternative Methode der Leistungswertbestimmung angewandt, so waren die Kovarianzen für Probanden beider Wortlisten im impliziten Wortstammerngänzungstest höher, als bei der gebräuchlichen Methode. Zu sehen ist ebenfalls ein leichter Anstieg der Standardabweichungen der Testhälften für Probanden beider Wortlisten im impliziten Wortstammerngänzungstest, von der gebräuchlichen Methode zur alternativen Methode. Für die explizite Form des Wortstammerngänzungstests ist der Anstieg der Kovarianzen von der gebräuchlichen zur alternativen Methode weniger ausgeprägt, was sich in dem mäßigen Anstieg des entsprechenden Testhalbierungskoeffizienten zeigt (siehe Tabelle 5).

Tabelle 6

Kovarianzen zwischen den Testhälften und die Standardabweichungen der Testhälften, wie sie für die Schätzung der Reliabilität in Experiment 2 verwendet wurden. Aufgeführt sind die Werte für beide Testformen, beide Altersgruppen und für beide Methoden der Leistungswertbestimmung.

	junge Probanden		ältere Probanden	
	Wortliste 1	Wortliste 2	Wortliste 1	Wortliste 2
impliziter Wortstammerngänzungstest (gebräuchliche Methode)				
$COV_{Th1,Th2}$	0.001	0.002	0.005	< 0.001
SD_{Th1}	0.09	0.10	0.12	0.09
SD_{Th2}	0.08	0.09	0.11	0.08
explizite Form des Wortstammerngänzungstests (gebräuchliche Methode)				
$COV_{Th1,Th2}$	0.011	0.022	0.021	0.029
SD_{Th1}	0.14	0.18	0.17	0.19
SD_{Th2}	0.11	0.15	0.15	0.17
impliziter Wortstammerngänzungstest (alternative Methode)				
$COV_{Th1,Th2}$	0.015	0.010	0.022	0.015
SD_{Th1}	0.16	0.12	0.17	0.14
SD_{Th2}	0.13	0.12	0.17	0.14
explizite Form des Wortstammerngänzungstests (alternative Methode)				
$COV_{Th1,Th2}$	0.033	0.027	0.032	0.016
SD_{Th1}	0.21	0.19	0.20	0.15
SD_{Th2}	0.18	0.16	0.19	0.13

3.4 Diskussion

Die Reliabilität des impliziten Wortstammergänzungstests war statistisch signifikant geringer als die Reliabilität der expliziten Form des Wortstammergänzungstests, wenn die gebräuchliche Methode zur Bestimmung der Leistungswerte angewandt wurde. Junge Probanden schnitten statistisch signifikant besser in den Leistungswerten der expliziten Form des Wortstammergänzungstests ab, während im Priming des impliziten Wortstammergänzungstests keine statistisch signifikanten Unterschiede zwischen jungen und älteren Probanden gefunden werden konnten, wenn die gebräuchliche Methode der Bestimmung der Leistungswerte angewandt wurde. Die Dissoziation zwischen dem impliziten Wortstammergänzungstest und der expliziten Form des Wortstammergänzungstests – wie sie z.B. von Light und Singh (1987) gefunden wurde – konnte damit repliziert werden. Beide Befunde decken sich mit den Voraussagen und sind konsistent mit der Vorstellung, dass Dissoziationen zwischen impliziten und expliziten Gedächtnistests Artefakte geringer Reliabilität impliziter Gedächtnistests sein könnten.

Übereinstimmend mit den Voraussagen kam es bei der Bestimmung der Leistungswerte mit der alternativen Methode zu einer statistisch signifikanten Steigerung der Reliabilitätsschätzung des impliziten Wortstammergänzungstests, gegenüber der Bestimmung der Leistungswerte mit der gebräuchlichen Methode. Die Reliabilitätsschätzung des impliziten Wortstammergänzungstests war jedoch immer noch statistisch signifikant geringer als die Reliabilitätsschätzung der expliziten Form des Wortstammergänzungstests. Die Unterschiede zwischen den Reliabilitätsschätzungen beider Tests sind jedoch von der gebräuchlichen zur alternativen Methode der Leistungswertbestimmung deutlich geringer geworden ($q = 0.90$ vs. $q = 0.41$). Die Reliabilitätsschätzung der expliziten Form des Wortstammergänzungstests erfuhr ebenfalls eine leichte Steigerung von der gebräuchlichen zur alternativen Methode der Leistungswertbestimmung, die jedoch rein deskriptiv war. Da es sich dabei um eine Steigerung im oberen Korrelationsbereich handelt, wäre an dieser Stelle noch anzumerken, dass Korrelationen nicht intervallskaliert sind. Damit sind Korrelationskoeffizienten nicht äquidistant und Vergleiche von Korrelationsdifferenzen auch nicht ohne weiteres zulässig. Gleiche Differenzen zwischen Korrelationskoeffizienten entsprechen im oberen Korrelationsbereich deutlich größeren Unterschieden, als im unteren Korrelationsbereich. Cohen's q stellt eine Effektgröße für Korrelationsunter-

schiede dar, die über die Differenz zweier Fisher Z-transformierter Korrelationen bestimmt wird. Sie ist äquidistant und erlaubt den direkten Vergleich von Korrelationsunterschieden. Korrelationen sind in ihrer Höhe auf den Wert eins begrenzt, da es sich um ein genormtes Maß für den Vergleich linearer Zusammenhänge handelt. Zur Bestimmung von Korrelationen werden Kovarianzen – als unstandardisierte lineare Zusammenhangsmaße – an dem Produkt der Standardabweichungen beider Merkmale normiert. Die Korrelation zweier Variablen lässt sich auch als Kovarianz beider z-standardisierter Variablen beschreiben und kann Werte zwischen null und eins annehmen.

Durch die alternative Methode der Leistungswertbestimmung und der damit verbundenen Reliabilitätserhöhung wurden keine statistisch signifikanten Gruppenunterschiede im Priming des impliziten Wortstammergänzungstests zwischen jungen und älteren Probanden aufgedeckt. Dies deckt sich mit einer der beiden möglichen Voraussagen und spricht dafür, dass es im impliziten Wortstammergänzungstest entweder keine Gruppenunterschiede zwischen jungen und älteren Probanden gibt und tatsächlich eine Dissoziation zwischen beiden Gedächtnistests besteht, oder dass es sich bei der Reliabilitätssteigerung durch die alternative Methode der Leistungswertbestimmung um ein Artefakt handelt. Bei den niedrigen Primingwerten des impliziten Wortstammergänzungstests könnte jedoch auch ein Bodeneffekt vorgelegen haben, weil die Aufgabe insgesamt zu «schwierig» war. Obwohl alle Primingwerte statistisch signifikant von null verschieden waren, könnten ältere Probanden sich in einem solchen Fall von jüngeren Probanden – aufgrund der niedrigen Primingwerte – kaum noch unterscheiden.

4 Experiment 3

In Experiment 2 konnte die Dissoziation zwischen dem impliziten Wortstammergänzungstest und der expliziten Form des Wortstammergänzungstests – wie sie z.B. von Light und Singh (1987) berichtet wurde – repliziert werden. Die Schätzung der Reliabilität bestätigte die Annahme, dass der implizite Wortstammergänzungstest deutlich unreliabler ist, als die explizite Form des Wortstammergänzungstests. Wurde jedoch statt der gebräuchlichen Methode, Leistungswerte zu bestimmen, die alternative Methode der Leistungswertbestimmung angewandt, so kam es zu einer statistisch signifikanten Steigerung der Reliabilität des impliziten Wortstammergänzungstests, die nun vergleichbar mit der Reliabilität expliziter Gedächtnistests war. Trotz der gesteigerten Messpräzision blieb die Dissoziation zwischen den beiden Gedächtnistests bestehen. Ziel von Experiment 3 war es – analog zu der Vorgehensweise in Experiment 2 – junge und ältere Probanden in einem weiteren impliziten und expliziten Gedächtnistest miteinander zu vergleichen und die Befunde aus dem zweiten Experiment zu überprüfen. Es sollte diesmal ein impliziter Gedächtnistest verwendet werden, der bedingt durch seine restriktiveren Testbedingungen eine höhere Reliabilität erwarten ließ.

4.1 Einleitung

Als impliziter Gedächtnistest wurde der Bildidentifikationstest in einer Klärifikationsprozedur verwendet. Als expliziter Gedächtnistest kam der Bilderkennungstest zur Anwendung. Eine Eigenschaft von impliziten Identifikationstests ist es – ähnlich wie auch bei expliziten Gedächtnistests –, dass die Testbedingungen ziemlich restriktiv sind (Buchner & Wippich, 2000). In einer typischen Klärifikationsprozedur sollen Probanden – so schnell und so genau wie möglich – fragmentierte visuelle Stimuli erkennen. Die Vielfalt möglicher kognitiver Prozesse, die zur Bewältigung der Aufgabe beitragen könnten, wird bei diesem Aufgabentyp stark eingeschränkt, was sich auch in einer höheren Reliabilität äußern sollte (Buchner & Wippich, 2000). Buchner und Wippich (2000) verglichen diese beiden Gedächtnistests in Bezug auf ihre Reliabilität miteinander. Die Autoren fanden in ihrem dritten Experiment, dass die Reliabilitätsschätzungen sich nicht zwischen den beiden Gedächtnistests unter-

scheiden. Sie schlussfolgerten, dass die Reliabilität impliziter Gedächtnistests sich nicht notwendigerweise von der Reliabilität expliziter Gedächtnistests unterscheiden muss, solange die Testbedingungen restriktiv genug sind. Eine Einschränkung in dem Experiment 3 von Buchner und Wippich (2000) jedoch war die geringe Reliabilität beider Gedächtnistests, was mit der geringen Anzahl der Items erklärt wurde.

In Experiment 3 der vorliegenden Arbeit wurde die Anzahl der Items, wie sie in der Studie von Buchner und Wippich (2000) verwendet wurden, zur Reliabilitätssteigerung verdoppelt. Junge und ältere Probanden bekamen in einer Lernphase Bilder von Objekten präsentiert, die benannt werden sollten. Nach einer Ablenkungsaufgabe folgte die Testphase. In dem Bilderkennungstest wurden Probanden die Bilder aus der Lernphase sowie die gleiche Anzahl neuer Bilder präsentiert. Die Aufgabe von Probanden war es, zu beurteilen, ob Bilder bereits in dem ersten Teil des Experimentes präsentiert worden sind oder ob sie neu waren. In dem Bildidentifikationstest wurden Probanden alte Bilder sowie die gleiche Anzahl neuer Bilder in einer fragmentierten Form präsentiert. Die Bilder wurden schrittweise immer deutlicher und die Aufgabe der Probanden war es, Objekte so schnell wie möglich zu benennen.

Für den impliziten Bildidentifikationstest wurde, aufgrund der restriktiven Testbedingungen, eine ähnlich hohe Reliabilität erwartet wie für den expliziten Bilderkennungstest, wenn die gebräuchliche Methode der Bestimmung der Leistungswerte angewandt wird. Gruppenunterschiede zwischen jungen und älteren Probanden sollten sowohl im expliziten Bilderkennungstest, als auch im impliziten Bildidentifikationstest aufgedeckt werden können. Wenn die alternative Methode der Leistungswertbestimmung angewandt wird, sollte dies zu einer weiteren Steigerung der Reliabilität des impliziten Bildidentifikationstests führen. Gruppenunterschiede zwischen jungen und älteren Probanden wurden in diesem Fall sowohl im expliziten Bilderkennungstest, als auch im impliziten Bildidentifikationstest erwartet. Wenn Gruppenunterschiede zwischen jungen und älteren Probanden durch einen reliablen impliziten Bildidentifikationstest aufgedeckt werden können, so spricht das dafür, dass Dissoziationen zwischen impliziten und expliziten Gedächtnistests Artefakte geringer Reliabilität impliziter Gedächtnistests sind. Sollten jedoch keine Gruppenunterschiede aufgedeckt werden können, so spricht das dafür, dass Dissoziationen zwischen impliziten und expliziten Gedächtnistests tatsächlich vorhanden sind.

4.2 Methode

In den kommenden Abschnitten wird der Aufbau des dritten Experimentes näher beschrieben. Im Einzelnen wird dabei auf die Stichprobe (4.2.1), das Material (4.2.2), die Prozedur (4.2.3) und auf das Design (4.2.4) des Versuchsaufbaus näher eingegangen.

4.2.1 Stichprobe

An Experiment 3 nahmen 336 Probanden teil. Die Hälfte der Probanden waren junge Erwachsene, die andere Hälfte der Probanden waren ältere Erwachsene. Die jungen Probanden waren zum überwiegenden Teil Studierende der Heinrich-Heine-Universität. Die Altersspanne der jungen Probanden reichte von 18 bis 48 Jahren ($M = 24.76$, $SD = 5.29$). Von den jungen Probanden waren 113 Personen weiblich und 55 Personen männlich. Die älteren Probanden, die ein Mindestalter von 60 Jahren aufweisen mussten, wurden durch Inserate in einer Düsseldorfer Tageszeitung angeworben. Die Altersspanne der älteren Probanden reichte von 60 bis 94 Jahren ($M = 71.67$, $SD = 6.20$). Von den älteren Probanden waren 106 Personen weiblich und 62 Personen männlich. Voraussetzungen für die Teilnahme der älteren Probanden waren ein altersgemäßes (korrigiertes) Seh- und Hörvermögen, sowie unbeeinträchtigte kognitive Funktionen. Die Voraussetzungen wurden in einem telefonischen Interview vor der eigentlichen Teilnahme an dem Experiment ermittelt, wobei bestimmte Vorerkrankungen und kognitive Beeinträchtigungen als Ausschlusskriterium galten. Zusätzlich dazu wurde im Anschluss an das Experiment ein Demenztest (DemTect) durchgeführt, aufgrund dessen Ergebnisse keiner der älteren Probanden ausgeschlossen werden musste. Die jungen Erwachsenen unterschieden sich – auf deskriptiver Ebene betrachtet – in ihrer Schulbildung von den älteren Erwachsenen. Von den jungen Probanden hatten 3 Personen den Realschulabschluss und 165 Personen das Abitur. Von den älteren Probanden hatten 61 Personen den Hauptschul- oder Volksschulabschluss, 63 Personen den Realschulabschluss und 44 Personen das Abitur. Gute Deutschkenntnisse wurden für die Teilnahme an dem Experiment vorausgesetzt. Junge und ältere Probanden wurden einzeln getestet und randomisiert entweder dem impliziten Bildidentifikationstest oder dem expliziten Bilderkennungstest zugewiesen. Ebenfalls randomisiert war die Zuweisung der Probanden entweder zu

Bilderset 1 und damit der Präsentation von Set 1 Bildern in der Lernphase, oder zu Bilderset 2 und der Präsentation von Set 2 Bildern. Die Variable «Bilderset» diente zur Bestimmung der Basisrate und ging nicht in das eigentliche Design des Experimentes mit ein. Insgesamt gab es somit die folgenden acht Bedingungen mit gleicher Probandenanzahl: *Bildidentifikation, jung, Bilderset 1 (n = 42)*; *Bildidentifikation, jung, Bilderset 2 (n = 42)*; *Bildidentifikation, alt, Bilderset 1 (n = 42)*; *Bildidentifikation, alt, Bilderset 2 (n = 42)*; *Bilderkennung, jung, Bilderset 1 (n = 42)*; *Bilderkennung, jung, Bilderset 2 (n = 42)*; *Bilderkennung, alt, Bilderset 1 (n = 42)*; und *Bilderkennung, alt, Bilderset 2 (n = 42)*. Die Versuchspersonen wurden am Ende des Experimentes durch Bezahlung entlohnt.

4.2.2 Material

Als Stimulusmaterial wurden 200 Bilder aus einem Bilderset ausgewählt, das von Snodgrass und Vanderwart (1980) veröffentlicht wurde. Die 200 Bilder wurden auf zwei Bildersets mit jeweils 100 Bildern aufgeteilt. Die Zuweisung der Bilder zu den beiden Bildersets erfolgte randomisiert. Das Experiment lief auf einem Windows PC, an dem ein zusätzlicher Bildschirm angeschlossen war. Der Bildschirm des Windows PC diente zur Steuerung des Experimentes. Er konnte nur von dem Versuchsleiter eingesehen werden, da zwischen dem Versuchsleiter und dem Probanden ein Sichtschutz aufgestellt war. Auf dem zweiten Bildschirm, der eine Diagonale von 14 Zoll hatte, bekamen Probanden sowohl Instruktionen als auch das Stimulusmaterial präsentiert. Der Abstand der Probanden zum Bildschirm betrug etwa 80 cm. Die Bilder wurden so skaliert, dass sie in einem Quadrat von etwa 11.8 x 11.8 cm Kantenlänge auf dem Bildschirm Platz fanden. In der Lernphase wurden die Objekte eines der beiden Bildersets präsentiert, je nachdem welchem Bilderset Probanden zugewiesen worden sind. Die Objekte wurden als schwarze Linienzeichnungen präsentiert. Zwischen Lernphase und Testphase kam eine Ablenkungsaufgabe zum memorieren von Zahlen unter Interferenzbedingungen zum Einsatz, die auf einem Apple iMac Computer mit einer Bildschirmdiagonalen von 20 Zoll präsentiert wurde. In der Testphase des Experimentes wurden die Objekte beider Bildersets in einer fragmentierten Form dargeboten, wobei die Fragmentierung dadurch erreicht wurde, dass jedes Bild von 15 verschiedenen Masken überlagert werden konnte. Jede der 15 Masken gab sukzessive einen größeren Anteil des darunter liegenden Objektes frei. Die Masken

hatten alle die gleiche Kantenlänge wie das eigentliche Bild. Am Anfang des Defragmentationsprozesses war jedes Objekt vollständig von der ersten Maske bedeckt. Am Ende jedoch, nach 15 Defragmentationsstufen, war jedes Objekt vollständig sichtbar. Die 15 Masken wurden so programmiert, dass der Bildausschnitt in 225 kleine Quadrate von knapp 0.79×0.79 cm Kantenlänge aufgeteilt wurde. Mit jeder Defragmentationsstufe wurden 15 weitere dieser kleinen Quadrate aus der Maske entfernt und damit immer mehr Sicht auf die darunter liegenden Objekte freigegeben. Die Position der Quadrate wurde für jede der Defragmentationsstufen mit einer Zufallsfunktion neu berechnet. Das Experiment fand unter lärmberuhigten Bedingungen statt.

4.2.3 Prozedur

Nach Eingabe der demographischen Daten wurde die Studie den Probanden als eine Bilderstudie zur Optimierung von Computer Monitoren vorgestellt. Die jungen und die älteren Probanden wurden einzeln getestet. Die Lernphase des Experimentes, die für alle Probanden gleich war, begann mit der Präsentation von Bildern aus einem der beiden Bildersets, je nachdem, zu welchem Bilderset die Probanden zugeordnet worden sind. Die Abfolge der Bilder in der Lernphase war für jeden Probanden randomisiert. Auch in der Testphase war die Abfolge der Bilder – bei der Aufgabe zur Bilderkennung – sowie die Abfolge der fragmentierten Bilder – bei der Aufgabe zur Bildidentifikation in einer Klarifikationsprozedur – für jeden Probanden randomisiert. In der Lernphase erschienen Bilder von Objekten, die benannt werden sollten. Unter jedem Bild war das Symbol eines Mikrofons eingeblendet, um die Probanden auf ihre Rückmeldung aufmerksam zu machen. Den Probanden wurde mitgeteilt, dass ihre Antwortzeiten erfasst werden. Sie erhielten die Anweisung, die Objekte möglichst zügig und spontan zu benennen. In diesem Experiment gab es jedoch keine Sprachaufzeichnung. Diese Anweisung diente lediglich dazu, eine Leistungsatmosphäre zu erzeugen. Der Versuchsleiter konnte auf seinem Monitor erkennen, welches Objekt Probanden gerade präsentiert wurde. Der nächste Durchgang wurde dann eingeleitet, wenn das jeweils aktuelle Objekt richtig benannt worden ist, wobei Synonyme erlaubt waren. Die Lernphase erstreckte sich über eine Dauer von etwa sieben Minuten.

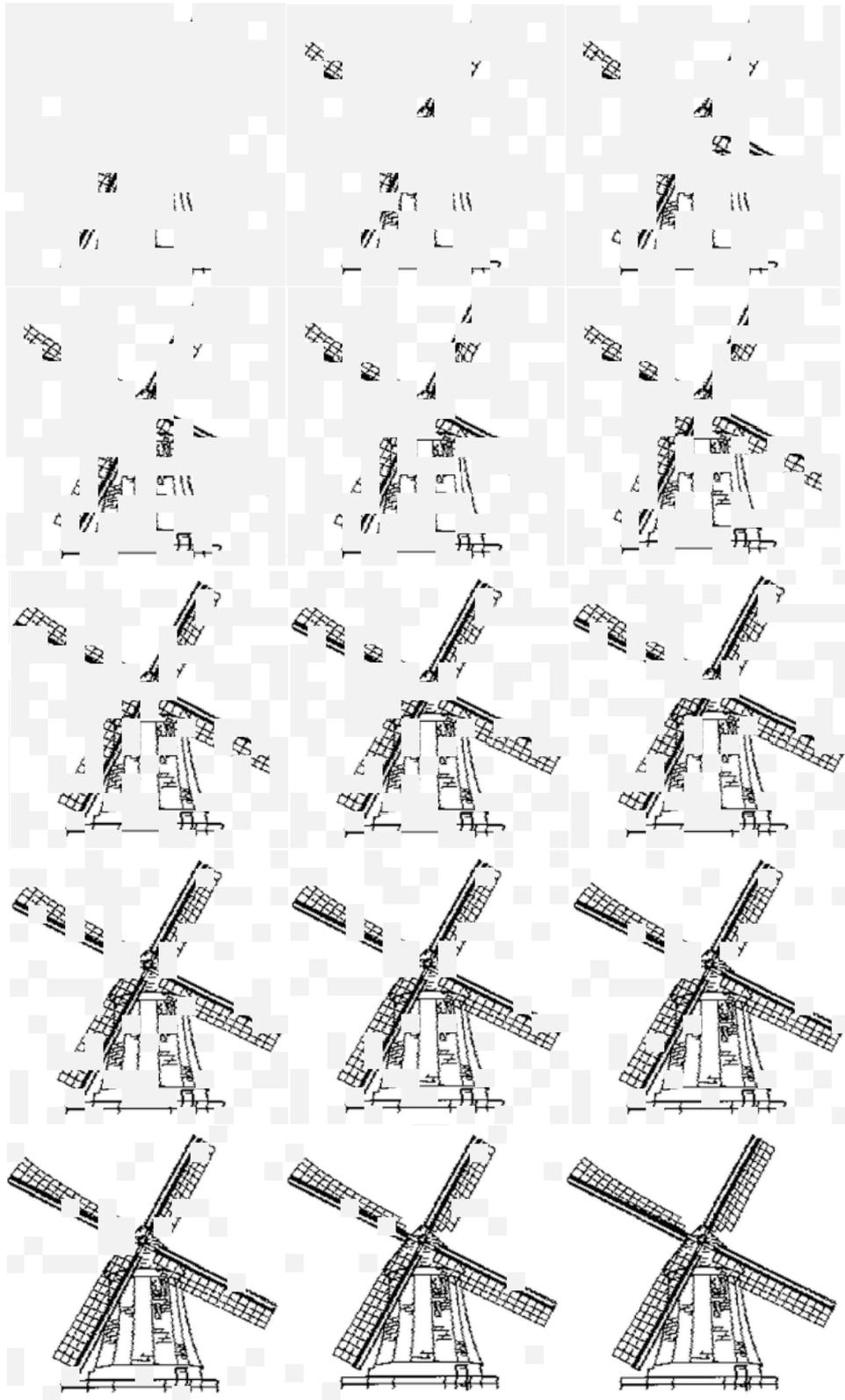


Abbildung 4: Beispiel einer Abfolge von Defragmentationsschritten bei der perzeptuellen Identifikation von Objekten in einer Klarifikationsprozedur.

Unmittelbar nach der Lernphase folgte eine Ablenkungsaufgabe zum memorieren von Zahlen. In der Ablenkungsaufgabe, die auf einem anderen Bildschirm präsentiert wurde, bekamen die Probanden eine Abfolge von Zahlen unter Interferenzbedingungen dargeboten. Über einen Kopfhörer wurde dazu eine Abfolge von Silben akustisch dargeboten, während Probanden Zahlen memorieren sollten, die nacheinander auf dem Bildschirm erschienen. Diese Zahlen sollten anschließend über den Zahlenblock der Tastatur wiedergegeben werden. Es wurde darauf geachtet, dass die akustisch dargebotenen Silben nicht mit dem Stimulusmaterial aus der Lernphase interferierten. Nach Beendigung der Ablenkungsaufgabe, die eine Dauer von etwa 20 Minuten hatte, folgte die Testphase.

Die Testphase unterschied sich in ihrem Ablauf für Probanden, die der Bildrekognitionsaufgabe zugeteilt worden sind, von dem Ablauf für Probanden, die der Bildidentifikationsaufgabe zugeteilt worden sind. In der Bildrekognitionsaufgabe wurden die Bilder des Bildersets aus der Lernphase präsentiert, sowie die gleiche Anzahl neuer Bilder. Je nachdem, zu welchem Bilderset Probanden in der Lernphase zugeteilt worden sind, waren die neuen Bilder in der Testphase aus dem jeweils anderen Bilderset. Den Probanden wurde mitgeteilt, dass eine Folge von Bildern von Objekten aus dem ersten Teil des Experimentes und Bilder von neuen Objekten erscheinen würden. Diese Objekte sollten möglichst schnell als «alt» oder als «neu» bewertet werden und es wurde gesagt, dass die Antwortzeiten erfasst werden. Auch in der Testphase wurde das Symbol eines Mikrofons unter jedem Objekt eingeblendet. Der Versuchsleiter kodierte die Antworten als «alt» oder «neu», indem er auf die entsprechenden Buttons auf seinem Kontrollmonitor klickte. Damit wurde der nächste Durchgang eingeleitet und das nächste Bild erschien auf dem Bildschirm des Probanden. In der Bildidentifikationsaufgabe mit Klarifikationsprozedur wurden sowohl Bilder aus dem Bilderset der Lernphase, als auch die neuen Bilder, in einer fragmentierten Form präsentiert (siehe Abbildung 4). Je nachdem, zu welchem Bilderset die Probanden in der Lernphase zugeteilt worden sind, waren die fragmentierten Bilder von neuen Objekten in der Testphase aus dem jeweils anderen Bilderset. Den Probanden wurde mitgeteilt, dass nun eine Folge von Bildern von Objekten erscheinen würde. Jedes Objekt würde zunächst in seiner höchsten Fragmentierungsstufe erscheinen und es würde dann mit jedem Schritt immer deutlicher zu sehen sein. Aufgabe der Probanden war es, die Objekte so schnell wie möglich und mit so wenig Defragmentationsschritten wie möglich, zu benennen. Das Symbol des Mi-

krofons erschien unter jedem Objekt. Wenn die Benennung von Objekten bei einer bestimmten Stufe der Defragmentation nicht möglich war, so gaben die Probanden ein Zeichen an den Versuchsleiter. Der Versuchsleiter, der auf seinem Kontrollbildschirm den Namen und die Defragmentationsstufe des jeweiligen Objektes erkennen konnte, rief dann die nächst deutlichere Defragmentationsstufe auf. Dies erfolgte so lange, bis die Probanden die Objekte richtig benennen konnten oder bis das Objekt die letzte Defragmentationsstufe erreicht hatte und der Versuchsleiter den nächsten Durchgang einleitete. Die Benennung der Objekte wurde auch dann als korrekt gewertet, wenn die gleichen Synonyme wie in der Lernphase verwendet wurden. Die Testphase nahm etwa 30 Minuten Zeit in Anspruch. Im Anschluss an die Testphase wurde an den älteren Probanden ein Demenztest (DemTect) durchgeführt. Insgesamt dauerte das Experiment etwa eine Stunde.

4.2.4 Design

Dem dritten Experiment lag ein 2×2 -faktorieller Versuchsplan ohne Messwiederholung zugrunde mit dem quasi-experimentellen Klassifikationsfaktor «Altersgruppe» (*jung* vs. *alt*) und dem experimentellen Treatmentfaktor «Testform» (*Bildidentifikationstest* vs. *Bilderkennungstest*). Dies führte zu den vier Testbedingungen (*Bildidentifikationstest, jung* vs. *Bildidentifikationstest, alt* vs. *Bilderkennungstest, jung* vs. *Bilderkennungstest, alt*) des zugrundeliegenden 2×2 -faktoriellen Versuchsplans. Hauptsächlichere Untersuchungsgegenstand war der Vergleich junger und älterer Probanden in beiden Testformen. Die abhängigen Variablen waren das Primingmaß des impliziten Bildidentifikationstests, sowie das Sensitivitätsmaß P_r des expliziten Bilderkennungstests.⁹ Weiterhin war der Vergleich der Reliabilität beider Gedächtnistests Gegenstand des zweiten Experimentes. Als Schätzungen der Reliabilität wurden Testhalbierungskoeffizienten verwendet.

Für die Stichprobenumfangsplanung wurde die Interaktion der beiden Faktoren zugrunde gelegt. Wenn Interaktionseffekte der Größe $f = 0.20$ in der Population aufgedeckt werden sollen («kleine» bis «mittelgroße» Effekte, nach Cohen, 1977), der

⁹ Die Definition des Primingmaßes, welches zur Erfassung der Gedächtnisleistungen im impliziten Bildidentifikationstest Verwendung fand und die Definition des Sensitivitätsmaßes P_r , welches zur Erfassung der Gedächtnisleistungen im expliziten Bilderkennungstest Verwendung fand, erfolgt bei der Darstellung der Ergebnisse.

Fehler 1. Art auf $\alpha = .05$ und der Fehler 2. Art auf $\beta = .05$ festgelegt wird, werden nach einer a priori Teststärkeanalyse mindestens $n = 82$ Probanden in jeder der vier Bedingungen benötigt. Tatsächlich konnten Daten von $n = 84$ Probanden pro Bedingung, bzw. von $N = 336$ Probanden insgesamt, erhoben werden. Damit ergibt sich für die Interaktion der beiden Faktoren eine Teststärke von $1 - \beta = .96$.

Wenn beim Vergleich der Reliabilität beider Gedächtnistests «mittelgroße» Unterschiede zwischen Korrelationen in der Population ($q = 0.40$, nach Cohen, 1977) aufgedeckt werden sollen, dann ergibt sich bei $\alpha = .05$, $n = 168$ Probanden für jede der beiden Testformen und bei einem einseitigen Testproblem eine Teststärke von $1 - \beta = .98$ für den Korrelationsvergleich.

4.3 Ergebnisse

Das Sensitivitätsmaß P_r wurde als Leistungsmaß für den expliziten Bildrekognitionstest verwendet. Probanden wurde in der Lernphase, je nach Setzuweisung, eines der beiden Bildersets präsentiert. In der Testphase wurden sowohl die Bilder des Bildersets aus der Lernphase, als auch die Bilder des jeweils anderen, «neuen», Bildersets präsentiert. Die Bearbeitung der Bilder des «alten» Bildersets diente zur Erfassung der Trefferrate. Die Bearbeitung der Bilder des «neuen» Bildersets diente zur Ermittlung der Rate falscher Alarme, die zur Bestimmung des Sensitivitätswertes P_r von der Trefferrate abgezogen wurde. Eine Korrektur der falschen Alarme ist notwendig, weil Bilder des «alten» Bildersets mit einer bestimmten Wahrscheinlichkeit auf der Basis von Rateprozessen als «alt» eingestuft werden, selbst wenn sie nicht erinnert werden. Zur Schätzung der Höhe dieser Wahrscheinlichkeit bedient man sich der Rate spontaner «alt» Einstufungen von Bildern des «neuen» Bildersets. Zur Ermittlung der Trefferrate wurde der Anteil derjenigen Bilder aus der Lernphase bestimmt, die durch Probanden in der Testphase als «alt» eingestuft worden sind, an allen Bildern aus der Lernphase. Davon abgezogen wurde der Anteil der als «alt» eingestuften Bilder des «neuen» Bildersets, an allen Bildern des «neuen» Bildersets. Auf diese Weise wurde für jeden Probanden ein Sensitivitätswert P_r ermittelt. Der Sensitivitätswert P_r der Skala wurde als Mittelwert der Sensitivitätswerte P_r aller Probanden bestimmt.

Das Priming wurde als Leistungsmaß für den impliziten Bildidentifikationstest herangezogen. Probanden wurden in der Lernphase, je nach Setzuweisung, eines der beiden Bildersets präsentiert. In der Testphase wurden sowohl die fragmentierten Bilder des Bildersets aus der Lernphase, als auch fragmentierte Bilder des jeweils anderen, «neuen», Bildersets präsentiert. Die Bearbeitung der fragmentierten Bilder des «alten» Bildersets diente zur Erfassung der Rohwerte des impliziten Gedächtnisses. Die Bearbeitung der fragmentierten Bilder des «neuen» Bildersets diente zur Erfassung der Basisrate, die zur Bestimmung des Primingwertes von dem Rohwert abgezogen wurde. Eine Korrektur der Basisrate ist notwendig, weil fragmentierte Bilder des «alten» Bildersets mit einer bestimmten Wahrscheinlichkeit auch spontan «früher» identifiziert werden, selbst wenn keine Erfahrungsnachwirkungen aus der Lernphase bestehen. Zur Ermittlung des Rohwertes wurde der Anteil der Defragmentationsstufen ermittelt, die zur Identifikation der Bilder des «alten» Bildersets aus der Lernphase benötigt wurden, an der Gesamtzahl der maximal möglichen Defragmentationsstufen aller Bilder des «alten» Bildersets. Davon abgezogen wurde der Anteil der Defragmentationsstufen, die zur Identifikation von Bildern des «neuen» Bildersets benötigt wurden, an der Gesamtzahl der maximal möglichen Defragmentationsstufen aller Bilder des «neuen» Bildersets. Auf diese Weise wurde für jeden Probanden ein Primingwert ermittelt. Der Primingwert der Skala wurde als Mittelwert der Primingwerte aller Probanden bestimmt.

Neben der bereits beschriebenen Methode der Bestimmung der Sensitivitätswerte P_r und der Primingwerte wurde auch eine alternative Methode der Bestimmung der Leistungswerte angewandt. Die beiden Methoden unterscheiden sich in der Ermittlung der Rate falscher Alarme bzw. der Basisrate voneinander. Die Ermittlung der Rate falscher Alarme bzw. der Basisrate bei der bereits beschriebenen Methode der Leistungswertbestimmung erfolgte an denselben Probanden, an denen auch die Trefferrate bzw. die Rohwerte ermittelt worden sind. Somit erfolgte die Bestimmung der Trefferrate bzw. der Rohwerte und der Rate falscher Alarme bzw. der Basisrate an Bildern *unterschiedlicher* Bildersets. Bei der alternativen Methode jedoch wurde die Rate falscher Alarme bzw. die Basisrate an Probanden ermittelt, die dem jeweils anderen Set zugewiesen worden sind. Zur Bestimmung der Sensitivitätswerte P_r oder der Primingwerte wurden Probanden beider Sets so gepaart, dass beispielsweise die Rate falscher Alarme bzw. die Basisrate des ersten Probanden eines Sets von der Trefferrate bzw. dem Rohwert des ersten Probanden des jeweils anderen Sets abge-

zogen wurde. Die gleiche Prozedur wurde für alle aufeinander folgenden Paare von Probanden beider Sets durchgeführt. Somit erfolgten die Bestimmung der Trefferrate bzw. der Rohwerte und die Bestimmung der Rate falscher Alarme bzw. der Basisrate an Bildern der *gleichen* Bildersets. Der Sensitivitätswert P_r bzw. der Primingwert der Skala wurde auch hier als Mittelwert der Sensitivitätswerte bzw. der Primingwerte aller Probanden bestimmt.

Die Primingwerte des impliziten Bildidentifikationstests und die Sensitivitätswerte P_r des expliziten Bilderkennungstests wurden zweifaktoriellen univariaten Varianzanalysen ohne Messwiederholung unterzogen. *Eine* Varianzanalyse wurde mit Daten durchgeführt, die mit der gebräuchlichen Methode ermittelt worden sind und eine *weitere* Varianzanalyse mit Daten, die mit der alternativen Methode ermittelt worden sind. Die Reliabilitätsanalysen wurden auf die gleiche Weise durchgeführt, wie dies bereits in Experiment 1 beschrieben worden ist.

Tabelle 7

Gruppenmittelwerte der Leistungsmaße, für beide Altersgruppen, beide Testformen und für beide Methoden der Bestimmung der Leistungswerte, in Experiment 3 (Die Werte in Klammern repräsentieren die Standardfehler der Mittelwerte).

	junge Probanden	ältere Probanden
impliziter Bildidentifikationstest Primingwert (gebräuchliche Methode)	1.28 (0.04)	1.24 (0.04)
expliziter Bilderkennungstest Sensitivitätswert P_r (gebräuchliche Methode)	.89 (.01)	.85 (.01)
impliziter Bildidentifikationstest Primingwert (alternative Methode)	1.28 (0.08)	1.24 (0.17)
expliziter Bilderkennungstest Sensitivitätswert P_r (alternative Methode)	.89 (.01)	.85 (.01)

In Tabelle 7 sind die Primingwerte des impliziten Bildidentifikationstests und die Sensitivitätswerte P_r des expliziten Bilderkennungstests, für beide Altersgruppen und für beide Methoden der Leistungswertbestimmung aufgeführt.¹⁰ Die Durchführung von Ein-Stichproben-t-Tests ergab, dass die Leistungswerte von Probanden

¹⁰ Die Variablen, die den Primingwerten und den Sensitivitätswerten P_r der entsprechenden Tests zugrunde liegen, sind der Tabelle C im Anhang zu entnehmen.

beider Bildersets, für beide Testformen und für beide Methoden der Leistungswertbestimmung, statistisch signifikant von null verschieden waren, $ts(41) > 4.57$, $ps < .001$, $\eta_p^2s > .34$. Die über Probanden beider Bildersets gemittelten Leistungswerte (siehe Tabelle 7) waren ebenfalls statistisch signifikant von null verschieden, $ts(83) > 7.27$, $ps < .001$, $\eta_p^2s > .39$.

Wurde die gebräuchliche Methode der Bestimmung der Leistungswerte angewandt, so ergab eine 2 x 2-faktorielle univariate Varianzanalyse ohne Messwiederholung mit den Faktoren «Altersgruppe» (*jung* vs. *alt*) und «Testform» (*Bildidentifikationstest* vs. *Bilderkennungstest*) keinen statistisch signifikanten Haupteffekt des Faktors «Altersgruppe», $F(1,332) = 1.61$, $p = .206$, $\eta_p^2 < .01$, jedoch einen statistisch signifikanten Haupteffekt des Faktors «Testform», $F(1,332) = 169.19$, $p < .001$, $\eta_p^2 = .34$. Die eigentlich interessante Interaktion zwischen den beiden Faktoren war nicht statistisch signifikant, $F(1,332) < 0.01$, $p = .964$, $\eta_p^2 < .01$. Post-hoc-Tests zeigten, dass sich die Primingwerte des impliziten Bildidentifikationstests nicht statistisch signifikant zwischen jungen und älteren Probanden unterschieden ($Ms = 1.28$ vs. 1.24 , $SEs = 0.04$ vs. 0.04 , $p = .388$, $\eta_p^2 < .01$). Die Sensitivitätswerte P_r des expliziten Bilderkennungstests unterschieden sich ebenfalls nicht statistisch signifikant zwischen jungen und älteren Probanden ($Ms = .89$ vs. $.85$, $SEs = .01$ vs. $.01$, $p = .354$, $\eta_p^2 < .01$). Wurde die alternative Methode der Bestimmung der Leistungswerte angewandt, so ergab eine 2 x 2-faktorielle univariate Varianzanalyse ohne Messwiederholung mit den Faktoren «Altersgruppe» (*jung* vs. *alt*) und «Testform» (*Bildidentifikationstest* vs. *Bilderkennungstest*) keinen statistisch signifikanten Haupteffekt des Faktors «Altersgruppe», $F(1,332) = 0.16$, $p = .691$, $\eta_p^2 < .01$, jedoch einen statistisch signifikanten Haupteffekt des Faktors «Testform», $F(1,332) = 16.70$, $p < .001$, $\eta_p^2 = .05$. Die eigentlich interessante Interaktion zwischen den beiden Faktoren war nicht statistisch signifikant, $F(1,332) < 0.01$, $p = .989$, $\eta_p^2 < .01$. Post-hoc-Tests zeigten, dass sich die Primingwerte des impliziten Bildidentifikationstests nicht statistisch signifikant zwischen jungen und älteren Probanden unterschieden ($Ms = 1.28$ vs. 1.24 , $SEs = 0.08$ vs. 0.17 , $p = .786$, $\eta_p^2 < .01$). Die Sensitivitätswerte P_r des expliziten Bilderkennungstests unterschieden sich ebenfalls nicht statistisch signifikant zwischen jungen und älteren Probanden ($Ms = .89$ vs. $.85$, $SEs = .01$ vs. $.01$, $p = .771$, $\eta_p^2 < .01$).

Tabelle 8

Testhalbierungskoeffizienten der Leistungsmaße, für beide Testformen und für beide Methoden der Bestimmung der Leistungswerte, in Experiment 3 (Die Werte in Klammern, repräsentieren die Spearman-Brown korrigierten Koeffizienten für die volle Testlänge).

	Bildidentifikationstest	Bilderkennungstest
gebräuchliche Methode der Leistungswertbestimmung (r_{tt})	.30 (.46)	.71 (.83)
alternative Methode der Leistungswertbestimmung (r_{tt})	.88 (.94)	.71 (.83)

In Tabelle 8 werden die Testhalbierungskoeffizienten der beiden Testformen präsentiert.¹¹ Es sind sowohl Testhalbierungskoeffizienten aufgeführt, die aus Leistungswerten geschätzt wurden, die mit der gebräuchlichen Methode ermittelt wurden, als auch Testhalbierungskoeffizienten, die aus Leistungswerten geschätzt wurden, die mit der alternativen Methode ermittelt wurden. Die Testhalbierungskoeffizienten von Probanden beider Bildersets und Probanden beider Altersgruppen wurden gemittelt, um die Teststärke für den entscheidenden Vergleich zwischen den Testformen zu erhöhen. Es sind sowohl die gemittelten Testhalbierungskoeffizienten, als auch die Spearman-Brown korrigierten Reliabilitätsschätzungen für den gesamten Test aufgeführt. Für den Vergleich der Reliabilitätsschätzungen zwischen den Testformen wurden die Testhalbierungskoeffizienten herangezogen. Wenn die gebräuchliche Methode der Bestimmung der Leistungswerte angewandt wurde, war der Testhalbierungskoeffizient des Bilderkennungstests (.71) statistisch signifikant größer, als der Testhalbierungskoeffizient des Bildidentifikationstests (.30), $z = 5.25$, $p < .001$, $q = 0.58$. Wenn die alternative Methode der Bestimmung der Leistungswerte angewandt wurde, war der Testhalbierungskoeffizient des Bildidentifikationstests (.88) statistisch signifikant größer, als der Testhalbierungskoeffizient des Bilderkennungstests (.71), $z = 4.44$, $p < .001$, $q = 0.49$. Wenn die Leistungswerte mit der alternativen Methode bestimmt wurden, war der Testhalbierungskoeffizient des Bildidentifikationstests (.88) statistisch signifikant größer, als wenn die Leistungswerte mit der gebräuchlichen Methode bestimmt wurden (.30), $z = 9.68$, $p < .001$, $q = 1.07$.

¹¹ Die statistischen Kennwerte, die den Testhalbierungskoeffizienten von Experiment 3 zugrunde liegen, sind Tabelle 9 zu entnehmen.

In Tabelle 9 sind die Kovarianzen zwischen den Testhälften und die Standardabweichungen der Testhälften aufgeführt, die den Testhalbierungskoeffizienten des dritten Experimentes zugrunde liegen. Wurde die alternative Methode der Leistungswertbestimmung angewandt, so waren die Kovarianzen und in geringerem Maße die Standardabweichungen, für Probanden beider Bildersets im impliziten Bildidentifikationstest höher, als wenn die gebräuchliche Methode der Leistungswertbestimmung angewandt wurde. Für den expliziten Bilderkennungstest ist kein Anstieg der Kovarianzen von der gebräuchlichen Methode der Leistungswertbestimmung zur alternativen Methode der Leistungswertbestimmung zu erkennen, was sich auch in dem unveränderten Wert des entsprechenden Testhalbierungskoeffizienten zeigt (siehe Tabelle 8).

Tabelle 9

Kovarianzen zwischen den Testhälften und die Standardabweichungen der Testhälften, wie sie für die Schätzung der Reliabilität in Experiment 3 verwendet wurden. Aufgeführt sind die Werte für beide Testformen, beide Altersgruppen und für beide Methoden der Leistungswertbestimmung.

	junge Probanden		ältere Probanden	
	Bilderset 1	Bilderset 2	Bilderset 1	Bilderset 2
impliziter Bildidentifikationstest (gebräuchliche Methode)				
$COV_{Th1,Th2}$	0.021	0.114	0.021	0.091
SD_{Th1}	0.36	0.41	0.40	0.52
SD_{Th2}	0.51	0.50	0.44	0.46
expliziter Bilderkennungstest (gebräuchliche Methode)				
$COV_{Th1,Th2}$	0.003	0.003	0.008	0.005
SD_{Th1}	0.06	0.07	0.09	0.09
SD_{Th2}	0.08	0.08	0.11	0.08
impliziter Bildidentifikationstest (alternative Methode)				
$COV_{Th1,Th2}$	0.633	0.350	2.318	2.454
SD_{Th1}	0.83	0.71	1.59	1.65
SD_{Th2}	0.93	0.68	1.57	1.59
expliziter Bilderkennungstest (alternative Methode)				
$COV_{Th1,Th2}$	0.004	0.003	0.008	0.005
SD_{Th1}	0.07	0.05	0.09	0.09
SD_{Th2}	0.08	0.08	0.11	0.08

4.4 Diskussion

Entgegen den Voraussagen unterschieden sich junge und ältere Probanden weder in den Primingwerten des impliziten Bildidentifikationstests, noch in den Sensitivitätswerten P_r des expliziten Bilderkennungstests statistisch signifikant voneinander, wenn die gebräuchliche Methode der Bestimmung der Leistungswerte angewandt wurde. Entgegen der Voraussage war dabei die Reliabilität des Bildidentifikationstests statistisch signifikant geringer, als die Reliabilität des Bilderkennungstests. Die Verdopplung der Itemanzahl im Vergleich zu Experiment 3 von Buchner und Wipich (2000), führte nicht zu der erhofften Reliabilitätssteigerung im Bildidentifikationstest. Auch die Vermutung, dass perzeptuelle Identifikationstests aufgrund ihrer restriktiveren Testbedingungen reliabler als andere implizite Tests sein könnten, konnte nicht bekräftigt werden. Überraschend war der Befund, dass junge und ältere Probanden sich nicht statistisch signifikant in den Sensitivitätswerten P_r des Bilderkennungstests voneinander unterschieden, wie es eigentlich zu erwarten gewesen wäre.

Wenn die alternative Methode zur Bestimmung der Leistungswerte angewandt wurde, kam es zu einer statistisch signifikanten Steigerung der Reliabilität des Bildidentifikationstests, im Vergleich zu der gebräuchlichen Methode. Dies entsprach den Voraussagen. Tatsächlich waren die Reliabilitätsschätzungen für den Bildidentifikationstest statistisch signifikant höher, als für den Bilderkennungstest. Statistisch signifikante Gruppenunterschiede zwischen jungen und älteren Probanden wurden jedoch weder in den Primingwerten des Bildidentifikationstests, noch in den Sensitivitätswerten P_r des Bilderkennungstests aufgedeckt. Dies entsprach nicht den Voraussagen. Die fehlenden Gruppenunterschiede zwischen jungen und älteren Probanden im expliziten Bilderkennungstest sind vermutlich durch einen Deckeneffekt zu erklären. Die fehlenden Gruppenunterschiede zwischen jungen und älteren Probanden im impliziten Bildidentifikationstest, wenn die alternative Methode der Leistungswertbestimmung angewandt wurde, sprechen dafür, dass tatsächlich keine Gruppenunterschiede bestehen. Wären Gruppenunterschiede vorhanden gewesen, hätten sie bei der hohen Reliabilität des Bildidentifikationstests mit hoher Wahrscheinlichkeit gefunden werden müssen.

5 Allgemeine Diskussion

In den folgenden Abschnitten werden Hintergründe und Ziele der Arbeit (5.1), sowie deren Ergebnisse (5.2), kurz zusammengefasst. Weiterhin werden die Befunde im Hinblick auf die zugrundeliegenden Fragestellungen und aufgestellten Hypothesen bewertet und in die Literatur eingeordnet (5.3). Als abschließender Punkt sollen offene Fragen und mögliche Inhalte zukünftiger Studien besprochen werden (5.4).

5.1 Hintergründe und Ziele der Arbeit

Hintergrund und Leitgedanke der vorliegenden Arbeit war die Vorstellung, dass Dissoziationen zwischen impliziten und expliziten Gedächtnistests ein Artefakt unzureichender Reliabilität impliziter Gedächtnistests, in Verbindung mit hoher Reliabilität expliziter Gedächtnistests sein könnten (Buchner & Wippich, 2000). In der Literatur werden typischerweise Dissoziationen gefunden, wenn implizite und explizite Gedächtnistests miteinander verglichen werden. Dies gilt für die Manipulation der verschiedensten experimentellen Variablen (z.B. Verarbeitungstiefe, Länge des Retentionsintervalls, Interferenz). Allgemeiner Tenor ist es, dass Dissoziationen zwischen impliziten und expliziten Gedächtnistests, funktionelle Unterschiede in den ihnen zugrunde liegenden Gedächtnissystemen (z.B. Tulving & Schacter, 1990) oder Gedächtnisprozessen (z.B. Roediger, 1990) widerspiegeln. Eine solche Erklärung von Befunden mit Dissoziationen ist jedoch nicht die einzig mögliche Interpretation. Die psychometrischen Eigenschaften des Messinstruments könnten ebenso die Wahrscheinlichkeit festlegen, mit der Gruppenunterschiede gefunden werden (Buchner & Wippich, 2000). Werden Dissoziationen beschrieben, ist es nämlich häufig so, dass große Gruppenunterschiede in expliziten Gedächtnistests gefunden werden, jedoch nur kleine und meist nicht signifikante Unterschiede in impliziten Gedächtnistests (z.B. Light & Singh, 1987). Eine methodische Erklärung solcher Befunde basiert auf dem typischen Reliabilitätsunterschied zwischen impliziten und expliziten Gedächtnistests. Explizite Gedächtnistests, in denen regelmäßig Gruppenunterschiede gefunden werden, haben meistens eine hohe Reliabilität. Damit sind sie in der Lage, eine hohe Messpräzision sowie eine hohe Wahrscheinlichkeit der Replikation von Befunden zu gewährleisten. Mangelnde Reliabilität impliziter Gedächtnistests und

geringe Teststärke können jedoch Gründe dafür sein, dass Effekte nicht gefunden werden können. Sie führt dazu, dass wahre Werte nur sehr ungenau geschätzt werden können, was bei fester Effektgröße eine höhere Teststärke erfordert, um Effekte gegenüber Zufallseinflüssen abzusichern. Kleine Stichproben und die damit verbundene geringe Teststärke verschärfen das Problem zusätzlich.

In einigen Studien wurde zur Einschätzung der Problematik die Reliabilität verschiedener impliziter Gedächtnistests ermittelt und im Allgemeinen als sehr gering eingestuft, vor allem im Vergleich zu Reliabilitätsschätzungen expliziter Gedächtnistests (z.B. Buchner & Brandt, 2003; Buchner & Wippich, 2000; Meier & Perrig, 2000). Als Gründe für die geringe Reliabilität impliziter Gedächtnistests werden vor allem wenig restriktive Testbedingungen angesehen, die eine hohe Variabilität von Gedächtnisprozessen zulassen und die zu einer Verringerung der Reliabilität führen (Buchner & Wippich, 2000). Da psychometrische Eigenschaften von Tests selten in Studien berichtet werden, gab es berechtigte Zweifel daran, dass zumindest ein Teil der bisherigen Befunde, die Dissoziationen zwischen impliziten und expliziten Gedächtnistests aufdecken konnten, mit psychometrischen Tests ermittelt wurden deren Reliabilität vergleichbar und ausreichend hoch war. Wenn die Reliabilität impliziter und expliziter Gedächtnistests nicht bekannt ist, kann nicht mit Sicherheit bestimmt werden, ob Dissoziationen tatsächlich bestehende Unterschiede zwischen verschiedenen Gedächtnissystemen oder Gedächtnisprozessen widerspiegeln, oder ob es sich dabei um ein methodisches Artefakt unterschiedlicher Reliabilität der verwendeten Gedächtnistests handelt (Buchner & Brandt, 2003).

Das Ziel der vorliegenden Arbeit bestand nicht darin, bestehende Theorien zu testen, sondern einer alternative Erklärungsmöglichkeit zur Entstehung von Dissoziationen zwischen impliziten und expliziten Gedächtnistests auf den Grund zu gehen. In einem ersten Experiment wurde der als recht unreliabel geltende implizite Wortstammerngänzungstest im Hinblick auf eine Reliabilitätssteigerung bearbeitet. Dieser Test und ein unbearbeiteter Wortstammerngänzungstest wurden Probanden vorgelegt. Anschließend wurden Reliabilitätsschätzungen beider Tests ermittelt und zur Beurteilung des Erfolgs der Maßnahmen verglichen. Weitere Maßnahmen zur Optimierung der Reliabilität wurden im Nachhinein durchgeführt. Im zweiten Experiment sollte eine klassische Dissoziation zwischen einem impliziten und einem expliziten Gedächtnistest repliziert werden. Dazu wurde ein unreliabler impliziter Wort-

stammerngänzungstest sowie eine reliable explizite Form des Wortstammerngänzungstests jungen und älteren Probanden vorgelegt. Neben der gebräuchlichen Methode der Bestimmung der Leistungswerte wurde eine alternative Methode angewandt, die eine Erhöhung der Reliabilität erwarten ließ. Es sollte festgestellt werden, ob die Erhöhung der Reliabilität des impliziten Wortstammerngänzungstests zur Aufdeckung signifikanter Gruppenunterschiede in den Primingwerten junger und älterer Probanden führt und damit zum Verschwinden der Dissoziation. In diesem Fall wäre zu vermuten, dass Dissoziationen Artefakte geringer Reliabilität impliziter Gedächtnistests sind. Falls keine Gruppenunterschiede gefunden werden, so spräche dies gegen die Erklärung von Artefakten geringer Reliabilität impliziter Gedächtnistests. In einem dritten Experiment sollten die Befunde aus dem zweiten Experiment auf einen weiteren impliziten und expliziten Gedächtnistest erweitert werden. Es wurde dazu ein impliziter Test verwendet, der durch die Restriktivität seiner Testbedingungen eine mit expliziten Tests vergleichbar hohe Reliabilität erwarten ließ. Junge und ältere Probanden wurden in dem impliziten Bildidentifikationstest und dem expliziten Bilderkennungstest miteinander verglichen. Primingwerte des impliziten Tests und Sensitivitätswerte P_r des expliziten Tests wurden sowohl mit der gebräuchlichen Methode, als auch mit einer alternative Methode bestimmt, die eine höhere Reliabilitätsschätzung des impliziten Tests erwarten ließ.

5.2 Zusammenfassung der Ergebnisse

Durch Maßnahmen zur Einschränkung von Gedächtnisprozessen bei der Enkodierung und dem Abruf von Wörtern, konnte keine statistisch signifikante Steigerung der Reliabilität des impliziten Wortstammerngänzungstests im ersten Experiment erreicht werden. Eine alternative Methode der Leistungswertbestimmung jedoch führte zu einer statistisch signifikanten Steigerung der Reliabilität gegenüber der gebräuchlichen Methode der Leistungswertbestimmung. Im zweiten Experiment konnte die Dissoziation zwischen dem impliziten Wortstammerngänzungstest und der expliziten Form des Wortstammerngänzungstests, als Funktion des Lebensalters, repliziert werden. Junge und ältere Probanden unterschieden sich in den Leistungswerten des expliziten Tests statistisch signifikant voneinander, jedoch nicht in den Primingwerten des impliziten Tests. Die Reliabilität des impliziten Wortstammerngänzungstests war dabei statistisch signifikant geringer als die Reliabilität der expliziten Form

des Wortstammergänzungstests, wenn die gebräuchliche Methode der Bestimmung der Leistungswerte angewandt wurde. Wurde die alternative Methode der Bestimmung der Leistungswerte angewandt, die zu einer statistisch signifikanten Steigerung der Reliabilität des impliziten Wortstammergänzungstests führte, blieb die Dissoziation auch weiterhin bestehen. Im dritten Experiment wurden weder statistisch signifikante Gruppenunterschiede junger und älterer Probanden in den Sensitivitätswerten P_r des expliziten Bilderkennungstests, noch in den Primingwerten des impliziten Bildidentifikationstests gefunden, wenn die gebräuchliche Methode der Bestimmung der Leistungswerte angewandt wurde. Die Reliabilität des impliziten Bildidentifikationstests war dabei statistisch signifikant geringer als die Reliabilität des expliziten Bilderkennungstests. Wurden die Leistungswerte mit der alternativen Methode bestimmt, kam es zu einer statistisch signifikanten Steigerung der Reliabilität des impliziten Bildidentifikationstests. Trotz dieser Reliabilitätssteigerung konnten in beiden Gedächtnistests keine statistisch signifikanten Gruppenunterschiede zwischen jungen und älteren Probanden aufgedeckt werden.

5.3 Bewertung der Befunde, und Schlussfolgerungen

Die Befunde der vorliegenden Arbeit lassen sich in zwei Gruppen aufteilen. Zunächst sollen Befunde zu den Reliabilitätsschätzungen der Tests diskutiert werden. Anschließend werden Befunde zum Vergleich der Leistungen junger und älterer Probanden besprochen.

5.3.1 Befunde zu Reliabilitätsschätzungen

In allen drei Experimenten der vorliegenden Arbeit wurde eine geringe Reliabilität impliziter Gedächtnistests gefunden, wenn Leistungswerte mit der gebräuchlichen Methode bestimmt wurden. Maßnahmen zur Einschränkung der Variabilität kognitiver Prozesse im impliziten Wortstammergänzungstest des ersten Experimentes führten nur auf deskriptiver Ebene zu einer Reliabilitätssteigerung. Auch eine hohe Anzahl von Items änderte nichts an den geringen Reliabilitätsschätzungen impliziter Gedächtnistests. Die Reliabilität expliziter Gedächtnistests jedoch war bedeutend höher und in einer Größenordnung angesiedelt, die sinnvolles Experimentieren zulässt.

Die Ergebnisse bestätigten zum großen Teil Befunde anderer Autoren. So fanden auch Buchner und Wippich (2000) geringe Reliabilitätsschätzungen für implizite Gedächtnistests, wie z.B. den Wortstammergänzungstest, verglichen mit Reliabilitätsschätzungen für explizite Gedächtnistests, wie den Worterkennungstest, oder die explizite Form des Wortstammergänzungstests. In dem dritten Experiment von Buchner und Wippich (2000) allerdings, war die Reliabilität des impliziten Bildidentifikationstests und des impliziten Wortidentifikationstests vergleichbar mit der Reliabilität des expliziten Bilderkennungstests und des expliziten Worterkennungstests. Die Reliabilitätsschätzungen waren zwar vergleichbar, befanden sich jedoch alle auf einem geringen Niveau. Die Autoren argumentierten, dass die Tests ihres dritten Experimentes nur aufgrund der geringen Itemanzahl (100 Items) wenig reliabel waren. Implizite Identifikationstests sollten demnach, aufgrund vergleichbar restriktiver Testbedingungen, ähnlich reliabel wie explizite Erkognitionstests sein. Eine Erhöhung der Itemanzahl dieser Tests sollte eine Steigerung der Reliabilität zur Folge haben (Buchner & Wippich, 2000). In dem dritten Experiment der vorliegenden Arbeit konnte jedoch, nach Verdopplung der Itemanzahl (200 Items), keine akzeptable Testreliabilität des impliziten Bildidentifikationstests gefunden werden. Rein deskriptiv war die Reliabilitätsschätzung des impliziten Bildidentifikationstests hier eher noch geringer als in dem dritten Experiment von Buchner und Wippich (2000). Die Höhe der Reliabilität des expliziten Bilderkennungstests allerdings war nach Verdopplung der Itemanzahl auf relativ hohem Niveau. Die Vermutung, dass perzeptuelle Identifikationstests aufgrund ihrer restriktiven Testbedingungen ähnlich reliabel sind wie explizite Gedächtnistests, konnte somit nicht bekräftigt werden. Meier und Perrig (2000) fanden ebenfalls geringe Reliabilitätsschätzungen impliziter Gedächtnistests, wie dem Bildidentifikationstest und dem Wortstammergänzungstest, im Vergleich zu Reliabilitätsschätzungen expliziter Gedächtnistests, wie dem freien Abruf von Wörtern, der expliziten Form des Wortstammergänzungstests und dem Worterkennungstest. Zusätzlich dazu waren die Interkorrelationen der von ihnen untersuchten expliziten Gedächtnistests höher als die Interkorrelationen impliziter Gedächtnistests. Buchner und Brandt (2003) fanden außerdem, dass neben den leistungsorientierten impliziten Gedächtnistests auch illusionsorientierte implizite Gedächtnistests von geringer Reliabilität betroffen sind.

Wurden Leistungswerte mit der alternativen Methode bestimmt, führte dies zu höheren Reliabilitätsschätzungen in den impliziten Gedächtnistests aller drei Experi-

mente dieser Arbeit, im Vergleich dazu, wenn die Leistungswerte mit der gebräuchlichen Methode bestimmt wurden. Eine höhere Reliabilitätsschätzung lässt sich ebenfalls erreichen, wenn Rohwerte zur Berechnung entsprechender Korrelationskoeffizienten herangezogen werden. Es ist jedoch fraglich, ob man damit eine realistische Schätzung der Reliabilität erhält. Eine Erhöhung der Reliabilitätsschätzungen für Rohwerte, im Vergleich zu Reliabilitätsschätzungen für Primingwerte, ist wahrscheinlich Folge von konsistenten Antworttendenzen (Buchner & Brandt, 2003). Eine Reliabilitätsbestimmung an Rohwerten würde die tatsächliche Reliabilität impliziter Gedächtnistests überschätzen und kann somit keine Lösung des Problems darstellen. Es stellt sich jedoch die Frage, weshalb die alternative Methode der Bestimmung von Leistungswerten ein besser geeignetes Verfahren zur Bestimmung von Leistungswerten impliziter Gedächtnistests sein sollte, als die gebräuchliche Methode der Leistungswertbestimmung.

Die Reliabilität impliziter Gedächtnistests wird typischerweise an Primingwerten ermittelt, die ihrerseits über die Differenz von Rohwert und Basisrate bestimmt werden. Eine Basisratenkorrektur der Rohwerte ist notwendig, weil ein gewisser Anteil von korrekten Itembeantwortungen auch ohne Erfahrungsnachwirkungen vorhanden ist. Dieser Anteil muss von den Rohwerten abgezogen werden, um eine Überschätzung der Probandenleistung zu vermeiden. Eine berechtigte Frage war jedoch, wie eine solche Basisratenkorrektur idealerweise durchzuführen ist. Dies führte zu dem alternativen Ansatz der Bestimmung der Leistungswerte. Wie bereits an anderer Stelle näher beschrieben, werden bei der gebräuchlichen Methode der Leistungswertbestimmung, Basisrate und Rohwerte an Items der jeweils anderen Itemliste bestimmt. Da Items verschiedener Itemlisten nicht zwingend gleiche Eigenschaften haben, kann dies die Rohwertdaten mit unsystematischer Varianz beaufschlagen. Unsystematische Varianz führt zu geringerer Kovarianz von Testteilen und sorgt dafür, dass die Varianz zwischen den Personen größer wird. Dies führt in der Konsequenz zu einer geringeren Testreliabilität. Einen Hinweis auf die Verschiedenheit der Wortlisten im Wortstammergänzungstest der ersten beiden Experimente liefern die unterschiedlichen Primingwerte von Probanden beider Wortlisten, wenn die gebräuchliche Methode der Leistungswertbestimmung angewandt wurde. Die beiden Wortlisten unterschieden sich in ihrer mittleren Itemschwierigkeit voneinander, der Wahrscheinlichkeit der Ergänzung von Wortstämmen, mit Wörtern der entsprechenden Wortliste. Dies schlug sich, trotz vollständiger Randomisierung bei der Zuweisung

von Wörtern zu Wortlisten und Probanden zu Wortlisten, in den unterschiedlichen Rohwerten und Basisraten der Probanden beider Wortlisten nieder. Liste 1 Wortstämme wurden beispielsweise häufiger mit Wörtern aus Liste 1 ergänzt, als Liste 2 Wortstämme mit Wörtern aus Liste 2 ergänzt wurden. Dies war sowohl dann der Fall, wenn entsprechende Wortlisten in der Lernphase präsentiert wurden, als auch dann, wenn sie nicht präsentiert wurden. Die Ergänzung von Liste 1 Wortstämmen führte damit zu höheren Rohwerten, jedoch auch zu höheren Basisraten, als die Ergänzung von Liste 2 Wortstämmen. Werden zur Bestimmung von Primingwerten Basisraten von Rohwerten abgezogen, die an unterschiedlichen Wortlisten ermittelt worden sind, so kommt es zu den beobachteten Abweichungen in den Primingwerten zwischen Probanden beider Wortlisten. Bei der alternativen Methode der Leistungswertbestimmung wurde die Basisrate an den gleichen Items ermittelt, an denen auch die Rohwerte ermittelt wurden. Die Ermittlung der Werte erfolgte jedoch an verschiedenen Probanden. Die gemeinsame Varianz der zu korrelierenden Testhälften war bedeutend höher, wenn die alternative Methode der Leistungswertbestimmung angewandt wurde, im Vergleich zu der gebräuchlichen Methode. Es kam folglich zu einer deutlichen Steigerung der Reliabilitätsschätzungen in den impliziten Gedächtnistests dieser Arbeit. Man sollte jedoch vorsichtig damit sein, einen Kausalzusammenhang zwischen dem Verschwinden von Primingwertdifferenzen beider Wortlisten in den ersten beiden Experimenten und der Steigerung der Reliabilität im impliziten Wortstammernergänzungstest herzustellen. Im dritten Experiment unterschieden sich die Primingwerte zwischen den Bildersets des Bildidentifikationstests nämlich nicht in höherem Maße, wenn die gebräuchliche Methode der Leistungswertbestimmung angewandt wurde, als wenn die alternative Methode der Leistungswertbestimmung angewandt wurde. Trotzdem kam es auch im dritten Experiment zu einer bedeutenden Steigerung der Reliabilität des impliziten Bildidentifikationstests, von der gebräuchlichen zu der alternativen Methode der Leistungswertbestimmung. Die Frage, wieso sich die Verschiedenheit von Personen deutlich weniger negativ auf die Reliabilität der impliziten Tests dieser Arbeit auswirkte als die Verschiedenheit von Wortlisten oder Bildersets, ist schwierig zu beantworten. Auf einer statistischen Ebene kann man sich die Verteilungen, die Dispersionsmaße und die Kovarianz der Testhälften für beide Methoden der Leistungswertbestimmung ansehen. Sie zeigen, dass die Verschiedenheit der Itemsets sich mehr auswirkt als die Verschiedenheit der Personen, liefern jedoch *keine* Erklärung dafür.

5.3.2 Befunde zum Vergleich von Leistungen junger und älterer Probanden

Im zweiten Experiment der vorliegenden Arbeit konnte eine Dissoziation zwischen dem impliziten Wortstammerngänzungstest und der expliziten Form des Wortstammerngänzungstests gefunden werden, wenn die gebräuchliche Methode der Bestimmung von Leistungswerten angewandt wurde. Damit konnte die Dissoziation, die Light und Singh (1987) in ihrem zweiten Experiment gefunden haben, repliziert werden. Da die Reliabilität des impliziten Wortstammerngänzungstests sehr gering war, besteht die Möglichkeit, dass die Dissoziation lediglich ein Artefakt geringer Reliabilität des impliziten Wortstammerngänzungstests in Verbindung mit hoher Reliabilität der expliziten Form des Wortstammerngänzungstests ist. Wenn Gruppenunterschiede zwischen jungen und älteren Probanden im impliziten Wortstammerngänzungstest tatsächlich vorhanden sind, dann sollten sie durch einen impliziten Wortstammerngänzungstest höherer Reliabilität auch gefunden werden. Wenn jedoch durch einen impliziten Wortstammerngänzungstest höherer Reliabilität keine Gruppenunterschiede zwischen jungen und älteren Probanden aufgedeckt werden können, so spräche dies gegen die Theorie der Artefakte geringer Reliabilität impliziter Tests. Es spräche dafür, dass sich junge und ältere Probanden tatsächlich nicht in ihrer Leistung voneinander unterscheiden.

Die Dissoziation zwischen dem impliziten Wortstammerngänzungstest und der expliziten Form des Wortstammerngänzungstests in Experiment 2 blieb auch dann bestehen, als die alternative Methode der Bestimmung der Leistungswerte angewandt wurde. Die dadurch erfolgte Steigerung der Reliabilität des impliziten Wortstammerngänzungstests, führte nicht zu der Aufdeckung von statistisch signifikanten Unterschieden in den Primingwerten junger und älterer Probanden. Dieser Befund spricht dafür, dass die Dissoziation beider Gedächtnistests kein Artefakt geringer Reliabilität des impliziten Tests sein kann. Meier (2001) fand im Gegensatz zu diesem Befund keine Dissoziation zwischen dem impliziten Bildidentifikationstest und einem expliziten Rekognitionstest. Die Reliabilität beider Tests war vergleichbar. Beide Tests in der Untersuchung von Meier (2001) zeigten einen Effekt der Verarbeitungstiefe, obwohl implizite Gedächtnistests in der Literatur oft als unabhängig von der Manipulation der Verarbeitungstiefe beschrieben werden (z.B. Roediger et al., 1992).

In dem dritten Experiment der hier vorliegenden Arbeit konnten keine statistisch signifikanten Unterschiede zwischen jungen und älteren Probanden im Priming des impliziten Bildidentifikationstests, sowie in den Sensitivitätswerten P_r des expliziten Bilderkennungstests, festgestellt werden. Dies gilt für die gebräuchliche Methode der Bestimmung von Leistungswerten, bei der die Reliabilität des impliziten Bildidentifikationstests deutlich geringer war, als die Reliabilität des expliziten Bilderkennungstests. Aber auch wenn Primingwerte und Sensitivitätswerte P_r mit der alternativen Methode bestimmt wurden und es zu einer Erhöhung der Reliabilität des impliziten Bildidentifikationstests kam, wurden keine statistisch signifikanten Gruppenunterschiede in den Primingwerten junger und älterer Probanden aufgedeckt. Diese Ergebnisse sprechen dafür, dass tatsächlich keine Gruppenunterschiede im impliziten Bildidentifikationstest vorhanden sind.

Die Befunde der Experimente dieser Arbeit sprechen dafür, dass Dissoziationen zwischen impliziten und expliziten Gedächtnistest keine Artefakte geringer Reliabilität impliziter Gedächtnistests in Verbindung mit hoher Reliabilität expliziter Gedächtnistests sind. Vergleichbar hohe Reliabilität impliziter und expliziter Gedächtnistests, wenn die alternative Methode der Bestimmung von Leistungswerten angewandt wurde, sowie hohe Stichprobenumfänge, führten nicht zur Aufdeckung von Unterschieden zwischen jungen und älteren Probanden in impliziten Gedächtnistests.

5.4 Ausblick

Unterschiedlichste Bemühungen, die Reliabilität des impliziten Wortstammergänzungstests zu steigern, blieben ohne Erfolg. Eine hohe Itemanzahl und Maßnahmen zur Einschränkung der Variabilität kognitiver Prozesse bei Enkodierung und Abruf, brachten die Reliabilitätsschätzung des impliziten Gedächtnistests nicht auf Augenhöhe mit Reliabilitätsschätzungen expliziter Gedächtnistests. Eine alternative Methode der Leistungswertbestimmung jedoch, bei der Rohwerte und Basisraten an Items der gleichen Itemliste bestimmt werden, brachte eine bedeutende Reliabilitätssteigerung in den impliziten Gedächtnistests der vorliegenden Experimente.

Implizite Gedächtnistests sind vielleicht gar nicht so unreliabel, wie bisher vermutet wurde. Die geringen Reliabilitätsschätzungen impliziter Gedächtnistests, wie sie bisher berichtet wurden, könnten auch durch eine ungeeignete Methode der Lei-

stungswertbestimmung zustande gekommen sein. Sollten zukünftige Studien dies bestätigen, so wäre es interessant herauszufinden, ob Dissoziationen zwischen impliziten und expliziten Gedächtnistests durch die Anwendung reliabler impliziter Gedächtnistests verschwinden.

6 Literaturverzeichnis

- Bazin, N. & Perruchet, P. (1996). Implicit and explicit associative memory in patients with schizophrenia. *Schizophrenia Research*, 22, 241-248.
- Bazin, N., Perruchet, P., Debonis, M. & Féline, A. (1994). The dissociation of explicit and implicit memory in depressed patients. *Psychological Medicine*, 24, 239-245.
- Begg, I. M., Anas, A. & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General*, 121, 446-458.
- Berry, D. C., Banbury, S. & Henry, L. (1997). Transfer across form and modality in implicit and explicit memory. *Quarterly Journal of Experimental Psychology*, 50A, 1-24.
- Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 657-668.
- Brown, A. S. & Mitchell, D. B. (1994). A reevaluation of semantic versus nonsemantic processing in implicit memory. *Memory & Cognition*, 22, 533-541.
- Buchner, A. & Brandt, M. (2003). Further evidence for systematic reliability differences between explicit and implicit memory tests. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 56A, 193-209.
- Buchner, A., Erdfelder, E. & Vaterrodt-Pluennecke, B. (1995). Toward unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *Journal of Experimental Psychology: General*, 124, 137-160.
- Buchner, A. & Jansen-Osmann, P. (2006). Implizites Gedächtnis. In J. Funke & P. A. Frensch (Hrsg.), *Handbuch der Allgemeinen Psychologie: Kognition* (S. 356-362). Göttingen: Hogrefe.

- Buchner, A. & Wippich, W. (2000). On the reliability of implicit and explicit memory measures. *Cognitive Psychology*, 40, 227-259.
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion* (2., aktualisierte Aufl.). München: Pearson.
- Butters, N., Heindel, W. C. & Salmon, D. P. (1990). Dissociation of implicit memory in dementia: Neurological implications. *Bulletin of the Psychonomic Society*, 28, 359-366.
- Chiarello, C. & Hoyer, W. J. (1988). Adult age differences in implicit and explicit memory: Time course and encoding effects. *Psychology and Aging*, 3, 358-366.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (überarbeitete Aufl.). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- Cohen, N. J. & Squire, L. R. (1980). Preserved learning and retention of pattern analyzing skills in amnesia: Dissociation of knowing how and knowing that. *Science*, 210, 207-210.
- Corkin, S. (1968). Acquisition of motor skill after bilateral medial temporal lobe excision. *Neuropsychologia*, 6, 255-265.
- Cowan, N. (1997). *Attention and memory: An integrated framework*. Oxford: Oxford University Press.
- Cowan, N. & Stadler, M. A. (1996). Estimating unconscious processes: Implications of a general class of models. *Journal of Experimental Psychology: General*, 125, 195-200.
- Craik, F. I. M., Moscovitch, M. & McDowd, J. M. (1994). Contributions of surface and conceptual information to performance on implicit and explicit memory tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 864-875.
- Curran, T. & Hintzman, D. L. (1995). Violations of the independence assumption in process dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 531-547.

- Curran, T. & Hintzman, D. L. (1997). Consequences and causes of correlations in process dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 496-504.
- Davis, H. P., Cohen, A., Gandy, M., Colombo, P., VanDusseldorp, G., Simolke, N. & Romano, J. (1990). Lexical priming deficits as a function of age. *Behavioral Neuroscience*, 104, 288-297.
- Debner, J. A. & Jacoby, L. L. (1994). Unconscious perception: Attention, awareness, and control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 304-317.
- Denny, E. R. & Hunt, R. R. (1992). Affective valence and memory in depression: Dissociation of recall and fragment completion. *Journal of Abnormal Psychology*, 101, 575-580.
- Dew, I. T. & Cabeza, R. (2011). The porous boundaries between explicit and implicit memory: Behavioral and neural evidence. *Annals of the New York Academy of Sciences*, 1224, 174-190.
- Diamond, R. & Rozin, P. (1984). Activation of existing memories in anterograde amnesia. *Journal of abnormal psychology*, 93, 98-105.
- Drury, J. L., Kinsella, G. J. & Ong, B. (2000). Age differences in explicit and implicit memory for pictures. *Neuropsychology*, 14, 93-101.
- Ebbinghaus, H. (1966). *Über das Gedächtnis*. Amsterdam: E. J. Bonset (Original veröffentlicht 1885).
- Faul, F., Erdfelder, E., Buchner, A. & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149-1160.
- 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, 175-191.

- Feustel, T. C., Shiffrin, R. M. & Salasoo, A. (1983). Episodic and lexical contributions to the repetition effect in word identification. *Journal of Experimental Psychology: General*, 112, 309-346.
- Fleischman, D. A., Wilson, R. S., Gabrieli, J. D. E., Bienas, J. L. & Bennett, D. A. (2004). A longitudinal study of implicit and explicit memory in old persons. *Psychology and Aging*, 19, 617-625.
- Forster, K. I. & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 580-698.
- Graf, P. & Komatsu, S. (1994). Process dissociation procedure: Handle with caution. *European Journal of Cognitive Psychology*, 6, 113-129.
- Graf, P. & Mandler, G. (1984). Activation makes words more accessible but not necessarily more retrievable. *Journal of Verbal Learning and Verbal Behavior*, 23, 553-568.
- Graf, P. & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 501-518.
- Graf, P. & Schacter, D. L. (1987). Selective effects of interference on implicit and explicit memory for new associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 45-53.
- Graf, P., Shimamura, A. P. & Squire, L. R. (1985). Priming across modalities and priming across category levels: Extending the domain of preserved functioning in amnesia. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 11, 385-395.
- Hallahan, M. & Rosenthal, R. (1996). Statistical power: Concepts, procedures, and applications. *Behavior Research and Therapy*, 34, 489-499.
- Halpern, A. R. & O'Connor, M. G. (2000). Implicit memory for music in Alzheimer's disease. *Neuropsychology*, 14, 391-397.

- Hertel, P. T. & Hardin, T. S. (1990). Remembering with and without awareness in a depressed mood: Evidence of deficits in initiative. *Journal of Experimental Psychology: General*, 119, 45-59.
- Ikier, S. & Hasher, L. (2006). Age differences in implicit interference. *Journal of Gerontology: Psychological Sciences*, 61b, 278-284.
- Ikier, S., Yang, L. & Hasher, L. (2008). Implicit proactive interference, age, and automatic versus controlled retrieval strategies. *Psychological Science*, 19, 456-461.
- Jacoby, L. L. (1983). Remembering the data: analyzing interactive processes in reading. *Journal of Verbal Learning and Verbal Behavior*, 22, 485-508.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30, 513-541.
- Jacoby, L. L. & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, 110, 306-340.
- Jacoby, L. L. & Kelley, C. M. (1987). Unconscious influences of memory for a prior event. *Personality and Social Psychology Bulletin*, 13, 314-336.
- Jacoby, L. L., Kelley, C. M., Brown, J. & Jasechko, J. (1989). Becoming famous overnight: Limits on the ability to avoid unconscious influences of the past. *Journal of Personality and Social Psychology*, 56, 326-338.
- Jacoby, L. L., Toth, J. P. & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General*, 122, 139-154.
- Jacoby, L. L. & Witherspoon, D. (1982). Remembering without awareness. *Canadian Journal of Psychology*, 36, 300-324.
- Jacoby, L. L., Woloshyn, V. & Kelley, C. M. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of Experimental Psychology: General*, 118, 115-125.

- Java, R. I. (1992). Priming and aging: Evidence of preserved memory function in an anagram solution task. *American Journal of Psychology*, 105, 541-548.
- Johnson, M. K. & Hasher, L. (1987). Human learning and memory. *Annual Review of Psychology*, 38, 631-668.
- Johnson, M. K., Kim, J. K. & Risse, G. (1985). Do Korsakoff amnesics develop affective reactions? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 22-36.
- Kandel, E. R., Kupfermann, I. & Iversen, S. (2000). Learning and memory. In E. R. Kandel, J. H. Schwartz & T. M. Jessell (Hrsg.), *Principles of neural science* (S. 1227-1245). New York: McGraw-Hill.
- Kelley, C. M. & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language*, 32, 1-24.
- Kinoshita, S. (2001). The role of involuntary aware memory in the implicit stem and fragment completion tasks: A selective review. *Psychonomic Bulletin & Review*, 8, 58-69.
- Kirsner, K. & Smith, M. C. (1974). Modality effects in word identification. *Memory & Cognition*, 2, 637-640.
- Kopriva, R. J. & Shaw, D. G. (1991). Power estimates: The effect of dependant variable reliability on the power of one-factor ANOVAs. *Educational and Psychological Measurement*, 51, 585-595.
- Krauth, J. (1995). *Testkonstruktion und Testtheorie*. Weinheim: Psychologie Verlags Union.
- Kunst-Wilson, W. R. & Zajonc, R. B. (1980). Affective discrimination of stimuli that cannot be recognized. *Science*, 207, 557-558.
- LeBel, E. P. & Paunonen, S. V. (2011). Sexy but often unreliable: The impact of unreliability on the replicability of experimental findings with implicit measures. *Personality and Social Psychology Bulletin*, 37, 570-583.

- Lienert, G. A. (1989). *Testaufbau und Testanalyse* (4. Aufl.). München: Psychologie Verlags Union.
- Light, L. L. & LaVoie, D. (1993). Direct and indirect measures of memory in old age. In P. Graf & M. Masson (Hrsg.), *Implicit memory: New directions in cognition, development, and neuropsychology* (S. 207-230). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Light, L. L., Prull, M. W., LaVoie, 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). Oxford, England: Oxford University Press.
- Light, L. L. & Singh, A. (1987). Implicit and explicit memory in young and old adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 531-541.
- Light, L. L., Singh, A. & Capps, J. L. (1986). The dissociation of memory and awareness in young and older adults. *Journal of Clinical and Experimental Neuropsychology*, 8, 62-74.
- Lustig, C. & Buckner, R. L. (2004). Preserved neural correlates of priming in old age and dementia. *Neuron*, 42, 865-875.
- Lustig, C. A. & Hasher, L. (2001a). Implicit memory is not immune to interference. *Psychological Bulletin*, 127, 618-628.
- Lustig, C. A. & Hasher, L. (2001b). Implicit memory is vulnerable to proactive interference. *Psychological Science*, 12, 408-412.
- Maki, P. M., Zonderman, A. B. & Weingartner, H. (1999). Age differences in implicit memory: fragmented object identification and category exemplar generation. *Psychology and Aging*, 14, 284-294.
- Marie, A., Gabrieli, J. D., Vaidya, C., Brown, B., Pratto, F., Zajonc, R. B. & Shaw, R. J. (2001). The mere exposure effect in patients with schizophrenia. *Schizophrenia Bulletin*, 27, 297-303.
- McBride, D. M. & Doshier, B. A. (1997). A comparison of forgetting in an implicit and explicit memory task. *Journal of Experimental Psychology: General*, 126, 371-392.

- McBride, D. M., Doshier, B. A. & Gage, N. A. (2001). A comparison of forgetting for conscious and automatic memory processes in word fragment completion tasks. *Journal of Memory and Language*, 45, 585-615.
- McKone, E. & Slee, J. A. (1997). Explicit contamination in implicit memory for new associations. *Memory & Cognition*, 25, 352-366.
- Meier, B. (2001). Verschwinden Dissoziationen zwischen implizitem und explizitem Gedächtnis, wenn die Reliabilität der Tests vergleichbar ist? Ein Beispiel. *Zeitschrift für Experimentelle Psychologie*, 48, 207-313.
- Meier, B. & Perrig, W. J. (2000). Low reliability of perceptual priming: Consequences for the interpretation of functional dissociations between explicit and implicit memory. *Quarterly Journal of Experimental Psychology*, 53, 211-233.
- Merikle, P. M. & Reingold, E. M. (1991). Comparing direct (explicit) and indirect (implicit) measures to study unconscious memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 224-233.
- Milner, B., Corkin, S. & Teubner, H. L. (1968). Further analysis of the hippocampal amnesic syndrome: Fourteen year follow-up study of H. M. *Neuropsychologia*, 6, 215-234.
- Mitchell, D. B. & Bruss, P. J. (2003). Age differences in implicit memory: conceptual, perceptual, or methodological? *Psychology and Aging*, 18, 807-822.
- 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, 519-533.
- Parkin, A. J. & Russo, R. (1990). Implicit and explicit memory and the automatic/effortful distinction. *European Journal of Cognitive Psychology*, 2, 71-80.
- Perruchet, P. & Baveux, P. (1989). Correlational analysis of explicit and implicit memory performance. *Memory & Cognition*, 17, 77-86.

- Petersen, S. E., Fox, P. T., Posner, M. I., Mintum, M. & Raichle, M. E. (1989). Positron emission tomographic studies of the processing of single words. *Journal of Cognitive Neuroscience*, 1, 153-170.
- Process Dissociation Tutorial. Zugriff am 15.10.2013, auf University of Arkansas <http://www.uark.edu/misc/lampinen/tutorials/jacoby.htm>
- Rabbit, P. M. A. (1982). How do old people know what to do next? In F. I. M. Craik & S. Trehub (Hrsg.), *Aging and cognitive processes* (S. 79-98). New York: Plenum Press.
- Rabbit, P. M. A. (1984). How old people prepare themselves for events which they expect. In H. Bouma & D. G. Bouwhuis (Hrsg.), *Attention and performance X: Control of language processes* (S. 515-527). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rappold, V. A. & Hashtroudi, S. (1991). Does organization improve priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 103-114.
- Reingold, E. M. & Merikle, P. M. (1988). Using direct and indirect measures to study perception without awareness. *Perception & Psychophysics*, 44, 563-575.
- Richardson-Klavehn, A. & Bjork, R. A. (1988). Measures of memory. *Annual Review of Psychology*, 39, 475-543.
- Roediger, H. L. (1990). Implicit memory. Retention without awareness. *American Psychologist*, 45, 1043-1056.
- Roediger, H. L., Marsh, E. J. & Lee, S. C. (2002). Varieties of memory. In D. L. Medin & H. Pashler (Hrsg.), *Stevens handbook of experimental psychology: Memory and cognitive processes* (3. Aufl., S. 1-41). New York: John Wiley & Sons.
- Roediger, H. L. & McDermott, K. B. (1994). The problem of differing false alarm rates for the process dissociation procedure: Comment on Verfaellie and Treadwell (1993). *Neuropsychology*, 8, 284-288.

- Roediger, H. L., Srinivas, K. & Weldon, M. S. (1989). Dissociations between implicit measures of retention. In S. Lewandowsky, J. C. Dunn & K. Kirsner (Hrsg.), *Implicit memory: Theoretical issues* (S. 67-84). Hillsdale, NJ: Erlbaum.
- Roediger, H. L., Weldon, M. S. & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger & F. I. M. Craik (Hrsg.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (S. 3-39). Hillsdale, NJ: Erlbaum.
- Roediger, H. L., Weldon, M. S., Stadler, M. A. & Riegler, G. H. (1992). Direct comparison of word stems and word fragments in implicit and explicit retention tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 1251-1269.
- Rogers, W. T. & Hopkins, K. D. (1988). Power estimates in the presence of a covariate and measurement error. *Educational and Psychological Measurement*, 48, 647-56.
- Salasoo, A., Shiffrin, R. M. & Feustel, T. C. (1985). Building permanent memory codes: Codification and repetition effects in word identification. *Journal of Experimental Psychology: General*, 114, 50-77.
- Sartori, G. & Job, R. (1988). The oyster with four legs: A neuropsychological study on the interaction of visual and semantic information. *Cognitive Neuropsychology*, 5, 105-132.
- Schacter, D. L. (1985). Priming of old and new knowledge in amnesic patients and normal subjects. *Annals of the New York Academy of Sciences*, 444, 41-53.
- Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 501-518.
- Schacter, D. L. (1992). Priming and multiple memory systems: Perceptual mechanisms of implicit memory. *Journal of Cognitive Neuroscience*, 4, 244-256.
- Schacter, D. L., Bowers, J. & Booker, J. (1989). Intention, awareness, and implicit memory: The retrieval intentionality criterion. In S. Lewandowsky, J. C. Dunn & K. Kirsner (Hrsg.), *Implicit memory: Theoretical issues* (S. 47-65). Hillsdale, NJ: Erlbaum.

- Schacter, D. L., Chiu, C.-Y. P. & Ochsner, K. N. (1993). Implicit Memory: A selective review. *Annual Review of Neuroscience*, 16, 159-182.
- Schacter, D. L. & Tulving, E. (1994). What are the memory systems of 1994? In D. Schacter & E. Tulving (Hrsg.), *Memory systems 1994* (S. 1-38). Cambridge, MA: MIT Press.
- Schwartz, B. L., Rosse, R. B. & Deutsch, S. I. (1993). Limits of the processing view in accounting for dissociations among memory measures in a clinical population. *Memory & Cognition*, 21, 63-72.
- Schwartz, M. F., Saffran, E. M. & Marin, O. S. (1980). The word order problem in agrammatism. *Brain and Language*, 10, 249-262.
- Shimamura, A. P. (1985). Problems with the finding of stochastic independence as evidence for multiple memory systems. *Bulletin of the Psychonomic Society*, 23, 506-508.
- Shimamura, A. P. (1986). Priming effects in amnesia: Evidence for a dissociable memory function. *Quarterly Journal of Experimental Psychology*, 38A, 619-644.
- Shimamura, A. P. & Squire, L. R. (1984). Paired-associate learning and priming effects in amnesia: A neuropsychological study. *Journal of Experimental Psychology: General*, 113, 556-570.
- Slamecka, N. J. & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory*, 4, 592-604.
- Sloman, S. A., Hayman, C. A. G., Ohta, N., Law, J. & Tulving, E. (1988). Forgetting in primed fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 223-239.
- Snodgrass, J. G. & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, 117, 34-50.

- 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 & Memory*, 6, 174-215.
- Squire, L. R. (1992). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. *Journal of Cognitive Neuroscience*, 4, 217-231.
- Squire, L. R. & Cohen, N. J. (1984). Human memory and amnesia. In G. Lynch, J. L. McGaugh & N. M. Weinberger (Hrsg.), *Neurobiology of learning and memory* (S. 3-64). New York: Guilford Press.
- Squire, L. R., Knowlton, B. & Musen, G. (1993). The structure and organization of memory. *Annual Review of Psychology*, 44, 453-495.
- Squire, L. R. & McKee, R. (1992). The influence of prior events on cognitive judgments in amnesia. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 106-115.
- Srinivas, K. & Roediger, H. L. (1990). Classifying implicit memory tests: Category association anagram solution. *Journal of Memory and Language*, 29, 389-412.
- Sutcliffe, J. P. (1958). Error of measurement and the sensitivity of a test of significance. *Psychometrika*, 23, 9-17.
- Tulving, E. & Schacter, D. L. (1990). Priming and human memory systems. *Science*, 247, 301-306.
- Tulving, E., Schacter, D. L. & Stark, H. A. (1982). Priming effects in word fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8, 336-342.
- Tulving, E. & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80, 352-373.
- Tunney, R. J. (2003). Implicit and explicit knowledge decay at different rates: A dissociation between priming and recognition in artificial grammar learning. *Experimental Psychology*, 50, 124-130.

- Warrington, E. K. & Weiskrantz, L. (1968). New methods of testing long-term retention with special reference to amnesic patients. *Nature*, 217, 972-974.
- Warrington, E. K. & Weiskrantz, L. (1970). Amnesic syndrome: Consolidation or retrieval? *Nature*, 228, 628-630.
- Warrington, E. K. & Weiskrantz, L. (1974). The effect of prior learning on subsequent retention in amnesic patients. *Neuropsychologia*, 12, 419-428.
- Weldon, M. S. & Jackson-Barrett, J. L. (1993). Why do pictures produce priming on the word fragment completion test? A study of encoding and retrieval factors. *Memory & Cognition*, 21, 519-528.
- Weldon, M. S. & Roediger, H. L. (1987). Altering retrieval demands reverses the picture superiority effect. *Memory & Cognition*, 15, 269-280.
- Wiggs, C. L. (1993). Aging and memory for frequency of occurrence of novel, visual stimuli: Direct and indirect measures. *Psychology and Aging*, 8, 400-410.
- Winocur, G., Moscovitch, M. & Bruni, J. (1996). Heightened interference on implicit, but not explicit, tests of negative transfer: Evidence from patients with unilateral temporal lobe lesions and normal old people. *Brain and Cognition*, 30, 44-58.
- Winograd, E., Goldstein, F. C., Monarch, E. S., Peluso, J. P. & Goldman, W. P. (1999). The mere exposure effect in patients with Alzheimer's disease. *Neuropsychology*, 13, 41-46.
- Witherspoon, D. & Moscovitch, M. (1989). Stochastic independence between two implicit memory tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 22-30.
- Zimbardo, P. G. & Gerrig, R. J. (2004). *Psychologie* (16., aktualisierte Aufl.). München: Pearson.

7 Anhang

Tabelle A

Gruppenmittelwerte der Variablen, die den Primingwerten der Tests aus Experiment 1 zugrunde liegen. Es sind Werte für beide Testformen und für beide Methoden der Leistungswertbestimmung aufgeführt (Die Werte in Klammern repräsentieren die Standardfehler der Mittelwerte).

	Wortliste 1 Probanden	Wortliste 2 Probanden	Alle Probanden
unbearbeiteter Wortstammergänzungstest (gebräuchliche Methode)			
Rohwert (Anteil alter Wörter richtig)	.57 (.01)	.49 (.01)	.53 (.01)
Basisrate (Anteil neuer Wörter richtig)	.35 (.01)	.41 (.01)	.38 (.01)
im Hinblick auf Reliabilitätssteigerung bearbeiteter Wortstammergänzungstest (gebräuchliche Methode)			
Rohwert (Anteil alter Wörter richtig)	.57 (.01)	.49 (.01)	.53 (.01)
Basisrate (Anteil neuer Wörter richtig)	.33 (.01)	.40 (.01)	.37 (.01)
unbearbeiteter Wortstammergänzungstest (alternative Methode)			
Rohwert (Anteil alter Wörter richtig)	.57 (.01)	.49 (.01)	.53 (.01)
Basisrate (Anteil neuer Wörter richtig)	.41 (.01)	.35 (.01)	.38 (.01)
im Hinblick auf Reliabilitätssteigerung bearbeiteter Wortstammergänzungstest (alternative Methode)			
Rohwert (Anteil alter Wörter richtig)	.57 (.01)	.49 (.01)	.53 (.01)
Basisrate (Anteil neuer Wörter richtig)	.40 (.01)	.33 (.01)	.37 (.01)

Tabelle B

Gruppenmittelwerte der Variablen, die den Primingwerten und den Leistungswerten der Tests aus Experiment 2 zugrunde liegen. Es sind Werte für beide Testformen und für beide Methoden der Leistungswertbestimmung aufgeführt (Die Werte in Klammern repräsentieren die Standardfehler der Mittelwerte).

	junge Probanden	ältere Probanden
impliziter Wortstammergänzungstest (gebräuchliche Methode)		
Rohwert (Anteil alter Wörter richtig)	.53 (.01)	.44 (.02)
Basisrate (Anteil neuer Wörter richtig)	.36 (.01)	.31 (.01)
explizite Form des Wortstammergänzungstests (gebräuchliche Methode)		
Rohwert (Anteil alter Wörter richtig)	.62 (.01)	.50 (.01)
Basisrate (Anteil neuer Wörter richtig)	.15 (.02)	.15 (.01)
impliziter Wortstammergänzungstest (alternative Methode)		
Rohwert (Anteil alter Wörter richtig)	.53 (.01)	.44 (.02)
Basisrate (Anteil neuer Wörter richtig)	.36 (.01)	.31 (.01)
explizite Form des Wortstammergänzungstests (alternative Methode)		
Rohwert (Anteil alter Wörter richtig)	.62 (.01)	.50 (.01)
Basisrate (Anteil neuer Wörter richtig)	.15 (.02)	.15 (.01)

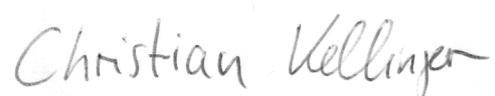
Tabelle C

Gruppenmittelwerte der Variablen, die den Primingwerten und den Sensitivitätswerten P_r der Tests aus Experiment 3 zugrunde liegen. Es sind Werte für beide Testformen und für beide Methoden der Leistungswertbestimmung aufgeführt (Die Werte in Klammern repräsentieren die Standardfehler der Mittelwerte).

	junge Probanden	ältere Probanden
impliziter Bildidentifikationstest (gebräuchliche Methode)		
Rohwert (Schritte bis Identifikation alter Bilder)	4.87 (0.06)	7.14 (0.12)
Basisrate (Schritte bis Identifikation neuer Bilder)	6.15 (0.08)	8.39 (0.13)
expliziter Bilderkennungstest (gebräuchliche Methode)		
Treffer (alte Bilder wiedererkannt)	.92 (.01)	.89 (.01)
Falsche Alarme (neue Bilder als alt eingestuft)	.03 (.01)	.04 (.01)
impliziter Bildidentifikationstest (alternative Methode)		
Rohwert (Schritte bis Identifikation alter Bilder)	4.87 (0.06)	7.14 (0.12)
Basisrate (Schritte bis Identifikation neuer Bilder)	6.15 (0.08)	8.39 (0.13)
expliziter Bilderkennungstest (alternative Methode)		
Treffer (alte Bilder wiedererkannt)	.92 (.01)	.89 (.01)
Falsche Alarme (neue Bilder als alt eingestuft)	.03 (.01)	.04 (.01)

Ich versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist. Die Dissertation wurde in der vorgelegten oder in ähnlicher Form noch bei keiner anderen Institution eingereicht. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Düsseldorf, den 01.03.2014

A handwritten signature in black ink that reads "Christian Kellinger". The script is cursive and somewhat slanted to the right.

(Christian Kellinger)