

UNIVERSITÄT MANNHEIM

**Ein Vergleich von sechs multinomialen
Modellen zur Erklärung des
Abruferlebens**

Inauguraldissertation zur Erlangung des akademischen
Grades eines Doktors der Sozialwissenschaften der
Universität Mannheim

vorgelegt von

Dipl.-Psych. Lutz Cüpper

Dekan Prof. Dr. Josef Brüderl

Gutachter Prof. Dr. Edgar Erdfelder
Prof. Dr. Hans Irtel

Tag der Disputation 14. Dezember 2007

Zusammenfassung

In der Testphase von Remember-Know-(RK-)Rekognitionsuntersuchungen entscheiden Versuchsteilnehmer für jedes Item, das sie als alt eingeschätzt haben, ob sie sich bewusst an die zugehörige Enkodierepisode erinnern können ('Remember') oder von der Reizpräsentation lediglich wissen ('Know'). Leider muss für die Antworten in RK-Untersuchungen angenommen werden, dass sie Rateeffekte unterliegen. Um Gedächtnis- und Rateprozesse zu dekonfundieren, wurde einerseits die Remember-Know-Guess-(RKG-)Prozedur vorgeschlagen; andererseits wurden zu diesem Zweck mehrere mathematische Modelle entwickelt. In dieser Arbeit wurden sechs RK-Modelle so erweitert, dass sie RKG-Daten erklären können. In einer nachfolgenden Modellelektionsstudie, die auf 46 Untersuchungen aus 32 Publikationen beruht, wurden die Modelle hinsichtlich wichtiger Modellgütekriterien miteinander verglichen: (1) Die deskriptive Adäquatheit der Modelle wurde anhand von herkömmlichen Likelihood-Quotienten-Verfahren und dem Vergleich von *AIC*-Differenzen bestimmt, (2) ein Maß für den Vorhersagefehler wurde durch ein nonparametrisches Bootstrapverfahren ermittelt, um zu untersuchen, inwiefern die deskriptive Adäquatheit der Modelle mit einer verringerten Generalisierbarkeit einherging, zudem wurde (3) die Konstruktvalidität der Modelle dadurch ermittelt, dass beobachtet wurde, in welchem Ausmaß Effekte von Manipulationen in jenen Parameterklassen wiedergefunden werden konnten, für die sie zu erwarten waren.

Die Ergebnisse dieser Arbeit liefern Evidenz für einige Zwei-Prozess-Modelle des Abruferslebens: Das SAC-Modell von Reder sowie das Four-States-Modell von Erdfelder zeigten eine bessere Leistung als das Ein-Prozess-Modell von Donaldson bzw. Hirshman, zwei Varianten des Zwei-Prozess-Signalentdeckungmodells von Yonelinas sowie das STREAK-Modell von Rotello. Analysen empirisch beobachteter ROC-Kurven unterstützen diese Schlussfolgerung.

Abstract

In the test phase of a typical Remember-Know (RK) recognition experiment, participants decide for each item they judged old whether they consciously 'remember' the event of encoding it or whether they just 'know' that the item was present in the learning phase. A straightforward statistical analysis of data obtained by this procedure is error-prone due to guessing processes that might influence the number of remember and especially know responses in absence of actual memory. In order to disentangle guessing from remembering and knowing, the Remember-Know-Guess (RKG) procedure has been suggested by Gardiner et al. (1996). However, mathematical models provide a better solution to the problem of separating task relevant memory processes from mere guessing.

In this work, six RK models were extended to account for data obtained by the RKG procedure. In a subsequent model selection study, 46 experiments from 32 different articles that had been published between 1996 and 2005 served as database. On that basis, the models' predictive ability was analyzed by means of techniques that allow to rank the models according to four model-selection criteria: (1) Descriptive adequacy was assessed by standard likelihood-ratio testing and comparisons of *AIC* differences, (2) a measure of prediction error was obtained by a nonparametric bootstrap technique to further explore, whether the models' descriptive adequacy was due to a lack of generalizability, and (3) construct validity was measured by testing whether effects of experimental manipulations are observed only in the parameter classes they are hypothesized to appear in.

The results of this work provide evidence for dual-process models of retrieval experiences. Reder's SAC model and Erdfelder's Four-States model did perform better than Donaldson's single-process signal-detection model, two versions of the Yonelinas' dual-process model, and Rotello's STREAK model. The analysis of empirically observed ROCs lend further support to this conclusions.

Danksagung

An dieser Stelle möchte ich all jenen Personen danken, die auf die eine oder andere Weise zum Gelingen dieser Arbeit beigetragen haben. Zuvorderst möchte ich Professor Dr. Edgar Erdfelder für seine Unterstützung und für die Freiheit danken, die er mir bei der Themenwahl und der Fertigstellung der Arbeit gewährte. Durch sein Engagement als Ansprechpartner und Diskutant übte er einen unverkennbaren Einfluss auf mein Arbeiten aus. Gleichmaßen dankbar bin ich für den Kontakt mit und die Anregungen von Professor Dr. Hans Irtel, der die Entstehungsgeschichte dieser Arbeit im Kolloquium für Kognitionspsychologie mitverfolgt und mitgestaltet hat.

Am Lehrstuhl für Psychologie III der Universität Mannheim fand ich während der Verfertigung dieser Arbeit eine akademische Heimat, hilfreiche Kolleginnen und Kollegen sowie eine sehr befruchtende Arbeitsatmosphäre. An dieser Stelle möchte ich Dr. Martin Brandt danken, der mir längst nicht nur fachlich zur Seite stand. Auch bei Professor Dr. Rüdiger Pohl und Dipl.-Psych. Tina-Sarah Auer fand ich Rat und Unterstützung.

Zudem bin ich für die praktische Hilfe dankbar, die ich durch die Hilfskräfte des Lehrstuhls erfahren habe. Insbesondere André Aßfalg hat in allen Stadien der Arbeit von der Literaturrecherche, über die Datenauslese bis hin zur Drucklegung durch seine tatkräftige Unterstützung einen wichtigen Beitrag zum Gelingen des Projektes geleistet.

Dank gebührt selbstverständlich auch Freunden, die mich während des Arbeitsprozesses unterstützt haben, und meiner Familie. An erster Stelle seien hier meine Eltern Ulrike und Rolf Cüpper erwähnt, ohne deren Zutun diese Arbeit sicherlich nicht geschrieben worden wäre. Weiterhin möchte ich Dipl.-Psych. Monika Undorf danken, die mit mir gelitten hat und sich mit mir freuen konnte. Glücklicherweise wurde sie dessen nie überdrüssig. Für die eindrucksvollen Momente, die wir an verlassenem Orten erleben konnten, danke ich Guido Hammes und seiner Frau Sandra.

Zu guter Letzt bitte ich um das Verständnis derjenigen, die ich nicht namentlich erwähnt habe, obwohl sie eine Erwähnung verdient hätten. Ich hoffe, dass sie sich durch die Lektüre meiner Arbeit entschädigen lassen.

Inhaltsverzeichnis

| | |
|---|------------|
| Zusammenfassung | iii |
| Abstract | v |
| Danksagung | vii |
| 1 Einleitung | 1 |
| I Theoretischer Hintergrund | 5 |
| 2 Gedächtnissysteme | 7 |
| 3 Episodisches Gedächtnis | 17 |
| 4 Messmodelle zur RKG-Prozedur | 33 |
| 4.1 SDM-Modell | 41 |
| 4.2 4-States-Modell | 44 |
| 4.3 Streak-Modell | 47 |
| 4.4 SAC-Modell | 53 |
| 4.5 HTSD-Modell | 57 |
| 4.6 HTUVSD-Modell | 61 |
| 5 Methoden der Modellselektion | 65 |
| 5.1 Modellgütekriterien | 65 |
| 5.2 Maximum-Likelihood-Parameterschätzung | 68 |
| 5.3 Bootstrapping und Modellselektion | 74 |
| II Empirische Modellvergleiche | 79 |
| 6 Datenbasis | 81 |

| | | |
|------------|---|------------|
| 7 | Plausibilität der Parameterwerte | 89 |
| 7.1 | Schätzungen für das SDM-Modell | 90 |
| 7.2 | Schätzungen für das HTSD-Modell | 92 |
| 7.3 | Schätzungen für das HTUVSD-Modell | 94 |
| 7.4 | Schätzungen für das 4-States-Modell | 96 |
| 7.5 | Schätzungen für das Streak-Modell | 97 |
| 7.6 | Schätzungen für das SAC-Modell | 99 |
| 7.7 | Diskussion | 100 |
| 8 | ROC-Analyse | 105 |
| 8.1 | ROC-Kurven für RKG-Untersuchungen | 106 |
| 8.2 | Modellvorhersagen | 109 |
| 8.3 | Empirische ROC-Kurven | 114 |
| 8.4 | Diskussion | 115 |
| 9 | Deskriptive Adäquatheit | 119 |
| 9.1 | Likelihood-Quotienten-Tests und AIC-Werte | 120 |
| 9.2 | Bootstrap-Modellselektion | 127 |
| 9.3 | Diskussion | 130 |
| 10 | Konstruktvalidität | 133 |
| 10.1 | Konstruktvaliditätsindex | 135 |
| 10.2 | Ergebnisse | 137 |
| 10.3 | Diskussion | 140 |
| 11 | Abschließende Diskussion | 143 |
| III | Anhang | 151 |
| A | Häufigkeitstabellen | 153 |
| B | ROC-Daten | 159 |
| C | Parameterschätzwerte | 163 |
| D | Datenanpassung: Hypothesenkonform | 189 |
| E | Datenanpassung: Hypothesendiskonform | 197 |
| F | Datenanpassung: Unrestringiert | 205 |
| | Literaturverzeichnis | 213 |

Tabellenverzeichnis

| | | |
|------|---|-----|
| 3.1 | Eigenschaften des episodischen Gedächtnisses | 19 |
| 3.2 | Ergebnisse einer Studie von Gregg und Gardiner (1994) | 25 |
| 3.3 | Steigungen von z -ROC-Geraden (nach Wixted & Stretch, 2004) | 32 |
| 5.1 | Modellselektionsindizes | 73 |
| 6.1 | Verwendete Studien | 83 |
| 6.2 | Klassifizierung experimenteller Manipulationen | 86 |
| 6.3 | Mittlere Antworthäufigkeiten für alte Reize und Distraktoren | 87 |
| 8.1 | Samplingschema der Simulationen | 111 |
| 8.2 | Ergebnisse der Simulation von ROC-Kurven | 112 |
| 8.3 | Empirische ROC-Parameter | 115 |
| 9.1 | Multiple Vergleiche der AIC -Rangpositionen. | 126 |
| 9.2 | Multiple Vergleiche nach Optimismus-Korrektur | 130 |
| 10.1 | Klassifizierung der Modellparameter | 134 |
| 10.2 | G^2 -Differenztests zur Analyse der Konstruktvalidität | 140 |

Abbildungsverzeichnis

| | | |
|------|---|-----|
| 2.1 | Langzeitgedächtnis | 9 |
| 2.2 | SPI-Modell nach Tulving (1995) | 13 |
| 3.1 | Zusammenhang zwischen A' -Werten und d' sowie C | 26 |
| 3.2 | Eindimensionales Kontinuum-Modell nach Donaldson (1996) | 28 |
| 4.1 | Antwortverhalten in RKG-Untersuchungen | 39 |
| 4.2 | SDM-Modell | 42 |
| 4.3 | 4-States-Modell | 46 |
| 4.4 | Streak-Modell | 50 |
| 4.5 | SAC-Modell | 55 |
| 4.6 | HTSD-Modell | 60 |
| 4.7 | HTUVSD-Modell | 62 |
| 6.1 | Anzahl der Reaktionen nach Itemtyp | 82 |
| 6.2 | Anteil der Abruferlebensarten an Alt-Antworten | 85 |
| 7.1 | Parameterschätzungen für das SDM-Modell | 91 |
| 7.2 | Parameterschätzungen für das HTSD-Modell | 93 |
| 7.3 | G^2 -Werte und extreme Parameterwerte (HTUVSD-Modell) | 94 |
| 7.4 | Parameterschätzungen für das HTUVSD-Modell | 95 |
| 7.5 | Parameterschätzungen für das 4-States-Modell | 97 |
| 7.6 | Parameterschätzungen für das Streak-Modell | 98 |
| 7.7 | Parameterschätzungen für das SAC-Modell | 100 |
| 8.1 | z-ROC-Geraden und deren Schnittwinkel φ | 108 |
| 9.1 | Relative Häufigkeit signifikanter G^2 -Differenztests | 121 |
| 9.2 | AIC -bestpassende Modelle | 122 |
| 9.3 | Bewährung im Sinne von AIC -Differenzen | 123 |
| 9.4 | Mittlere AIC -Ränge | 125 |
| 9.5 | Mittlerer Modelloptimismus | 128 |
| 9.6 | Mittlerer Rang nach Optimismus-Korrektur | 129 |
| 10.1 | Konstruktvalidität | 138 |

Kapitel 1

Einleitung

In der folgenden Arbeit wird ein wichtiger Bereich der Erforschung des menschlichen Gedächtnisses näher beleuchtet. Dabei handelt es sich um den Erlebensaspekt des Gedächtnisabrufes und die kognitiven Prozesse, aus denen das Erleben von Bewusstheit folgt. In der modernen Gedächtnisforschung hat das Abruferleben in der Debatte um Zwei-Prozess-Modelle des Rekognitionsgedächtnisses eine zentrale Rolle eingenommen (vgl. Roediger, Marsh & Lee, 2002; Yonelinas, 2002). Daher ist es nicht weiter verwunderlich, dass sich in der Literatur eine Vielzahl von Arbeiten finden lassen, die sich der Frage nach dem Erleben von Bewusstheit bei der Rekognition widmen (für einen Überblick siehe bspw. Dunn, 2004 oder Gardiner, 2002).

Diese Arbeiten lassen sich in drei Gruppen untergliedern: Die meisten dieser Forschungsarbeiten zeichnen sich dadurch aus, dass sie sich der Abfrage des Abruferlebens bedienen, um empirische Daten zu erzeugen, die helfen sollen, das Funktionieren des menschlichen Gedächtnisses näher zu verstehen. Diese erste Gruppe von Arbeiten bedient sich der Remember-Know-Prozedur als einer Technik, um psychologisches Wissen zu anderen Gegenstandsbereichen zu erweitern.

Ein weiterer Teil von Arbeiten widmet sich einer genaueren Klärung der theoretischen Basis dessen, was unter dem Abruferleben an sich zu verstehen ist. Eine solche Klärung lässt sich auf zweierlei Weise vornehmen: Zum einen können die theoretischen Grundlagen des Abruferlebens experimentell exploriert werden und zum anderen lässt sich die psychologische Theorie in mathematische Strukturen übersetzen, die helfen den jeweiligen Sachbereich zu erhellen. Dem letzteren Zugang folgend, wurden in neuerer Zeit sechs mathematische Modelle vorgeschlagen, die dazu dienen sollen, das Zustandekommen des Abruferlebens zu erklären.

Das Ziel dieser Arbeit besteht darin, jene sechs konkurrierenden Modelle metaanalytisch anhand von Daten, die publizierten Fachartikeln entnommen wurden, auf ihre empirische Adäquatheit hin zu untersuchen. Dabei werden hauptsächlich Modellselektionsverfahren eingesetzt, die auf Maximum-Likelihood(ML)-Schätzverfahren beruhen, wie sie gelegentlich in der psychologischen Forschung eingesetzt werden (vgl. Myung, 2000). Zusätzlich werden diese ML-basierten Verfahren durch Analysen von Receiver-Operating-Characteristics, sogenannten ROC-Kurven, ergänzt.

Die Arbeit gliedert sich in zwei Teile: Der erste dieser Teile umfasst insgesamt

vier Kapitel und befasst sich mit dem theoretischen Hintergrund des Modellvergleichs. Dabei wird in Kapitel 2 auf Systemtheorien des Gedächtnisses eingegangen, wie sie von Squire, Knowlton und Musen (1993) und Tulving (1995) entwickelt wurden. Systemtheorien haben einen wichtigen Einfluss auf die neuropsychologischen Theoriebildung, in der zwischen semantischem und episodischem Langzeitgedächtnis unterschieden wird.

Da die Erforschung des Abruferlebens in der Unterscheidung zwischen semantischem und episodischem Gedächtnis insofern ihren Anfang nahm, als das Erleben eines bewussten Abrufes dem episodischen Langzeitgedächtnis zugeschrieben wurde, während das Erleben von Vertrautheit als für das semantische Langzeitgedächtnis charakteristisch angesehen wurde (vgl. Tulving, 2002a), widmet sich das Kapitel 3 dem episodischen Gedächtnissystem. In diesem Kapitel wird die Remember-Know-Prozedur als eine Methode zur Erforschung des episodischen Gedächtnisses eingeführt. Hierbei wird besonderes Augenmerk auf Argumente gelegt, die die Richtigkeit einer Zwei-Komponenten-Interpretation der Ergebnisse von Remember-Know-Untersuchungen in Frage stellen.

In Kapitel 4 werden dann sechs mathematische Modelle vorgestellt, die dazu dienen, die Antworthäufigkeiten in Remember-Know- und Remember-Know-Guess-Untersuchungen in spezifische Bestandteile zu dekomponieren, denen in zugrunde liegenden kognitiven Theorien eine bestimmte Bedeutung zugeschrieben wird. Im Zuge dessen wird gezeigt, dass sich durch die theoriebasierte Modellbildung spezifische Probleme beseitigen lassen, die es bis dato unmöglich gemacht haben, zwischen einer Ein-Komponenten-Interpretation der Remember-Know-(Guess-)Prozedur und Zwei-Komponenten-Theorien zum selben Gegenstandsbereich zu unterscheiden (vgl. Dunn, 2004).

Eine Möglichkeit, die Adäquatheit der Modelle miteinander zu vergleichen und auf diese Weise die Korrektheit einzelner theoretischer Rahmenmodelle zu ermitteln, bieten jene Methoden der Modellselektion, die in Kapitel 5 vorgestellt werden sollen. Dabei handelt es sich überwiegend um Maximum-Likelihood-Verfahren. Diese Methoden sind insofern nützlich, als sie es erlauben, die wahrscheinliche Größe theoretischer Entitäten aus empirischen Daten zu schätzen und gleichzeitig ein Maß für die Wahrscheinlichkeit liefern, mit der Daten vor dem Hintergrund eines gegebenen Modells gültig sind. Verschiedentlich wird diese Wahrscheinlichkeit auch genutzt, um die Güte von Modellen abzuschätzen (vgl. Erdfelder, Cüpper & Auer, 2006; Myung, 2003).

Der zweite Teil der Arbeit widmet sich dem empirischen Vergleich der mathematischen Modelle anhand von Daten aus Remember-Know-Guess-Untersuchungen, die eine umfassende Literaturrecherche in den für dieses Forschungsfeld relevanten Literaturdatenbanken zutage gefördert hat. Die verwendete Datenbasis und ihre Eigenheiten werden in Kapitel 6 näher beleuchtet. Daran schließt sich in Kapitel 7 eine Erörterung der Plausibilität der Parameterschätzungen an, die die einzelnen Modelle für jene Untersuchungsbedingungen lieferten, die Teil der analysierten Datenbasis waren.

In Kapitel 8 werden die Daten der einzelnen Untersuchungsbedingungen dazu genutzt, um empirische ROC-Kurven mit Modellvorhersagen zu vergleichen. Dieses Ka-

pitel nimmt insofern eine Sonderstellung ein, als bei der Analyse der ROC-Kurven der statistische Rahmen der Maximum-Likelihood-Methoden verlassen wird. Die Ergebnisse dieser Analysen können daher als Test verstanden werden, inwiefern die Ergebnisse des Modellvergleichs unabhängig von den verwendeten statistischen Methoden Bestand haben.

In Kapitel 9 werden die Modelle hinsichtlich ihrer deskriptiven Adäquatheit miteinander verglichen. Die Fähigkeit von Modellen, empirische Daten zu beschreiben, stellt neben der in Kapitel 7 behandelten Plausibilität der Modellannahmen ein zweites wichtiges Modellgütekriterium dar. Da die deskriptive Adäquatheit von Modellen in einem Spannungsverhältnis zu einer weiteren wünschenswerten Eigenschaft brauchbarer Modelle steht, nämlich der Generalisierbarkeit von Modellvorhersagen (Pitt, Kim & Myung, 2003), ist dieses Kapitel in zwei Teile untergliedert, die sich in der Art der Korrektur für die Modellkomplexität unterscheiden. Die Modellkomplexität stellt den zentralen Einflussfaktor dar, der förderlich auf die deskriptive Adäquatheit wirkt und gleichzeitig abträglich für die Generalisierbarkeit von Modellvorhersagen ist. Während die üblichen statistischen Verfahren zum Vergleich von Modellen, wie sie auch im ersten Teil dieses Kapitels verwendet werden, die mathematische Komplexität vernachlässigen, wird im zweiten Teil dieses Kapitels auf die Auswirkungen eingegangen, die die Berücksichtigung selbiger auf die geschätzte deskriptive Adäquatheit der Modelle hat.

Ein weiteres wichtiges Gütekriterium stellt die Konstruktvalidität der Modelle dar. Dieses Thema bildet entsprechend das Zentrum der in Kapitel 10 zusammengefassten Ergebnisse.

In Kapitel 11 werden die wesentlichen Ergebnisse des empirischen Teils zum Abschluss noch einmal zusammengefasst und ihre Implikationen für die zukünftige Forschung in diesem Feld dargestellt.

Teil I

Theoretischer Hintergrund

Kapitel 2

Gedächtnissysteme

Es liegt nun über 100 Jahre zurück, dass Ebbinghaus (1885) mit seinem Werk *Über das Gedächtnis* den Grundstein für die moderne Gedächtnisforschung legte. Seit dieser Zeit hat sich der Gegenstandsbereich der von Ebbinghaus begründeten Disziplin vervielfältigt; das zunehmende Wissen über das Gedächtnis führte dazu, dass man sich von der ursprünglichen Vorstellung eines einheitlichen Konstrukts abwandte und dazu überging, das Gedächtnis in eine Vielzahl von Subsystemen zu untergliedern. Auch wenn sich diese Untergliederung schon sehr früh, z. B. in den *Principles of Psychology* von James (1890) ankündigte, so ist, wie Baddeley (2002) ausführte, als erste tatsächlich vorgenommene Unterteilung Hebbs (1949) Unterscheidung zwischen Kurzzeitgedächtnis und Langzeitgedächtnis anzusehen.

Erstmalig fanden Brown (1958) und Peterson und Peterson (1959) empirische Evidenz für diese Unterscheidung. Sowohl Brown (1958) als auch Peterson und Peterson (1959) beschäftigten sich mit dem schnellen Vergessen, das beobachtet werden kann, wenn beim Behalten kleiner Informationsmengen das Memorieren der Informationen verhindert wird. Sie erklärten das in dieser Aufgabe zu beobachtende Vergessen durch den Zerfall einer kurzfristigen Gedächtnisspur und unterschieden davon das Vergessen aus dem Langzeitgedächtnis, das auf Interferenz beruhen sollte. Ob dem kurzfristigen Vergessen tatsächlich Spurenerfall zugrunde liegt, wie es diese Autoren postulieren, oder Interferenz, wie es andere Forscher behaupten, ist bis heute unklar (vgl. Cowan, Nugent, Elliott & Geer, 2000; Service, 1998). Allerdings kamen während der 1960er Jahre experimentelle und neuropsychologische Belege hinzu, die für eine Trennung zwischen Kurzzeit- und Langzeitgedächtnis sprechen.

Neuropsychologisch sind für die Unterteilung des Gedächtnisses in verschiedene Subkomponenten Beobachtungen, die an amnestischen Patientinnen und Patienten vorgenommen worden sind, von großer Relevanz. Amnestiker zeichnen sich im Regelfall dadurch aus, dass sie ein Gedächtnisdefizit aufweisen, obwohl andere kognitive Leistungsparameter wie die allgemeine Intelligenz unauffällig sind (Baddeley & Warrington, 1970). Betrachtet man das amnestische Syndrom genauer, so lassen sich die meisten Amnesien in zwei Bestandteile untergliedern, die mit den Begriffen der anterograden Amnesie einerseits und der retrograden Amnesie andererseits belegt werden. Unter anterograder Amnesie versteht man ein beobachtbares Defizit bei der Speiche-

rung neuer Informationen. Sie lässt sich dementsprechend umschreiben als ein Lerndefizit, wobei sich dieses Defizit nicht in allen Bereichen gleichermaßen zeigt; so ist es dem betroffenen Patienten in aller Regel möglich, prozedural oder auch implizit zu lernen (z. B. O'Connor, Verfaellie & Cermak, 1995). Die retrograde Amnesie hingegen stellt sich als ein Abrufproblem dar, wobei sich dieses darin offenbart, dass es den Betroffenen unmöglich ist, sich an spezifische Ereignisse zu erinnern, die sie vor dem Eintreten ihrer Schädigung erlebt haben, oder auf Wissen zurückzugreifen, das sie vor dem schädigenden Ereignis aufgenommen haben.

Für die Trennung zwischen Kurzzeit- und Langzeitgedächtnis wesentlich ist die bereits erwähnte Beobachtung, dass Patientinnen und Patienten, die unter dem klassischen amnestischen Syndrom leiden, zwar in Langzeitgedächtnisaufgaben Leistungseinbußen hinzunehmen haben, häufig aber ein vollkommen intaktes Kurzzeitgedächtnis aufweisen (Baddeley & Warrington, 1970). Betrachtet man diesen Befund vor dem Hintergrund der komplementären Beobachtung von Patienten, die ein intaktes Langzeitgedächtnis aufweisen, während ihr Kurzzeitgedächtnis defizitär ausgeprägt ist (Shallice & Warrington, 1970), so liegt der Schluss nahe, dass Kurzzeit- und Langzeitgedächtnis zwei eigenständige Entitäten im Gefüge des menschlichen Geist-Gehirn-Systems darstellen.

Entsprechend wurden in diesem Zeitraum eine Reihe von Zwei-Komponenten-Modellen publiziert, die auf der Annahme zweier getrennter Speicher, einem Kurzzeitspeicher und einem Langzeitspeicher, basieren. Der wohl bekannteste Vertreter aus dieser Reihe ist das von Atkinson und Shiffrin (1968) vorgeschlagene Gedächtnismodell, das wegen der breiten Zustimmung, auf die es stieß, auch als modales Gedächtnismodell bekannt wurde. Diesem Modell zufolge werden Informationen, die aus der Umwelt auf das Individuum treffen, von einem Kurzzeitspeicher aufgenommen und von dort in den Langzeitspeicher überführt werden. Die Lernwahrscheinlichkeit, also die Wahrscheinlichkeit, mit der eine Information in den Langzeitspeicher aufgenommen wird, soll von der Verweildauer der Information im Kurzzeitspeicher abhängen. Gleichzeitig fungiert der Kurzzeitspeicher als Arbeitsgedächtnis, das während der Ausführung komplexer kognitiver Tätigkeiten, wie etwa dem Verstehen oder Denken, die dazu notwendigen Informationen bereithält.

Obwohl das Atkinson-Shiffrin-Modell mit den damals verfügbaren Daten übereinstimmte und bis heute in vielen Lehrbüchern der kognitiven Psychologie gefunden werden kann, bekam es recht schnell substantielle Probleme. So konnten Baddeley und Hitch (1974) in Untersuchungen, die einem Vorgehen entsprechen, das später als Zweitaufgaben-Paradigma bekannt geworden ist, zeigen, dass die Annahme eines einheitlichen Kurzzeitspeichers eine unzulässige Vereinfachung ist, und setzten an seine Stelle ein Arbeitsgedächtnis, das sich aus mehreren Komponenten zusammensetzt. Mit der Annahme zweier Sklavensysteme, der phonologischen Schleife einerseits und dem visuell-räumlichen Notizblock andererseits, die Baddeley und Hitch (1974) trafen, lässt sich zum Beispiel erklären, weshalb eine Kurzzeitspeicherbelastung wie das Merken einer Ziffernfolge (z. B. 4638) keine Auswirkungen auf die Leistung in einer visuellen Entscheidungsaufgabe hatte (vgl. Baddeley, 1986). Konfrontiert mit Buchstabenkombinationen (z. B. AB) sollte in letzterer Aufgabe entschieden werden, ob

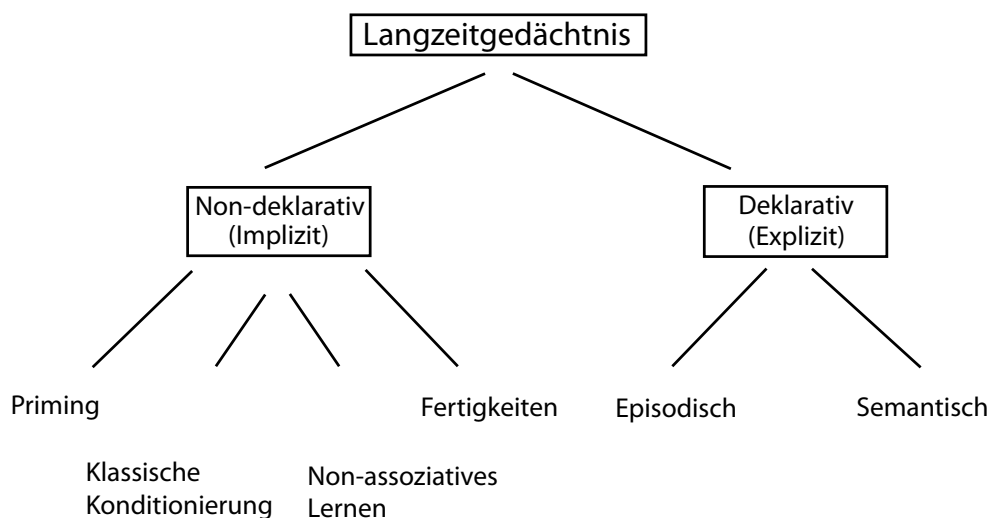


Abbildung 2.1: Dargestellt wird die hierarchische Unterteilung des Langzeitgedächtnisses in zwei Funktionsbereiche, die sich wiederum in verschiedene Unterkategorien gliedern lassen (Squire, 1992).

Aussagen des Typs 'A steht vor B' richtig oder falsch sind. Wegen der verbalen Kodierbarkeit der Ziffernfolge soll diese im Arbeitsgedächtnis sensu Baddeley (1986) mit Hilfe der phonologischen Schleife für einen späteren Abruf bereitgehalten werden können, während die Entscheidungsaufgabe aufgrund ihres räumlichen Charakters den visuell-räumlichen Notizblock benötigt, nicht aber auf die phonologische Schleife angewiesen ist.

Auch im Bereich des Langzeitgedächtnisses zeichnete sich eine Aufteilung in Subkomponenten ab. So schlugen Tulving und Donaldson (1972) vor, theoretisch zwischen semantischem und episodischem Gedächtnis zu unterscheiden. Damit reagierten sie auf die Arbeiten von Landauer und Freedman (1968) sowie Collins und Quillian (1969), die sich eingehend mit dem Format auseinandersetzten, in dem Gedächtnisinhalte abgelegt werden. Im Vordergrund dieser Untersuchungen stand die Speicherung von Fakten, so dass sich diese Arbeiten laut Tulving und Donaldson (1972) mit dem semantischen Gedächtnis und der Frage, wie semantische Informationen gespeichert werden, beschäftigten, während die klassische Gedächtnispsychologie der Frage nach der Funktionsweise des episodischen Gedächtnisses nachginge. Diese Unterscheidung von episodischem und semantischem Gedächtnis lässt sich auch in Abbildung 2.1 erkennen, die die Einteilung des Langzeitgedächtnisses in ineinander geschachtelte Komponenten wiedergibt, wie sie von Squire (1992) vertreten wurde (s. a. Squire et al., 1993).

In jüngerer Zeit schlug Squire (1992) vor, das Langzeitgedächtnis auf oberster Ebene in zwei Komponenten einzuteilen: Er unterschied ein System, das deklarative Inhalte speichert, von einem zweiten, das auf prozeduralem, non-deklarativem Lernen beruht. Evidenz für eine entsprechende Unterteilung ließ sich schon in den frühen 1980er Jahren ausmachen (z. B. Cohen & Squire, 1980; Graf, Squire & Mandler, 1984;

Jacoby & Witherspoon, 1982; Malamut, Saunders & Mishkin, 1984; Tulving, Schacter & Stark, 1982; Warrington & Weiskrantz, 1982). Befunde an gesunden erwachsenen Menschen, amnestischen Patienten und am Tiermodell sprachen dafür, dass eine Trennung vorgenommen werden muss zwischen dem bewussten Gedächtnis für Fakten und Ereignisse und einer Reihe von Funktionen, die unbewusstes Gedächtnis ermöglichen, wie z. B. den Erwerb von Fertigkeiten, das Trainieren von Gewohnheiten, die klassische Konditionierung sowie das Phänomen des Primings.

Knowlton, Ramus und Squire (1992) konnten beispielsweise zeigen, dass sich Leistungsdifferenzen zwischen Amnestikern und gesunden Kontrollprobanden in Aufgaben, die dem Paradigma des impliziten Grammatiklernen folgten (vgl. Reber, 1967), nur in Leistungsmaßen zeigten, die das deklarative Gedächtnis beanspruchen – in diesem Falle der Akkuratheit, mit der die Probanden bereits gesehene Instanzen als 'Alt' klassifizieren konnten. Bei Maßen für implizites Lernen, hier der Klassifikation nicht gelernter Reize als im Sinne der implizit gelernten Grammatik korrekt bzw. inkorrekt, zeigte sich hingegen kein bedeutsamer Leistungsvorteil für die gesunden Kontrollprobanden. Ebenso konnten Cave und Squire (1992) in einer Reihe von Primingexperimenten keine Leistungsunterschiede zwischen einer Gruppe von Amnestikern und einer Gruppe gesunder Kontrollprobanden finden. Bestand beispielsweise die Aufgabe der Probanden darin, Bilder zu benennen, so zeigte sich, dass die Benennungen neuer Bilder im Vergleich zu Bildern, die die Probanden vor zwei bzw. sieben Tagen schon einmal bearbeitet hatten, deutlich langsamer erfolgten. Dieses sogenannte Wiederholungspriming konnte sowohl bei der gesunden als auch bei der amnestischen Stichprobe nachgewiesen werden.

In beiden geschilderten Untersuchungen wurde die den Amnestikern verbleibende Gedächtnisfunktion dem unbewussten non-deklarativen Gedächtnis zugeschrieben, während davon ausgegangen wurde, dass die amnestischen Probanden deutliche Leistungseinbußen auf der Seite des deklarativen Gedächtnisses hinnehmen mussten.

Fraglich ist allerdings, welche Komponenten des deklarativen Gedächtnisses in welchem Ausmaß von einer Amnesie betroffen sind. Squire (1992) wies darauf hin, dass sich das Gedächtnis für deklarative Inhalte in zwei weitere Bestandteile untergliedern lässt, die sich darin unterscheiden, welche Art von Informationen sie bearbeiten. Bei diesen beiden von Squire unterschiedenen Informationstypen handelte es sich zum einen um Ereignisse bzw. Episoden und zum anderen um Fakten und Weltwissen. In dieser Unterscheidung lässt sich unschwer die von Tulving und Donaldson (1972) vorgeschlagene Differenzierung zwischen episodischem und semantischem Gedächtnis erkennen.

Für die Unterteilung des deklarativen Gedächtnisses in einen episodischen und einen semantischen Teil sprechen Untersuchungen an K.C., einem Patienten mit amnestischer Symptomatik. K.C. entwickelte aufgrund von Kopfverletzungen, die er bei einem Verkehrsunfall davongetragen hatte, ein amnestisches Syndrom, in dessen Folge er sich, wie Tulving (1991) ausführte, weder an Episoden aus seinem Leben vor dem Unfall erinnern konnte, noch in der Lage war, sich neue Episoden einzuprägen. Trotz seiner generellen Unfähigkeit zur episodischen Erinnerung, die sich in einer ausgeprägten retrograden wie auch anterograden Amnesie im Bereich seines episodischen

Gedächtnisses äußerte, ergab sich bezogen auf das Faktenwissen, über das K.C. noch verfügte, ein anderes Bild. Das vor seinem Unfall erworbene Wissen blieb dem Patienten dem Bericht von Tulving (1991) zufolge unbeschadet erhalten. In diesem Bereich ließen sich weitaus weniger Anzeichen einer retrograden Amnesie feststellen. Zudem konnten Tulving, Hayman und Macdonald (1991) zeigen, dass K.C. in der Lage war, neues Wissen zu erwerben. Hierzu ließen sie ihn in einer sich über Monate erstreckenden Untersuchung zufällig zusammengestellte Drei-Wort-Sätze wie z. B. den Satz *MEDICINE cured HICCUP* lernen. Bei einem Test, der darauf beruhte, dass der Patient vom Versuchsleiter vorgegebene Satzfragmente, im Beispiel *MEDICINE cured _____*, um das fehlende Objekt ergänzen sollte, offenbarte K.C. seine Fähigkeit, sich neue semantische Informationen einzuprägen.

In dieser, wenn auch rudimentären, Fähigkeit zu semantischem Lernen unterschied sich K.C. deutlich von dem allgemeinen Bild eines Amnestikers, das durch den Bericht von Scoville und Milner (1957) über einige Patienten geprägt ist, die nach beidseitiger Läsion des medialen Temporallappens einen ausgeprägten Gedächtnisverlust entwickelten. Der wohl berühmteste Fall aus diesem Bericht ist der Patient H.M., dem aufgrund seiner schweren Epilepsie im Jahre 1953 beidseitig große Teile des medialen Temporallappens, darunter der Hippocampus entfernt worden waren. Nach der Operation war der Patient nicht mehr in der Lage, das Klinikpersonal zu erkennen oder den Weg zu seinem Badezimmer zu finden. Wie Scoville und Milner berichteten, verbesserte sich H.M.'s Situation bis zur Vorlage ihres Berichts nicht. Für sein fortwährendes Unvermögen, sich neues Wissen anzueignen, lässt sich eine Vielzahl von Beobachtungen anführen. So hatte sich der Patient auch zehn Monate nach dem Umzug die neue Adresse seines Elternhauses nicht einprägen können. Wie seine Mutter berichtete, las H.M. Zeitschriftenartikel wieder und wieder ohne auch nur ein Anzeichen von Vertrautheit oder Wiedererkennen zu zeigen. Scoville und Milner (1957) berichteten zudem von gemeinsamen Mittagessen mit H.M., an die sich dieser schon eine halbe Stunde später nicht mehr erinnern konnte.

Anders als bei K.C. scheint bei H.M. sowohl das episodische als auch das semantische Gedächtnis von einer anterograden Amnesie betroffen gewesen zu sein. Verglichen mit der Symptomatik von H.M. wies die Symptomatik von K.C. damit zwei charakteristische Besonderheiten auf. Diese Besonderheiten sind zum einen die oben geschilderte Fähigkeit zu semantischem Lernen und zum anderen eine deutlich asymmetrische retrograde Amnesie. Diese Asymmetrie bestand darin, dass K.C. sich im Alltag zwar des vor seinem Unfall erworbenen Wissens bedienen konnte, ihm allerdings seine autobiografischen Erinnerungen aus demselben Zeitraum unzugänglich waren. Während also H.M. in Folge seiner Operation einen völligen Verlust seiner deklarativen Lernfähigkeit hinnehmen musste, scheint das in dieser Totalität bei K.C. nur für den Bereich der episodischen Erinnerungen zuzutreffen.

Wenngleich die Leistungsunterschiede zwischen semantischem und episodischem Gedächtnis bei K.C. besonders ausgeprägt waren, so ist dies nicht der einzige dokumentierte Fall, bei dem eine solche Symptomatik vorliegt. Die geschilderte Asymmetrie zwischen episodischem und semantischem Gedächtnis kann in einer Reihe von Berichten leicht wieder gefunden werden (z. B. Calabrese et al., 1996; Cermak &

O'Connor, 1983; Della Sala, Laiacona, Spinnler & Trivelli, 1993; Kopelman, Stanhope & Kingsley, 1999; Levine et al., 1998; Markowitsch et al., 1993; Viskontas, McAndrews & Moscovitch, 2000).

Zentrale Fragen, die noch zu klären sind, lauten, wie die beiden deklarativen Gedächtnissysteme ineinander greifen und wie sie miteinander interagieren. Auf diese Fragen gaben Squire et al. (1993) sowie Tulving (1995) dezidierte Antworten, die sich, wie im Folgenden deutlich wird, über weite Strecken ähneln.

Squire et al. (1993) sprachen sich dafür aus, dass das episodische Gedächtnissystem auf semantischen Inhalten aufbaut und diese als basale Bausteine der Erinnerungen assoziativ in Episoden zusammenführt. Das episodische Gedächtnis wäre demnach auf das semantische Gedächtnis angewiesen, um funktionieren zu können, während letzteres unabhängig von der Funktionstüchtigkeit des episodischen Gedächtnisses wäre. Daraus folgt direkt, dass Menschen ein funktionsfähiges semantisches Gedächtnis aufweisen können, ohne in der Lage zu sein, Episoden aus ihrem Leben zu behalten, wie es bei dem Patienten K.C. der Fall war. Ebenso können sowohl semantisches als auch episodisches Gedächtnis beeinträchtigt sein, mitunter bis zur vollständigen Funktionsunfähigkeit, wie sie bei H.M. beobachtet werden konnte. Im Sinne der Theorie von Squire et al. (1993) nicht erlaubt ist hingegen die Beobachtung eines funktionstüchtigen episodischen Gedächtnisses, wenn gleichzeitig Leistungseinbußen im semantischen Bereich vorliegen.

Unterstützt werden diese Annahmen durch das Phänomen der Quellenamnesie. Darunter versteht man einen selektiven Verlust des Wissens um das 'Wann' und das 'Wo' eines Enkodierereignisses, während die Fähigkeit, sich an das 'Was' – also den Inhalt der aufgenommenen Information – zu erinnern, unverändert erhalten bleibt. Shimamura und Squire (1991) konnten zeigen, dass Amnestiker, die eine große Zahl an Quellenfehlern begehen, bei Multiple-Choice-Tests, in denen die Erinnerung an die Episodeninhalte abgefragt wurde, nicht schlechter abschnitten als Amnestiker, die keine Beeinträchtigung des Quellengedächtnisses aufwiesen. Hier schien also beiden Gruppen das 'Was' zur Verfügung zu stehen, während nur die eine Gruppe in der Lage war, die kritische Verknüpfung der einzelnen Informationen zu Episoden vorzunehmen.

Eine detaillierte Beschreibung des Zusammenspiels von semantischem und episodischem Gedächtnis und der Art und Weise, wie diese mit weiteren Gedächtniskomponenten in funktionaler Weise verbunden sind, lieferte auch Tulvings SPI-Modell (Tulving, 1995; 2002b; Tulving & Markowitsch, 1998).

In diesem Modell werden zunächst Gedächtnissysteme und Gedächtnisprozesse voneinander unterschieden. Bei den zentralen Gedächtnisprozessen handelt es sich um das Enkodieren, das Speichern und das Abrufen von Informationen. Alle drei Prozesse laufen in den zentralen Gedächtnissystemen ab. Diese bestehen nach Tulving (2002b) aus dem semantischen und dem episodischen Gedächtnissystem sowie dem perzeptuellen Repräsentationssystem (PRS). In Tulvings Modell lassen sich damit beide von Squire (1992) vorgeschlagenen Komponenten des deklarativen Gedächtnisses wiederfinden. Darüber hinaus sind mit dem perzeptuellen Repräsentationssystem im SPI-Modell auch non-deklarative Gedächtnisinhalte vertreten. Allerdings nimmt Tul-

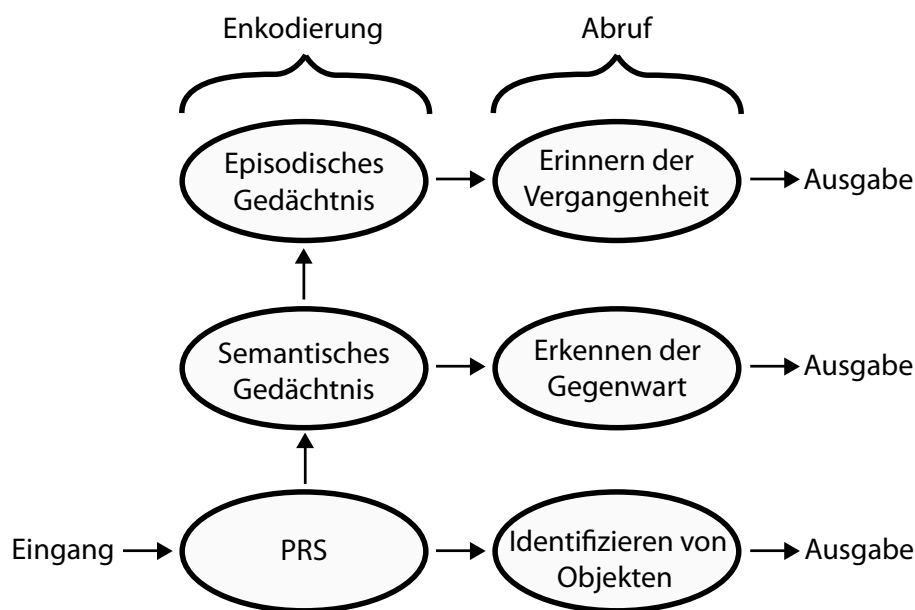


Abbildung 2.2: Dargestellt wird das SPI-Modell des Langzeitgedächtnisses nach Tulving (1995). Es werden drei Gedächtnissysteme mit spezifischen Aufgaben unterschieden. Pfeile geben die angenommene Richtung des Informationsflusses von der Enkodierung bis zum Abruf wieder.

ving (2002b) ausschließlich den perzeptuellen Anteil des non-deklarativen Gedächtnisses auf und vernachlässigt das an anderer Stelle ebenfalls postulierte prozedurale Gedächtnissystem, da dieses seiner Ansicht nach für das Verständnis des episodischen Gedächtnisses nicht von entscheidender Bedeutung ist. Das perzeptuelle Repräsentationssystem hingegen spielt als Vorstufe bei der Verarbeitung deklarativer Informationen eine entscheidende Rolle: So gelangen Tulving zufolge die perzeptuellen Eigenschaften physikalischer Objekte in das PRS. Dabei übernimmt das PRS zum einen die Aufgabe eines Speichers für diese perzeptuellen Eigenschaften und zum anderen Hilfsfunktionen für die übrigen Gedächtnissysteme, indem es bei der Enkodierung die Informationen für das nachfolgende System bereithält und zugänglich macht. Das perzeptuelle Repräsentationssystem für sich genommen ist u. a. verantwortlich für einfaches Stimulus-Response-Lernen, das auf der bloßen Assoziation eines physikalischen Auslösers und einer Verhaltensreaktion beruht. Darüber hinaus wurde angenommen, dass das PRS von entscheidender Bedeutung für das Phänomen des perzeptuellen Primings ist, das sich als eine erfahrungsbasierte Wahrnehmungserleichterung äußert, wie sie z. B. bei wiederholter Objektidentifikation beobachtet werden kann (Schacter, 1990; Tulving & Schacter, 1990).

Bis zu diesem Punkt unterscheidet sich die Skizze, die das SPI-Modell von der Struktur des menschlichen Langzeitgedächtnisses zeichnet, nicht deutlich von dem Bild, das Squire et al. (1993) davon entworfen. Eine theoretische Neuerung lässt sich hingegen in einer Reihe von Annahmen erkennen, die die Zusammenarbeit der drei

Gedächtnissysteme während des Enkodierens, Speicherns und Abrufens spezifizieren. Diese Annahmen über die Art des Informationsflusses bzw. die Interaktion der drei entscheidenden Gedächtnisprozesse sind so zentral, dass sie im Namen des Modells wieder gefunden werden können. Wie in Abbildung 2.2 dargestellt, sollen Informationen bei der Enkodierung seriell (*S*) verarbeitet werden, wobei die Verarbeitung im PRS beginnt. Die Ergebnisse dieser Stufe werden dann an das semantische System weitergereicht, bevor die semantischen Inhalte an das episodische System weitergeleitet werden. Während das *S* in dem Modellnamen SPI also für *seriell* steht und damit den Operationsmodus der Gedächtnissysteme bei der Enkodierung kennzeichnet, steht das *P* für *parallel* und kennzeichnet den Operationsmodus der Systeme bei der Speicherung. Parallelität während der Speicherung bedeutet, dass die Informationen in allen drei Systemen zeitgleich gespeichert werden, was dazu führt, dass einzelne Informationen auf den verschiedenen Ebenen mitunter mehrfach repräsentiert sind. Dabei gilt, dass in jedem einzelnen System spezifische Facetten der Reize gespeichert sind, d. h. im PRS die perzeptuellen Eigenschaften, im semantischen Gedächtnis die semantischen Anteile und im episodischen Gedächtnis die Assoziationen mit dem räumlichen und zeitlichen Kontext. Zuletzt steht das *I* für *unabhängig (independent)*; es besagt, dass die drei Systeme beim Abrufen unabhängig voneinander agieren können. Damit ist gemeint, dass Abrufergebnisse aus dem einen System den Abruf aus einem anderen System nicht beeinflussen müssen. Zwar werden die aus den unterschiedlichen Systemen abgerufenen Inhalte häufig wieder zusammengeführt, eine Notwendigkeit dazu besteht allerdings nicht. Beim Abruf ist also theoretisch jedes Gedächtnissystem für sich funktionsfähig und nicht auf die Funktionstüchtigkeit eines der anderen Systeme angewiesen.

Trotz seiner Einfachheit wartet das SPI-Modell mit interessanten Vorhersagen auf, die dem traditionellen Verständnis davon, was das Wesen der Kommunikation zwischen den Gedächtnissystemen ausmacht, zuwiderlaufen. Um der Leserin bzw. dem Leser ein besseres Verständnis für das SPI-Modell zu vermitteln, werden in der Folge einige dieser Hypothesen näher dargestellt und mit empirischen Argumenten untermauert.

Zuerst einmal lässt sich aus den Annahmen des Modells ableiten, dass ein Gedächtnissystem auf jeder Stufe funktionstüchtig sein kann, vorausgesetzt die jeweils niedrigeren Stufen sind intakt. Da dies unabhängig davon gilt, ob eine Schädigung auf einer höheren Stufe vorliegt oder diese Stufe in dem gegebenen Organismus gar nicht erst vorhanden ist, lassen sich Beobachtungen am gesunden Gedächtnis und am amnestischen Patienten mit diesem Modell genauso erklären wie Beobachtungen an Tieren und Säuglingen.

Häufig wird bei Tieren, aber auch bei Säuglingen, die sich noch keiner Sprache bedienen können, die perzeptuelle Rekognitionsleistung mit der sogenannten *delayed-non-matching-to-sample*-Aufgabe untersucht (Sharbaugh, Viet, Fraser & McMaster, 2003). Ebenso wie beim Einsatz ähnlicher Verfahren, die auf reiner Objektrekognition beruhen, zeigen die untersuchten Subjekte darin Gedächtnisleistungen, ohne auf das semantische Gedächtnis zurückgreifen zu müssen (Tulving, 2002a).

Neben Untersuchungen, die belegen, dass amnestische Patienten trotz ihres stark

eingeschränkten episodischen Gedächtnisses durchaus fähig sein können, neues Weltwissen zu erwerben, spricht die Beobachtung, dass junge Kinder bereits Wissen über die sie umgebende Welt erwerben können, bevor sie die Fähigkeit erlangen, sich an spezifische Ereignisse aus ihrer Vergangenheit zu erinnern (Nelson, 1993; Perner & Ruffman, 1995), für die Unabhängigkeit des Wissenserwerbs vom episodischen Gedächtnis.

Aufgrund der parallelen Speicherung und der Unabhängigkeit des Abrufes sagt das SPI-Modell anders als die Theorie von Squire et al. (1993) für retrograde Amnesien doppelte Dissoziationen zwischen semantischem und episodischem Gedächtnis vorher. Eine retrograde Amnesie kann sich also entweder in einer fast ausschließlichen Beeinträchtigung des episodischen bzw. des semantischen Gedächtnisses äußern, oder aber eine gleichmäßige Beeinträchtigung beider Systeme mit sich bringen, wie es in den meisten Fällen berichtet wurde. Für die anterograde Amnesie hingegen sind solche Dissoziationen ausgeschlossen. Da bei der anterograden Amnesie das Enkodieren beeinträchtigt ist, betrifft – aufgrund der Serialitätsannahme – jede Störung auf unterer Ebene die Leistung der höheren Systeme. Entsprechend kann bei der anterograden Amnesie nur eine einfache Dissoziation beobachtet werden: Entweder ist das episodische Gedächtnis geschwächt und das semantische Gedächtnis intakt oder sowohl das semantische als auch das episodische Gedächtnis weisen eine verminderte Leistungsfähigkeit auf. Der dritte Fall, d. h. ein intaktes episodisches Gedächtnis bei defizitärem semantischem Gedächtnis, ist mit den Modellannahmen dagegen nicht vereinbar.

Zuletzt lässt sich aus den Modellannahmen die Behauptung ableiten, dass die Gedächtnisleistung, wie sie in Rekognitionsuntersuchungen erfasst wird, nicht durch ein einheitliches Konstrukt wiedergegeben werden sollte, wie es in signalentdeckungstheoretischen Rekognitionstheorien der Fall ist, die die Gedächtnisleistung durch Maße wie d' (Banks, 1970), A' (Pollack & Norman, 1964) oder P_r (Snodgrass & Corwin, 1988) abbilden. Tulving (2002b) postulierte vielmehr, dass sich die Rekognitionsleistung aus der Leistung zweier Bestandteile zusammensetzt, nämlich zum einen der des episodischen und zum zweiten der des semantischen Gedächtnisses. Eine zentrale Eigenschaft des SPI-Modells kann also darin gesehen werden, dass es zwei inhaltlich unterschiedliche Komponenten benutzt, um die Ergebnisse von Rekognitionsuntersuchungen zu erklären. Damit unterscheidet es sich deutlich von einer ganzen Reihe sparsamerer Ein-Komponenten-Modelle. Um die Beiträge von semantischem und episodischem Gedächtnis zur Rekognitionsleistung getrennt erfassen zu können, ist allerdings eine genauere Beschreibung dieser beiden Konzepte von Nöten. Der Schwerpunkt sollte dabei auf Unterschiede zwischen semantischem und episodischem Gedächtnis gelegt werden, da sich erst nach einer genauen Klärung der Unterschiede Methoden entwickeln und verstehen lassen, die dazu geeignet sind, die Beiträge der beiden Systeme getrennt voneinander zu erfassen.

Entsprechend wird im folgenden Kapitel das Konzept des episodischen Gedächtnisses näher beleuchtet, um es von dem semantischen Gedächtnis abgrenzen zu können. Zudem wird darin mit dem Remember-Know-Paradigma eine Methode vorgestellt, die verstanden werden kann als ein Weg zur Trennung zwischen Leistungsan-

teilen, die auf das episodische Gedächtnis zurückgeführt werden können, und solchen, die auf einem Abruf aus dem semantischen Gedächtnis fußen. Der Systemansatz, wie er bisher dargelegt wurde, ist allerdings nicht der einzige theoretische Rahmen, innerhalb dessen das Remember-Know-Paradigma fruchtbar eingesetzt bzw. verstanden werden kann. Eine alternative Theorie bedient sich eines Zwei-Prozess-Ansatzes zum Gedächtnis. Dieser Ansatz wird zum Abschluss des nächsten Kapitels erörtert werden.

Kapitel 3

Episodisches Gedächtnis

Im vorherigen Kapitel ist der Aufbau des menschlichen Gedächtnisses skizziert worden. Dabei wurde ein Schwerpunkt auf das Langzeitgedächtnis gelegt. Wie dargestellt wurde, lässt sich dieses in verschiedene Subkomponenten unterteilen, zu denen das semantische und das episodische Gedächtnis gezählt werden. Während das semantische Gedächtnis Faktenwissen speichert und zum Abruf bereithält, ermöglicht das episodische Gedächtnis ein autobiografisches Erinnern an Ereignisse der persönlichen Geschichte und Episoden des täglichen Lebens. Des Weiteren wurden zwei Rahmenmodelle vorgestellt, in die das semantische und das episodische Gedächtnis eingebettet werden können. Diese beiden Theorien waren zum einen das Gedächtnismodell von Squire (1992) und zum anderen das SPI-Modell von Tulving (1995). Beim Vergleich der beiden Rahmenmodelle ließ sich unschwer erkennen, dass diese sich weniger darin unterscheiden, aus welchen basalen Komponenten sich das Gedächtnis ihnen zufolge zusammensetzen soll; vielmehr unterschieden sich die beiden theoretischen Konzepte darin, wie sie das Zusammenarbeiten der Gedächtnissysteme beschrieben.

In dem nun folgenden Abschnitt wird das sogenannte Remember-Know-Paradigma im Mittelpunkt stehen. Das Remember-Know-Paradigma ist ein Forschungsparadigma, mit dessen Hilfe Wissenschaftler die Eigenheiten des semantischen und des episodischen Gedächtnisses experimentell näher zu beleuchten versuchen. Bevor aber diese Methode erläutert wird, wird erst einmal dessen theoretische Grundlage näher betrachtet werden, bei der es sich, wie bereits erwähnt, um die Eigenschaften des episodischen Gedächtnisses handelt. Dabei interessieren zum einen die Eigenschaften dieses Gedächtnissystems im Allgemeinen und zum anderen jene Eigenschaften des Systems, die es von den verwandten Systemen, insbesondere dem semantischen Gedächtnis, unterscheiden.

Tulving (2005) legte eine in diesem Sinne sehr hilfreiche Liste vor, die auf Arbeiten von Wheeler, Stuss und Tulving (1997), Tulving und Markowitsch (1998) sowie Griffiths, Dickinson und Clayton (1999) zurückgeht und die beobachtbaren Eigenschaften des semantischen sowie des episodischen Gedächtnisses enthält. Diese in Tabelle 3.1 wiedergegebene Liste besteht ihrerseits aus zwei Unterlisten; während in der oberen Hälfte solche Eigenschaften zu finden sind, die dem semantischen Gedächtnis zu Eigen sind, finden sich in der unteren Hälfte Eigenschaften, die genuin dem episodischen

Gedächtnis zugeschrieben werden können.

Nach Tulving (1993) lassen sich die entscheidenden Merkmale des episodischen Gedächtnisses in den folgenden drei Punkten ausmachen, die im unteren Teil von Tabelle 3.1 an prominenter Stelle wieder gefunden werden können: Erstens soll es dem Menschen ermöglichen, sich an erlebte Ereignisse als solche zu erinnern. Zweitens soll der Mensch nicht nur in der Lage sein, die zeitliche Organisation sonst gänzlich zusammenhangloser Ereignisse zu erinnern, sondern zudem mentale Zeitreisen vollziehen können. Drittens sollen sich episodisches und semantisches Gedächtnis fundamental in der Art des Bewusstheitserlebens unterscheiden, das den Informationsabruf begleitet. Das den episodischen Abruf begleitende Bewusstheitserleben nennt Tulving (2002a) *autonoetische* (selbstwissende) Bewusstheit und hebt es ab von *noetischer* (wissender) Bewusstheit, die sich einstellt, wenn eine Information ausschließlich semantisch abgerufen werden kann.

Eine Erinnerung, derer man sich in *autonoetischer* Form bewusst ist, zeichnet sich durch eine Beteiligung des Selbst aus, wie man sie erlebt, wenn man Ereignisse mental rekonstruiert, die man in der Vergangenheit erlebt hat. Dabei ist das Bewusstheitserleben beim Erinnern dem Gegenwartserleben einerseits so ähnlich, dass Menschen sich in der Vergangenheit Erlebtes vorstellen können, beinahe als würden sie die gleiche Situation erneut durchleben, und andererseits doch so deutlich von selbigem unterscheidbar, dass Menschen im Allgemeinen sehr gut zwischen einer Erinnerung an ein vergangenes Ereignis und dem Erleben der Gegenwart unterscheiden können. Im Unterschied dazu äußert sich das Erleben *noetischer* Bewusstheit, das Tulving zufolge den Abruf von Wissen aus dem semantischen Gedächtnis begleitet, in einem Eindruck von Vertrautheit oder Bekanntheit, ohne dass das Selbst in die Erinnerung eingebunden oder ein Wiedererleben verspürbar wäre. Zur Beschreibung dieser beiden Bewusstheitsformen führt Tulving (1983) spezifische Termini *technici* ein: Für die *autonoetische* Bewusstheit verwendet er den Begriff *Remember* respektive *Remembering*, während er die *noetische* Bewusstheit mit dem Begriff *Know* bzw. *Knowing* belegt.

Diese Fachbegriffe sind es auch, denen das von Tulving (1985b) vorgeschlagene *Remember-Know* (RK)-Paradigma seinen Namen verdankt. Im Rahmen dieses Paradigmas wird die Form des Bewusstheitserlebens durch eine an die Probanden gerichtete Nachfrage erfasst: Die Probanden sollen bei erinnerten Reizen äußern, ob sie sich an die Präsentation des infrage stehenden Reizes in der Lernphase bewusst erinnern (*Remember*, R) oder ob sie lediglich wissen, dass der Stimulus in der Lernphase vorgekommen ist, ohne sich bewusst an die Lernsituation zu erinnern (*Know*, K).

Meistens wird das *Remember-Know*-Paradigma zusammen mit *Ja/Nein*-Rekognitionstests respektive *Alt-Neu*-Rekognitionstests eingesetzt, wobei die Probanden nur im Falle von *Alt*- bzw. *Ja*-Antworten um ein entsprechendes Bewusstheitsurteil gebeten werden. Im weiteren Verlauf soll deshalb unter der klassischen ‘*Remember-Know*-Prozedur’ ein Gedächtnisexperiment verstanden werden, das die folgenden beiden Bedingungen erfüllt:

1. Die Nachfrage nach dem Erleben der Bewusstheit beim Abruf wurde eingebettet in einen Rekognitionstest;

Tabelle 3.1: *Aufgelistet werden Eigenschaften des episodischen Gedächtnisses nach Tulving (2005). Im oberen Teil finden sich Eigenschaften, die primär das semantische Gedächtnis charakterisieren, während der untere Teil solche Eigenschaften enthält, durch die sich das episodische Gedächtnis vom semantischen Gedächtnis unterscheiden lässt.*

Semantisches Gedächtnis

- Schlüsselfunktion: Registrieren, Speichern und Abrufen nutzbaren Wissens
- Multimodaler Input
- Schnelle Enkodieroperationen ermöglichen Single-Trial-Lernen
- Komplexer und hochstrukturierter Speicher
- Speicherinhalt ist gegenständlich
- Speicherinhalt ist in Propositionen überführbar
- Speicherinhalt hat prinzipiell einen Wahrheitswert
- Informationsverarbeitung ist ausgesprochen kontextsensitiv
- Möglichkeit zu flexiblem Abruf
- Symbolische Repräsentation gespeicherter Information
- System ist kognitiv – Inhalte können reflektiert werden
- Verhaltensäußerung ist optional und nicht obligatorisch
- Bewusste Beachtung der Zeit ist nicht zwingend erforderlich
- Erleben noetischer Bewusstheit
- System interagiert mit anderen neurokognitiven Systemen
- Abhängig von weit verzweigten kortikalen und subkortikalen neuronalen Systemen
- Bei einer großen Zahl von Tieren zu finden; hoch entwickelt bei Säugetieren und Vögeln

Episodisches Gedächtnis

- Schlüsselfunktion: Bewusstes Erinnern von Geschehnissen in subjektiver Zeit
 - Ermöglicht mentale Zeitreisen in Vergangenheit und Zukunft
 - Aktivität wird von auto-noetischer Bewusstheit begleitet
 - Abhängigkeit von erinnerndem Selbst
 - Entwickelte sich in phylogenetisch jüngerer Zeit
 - Ontogenetische Entwicklung ist später als die anderen Gedächtnissystemen
 - Anfälliger gegenüber Schädigungen
 - Operationen bedürfen besonderer neurokognitiver Strukturen
 - Operationen sind auf semantisches Gedächtnis angewiesen
 - Episodische Erinnerung setzt Wissen voraus, Wissen impliziert hingegen nicht Erinnerung
 - Vergleichsweise stark auf den präfrontalen Kortex angewiesen
 - Kommt möglicherweise nur beim Menschen vor
-

2. Die Abfrage von Rekognitionsurteil und Remember-Know-Urteil erfolgte in zwei Stufen: Zuerst wurde ein Rekognitionsurteil (Ja/Nein oder Alt/Neu) erbeten und anschließend wurde im Falle einer Bejahung ein Remember-Know-Urteil notwendig.

In den zwanzig Jahren seitdem Tulving (1985b) erstmals zeigte, dass Probanden zwischen den erläuterten Bewusstseinsformen unterscheiden und sie in Remember- und Know-Antworten wiedergeben können, ist die Remember-Know-Prozedur in einer Vielzahl von Untersuchungen eingesetzt worden, die sich sehr unterschiedlichen Fragestellungen widmeten. Aus theoretischer Sicht allerdings können drei Fragen in den Vordergrund gerückt werden, die viele Wissenschaftler beschäftigten. Viele der frühen Arbeiten mit der Remember-Know-Prozedur zielten darauf ab, Remember- und Know-Antworten gezielt durch experimentelle Manipulationen zu dissoziieren. Eine zweite Forschungslinie offenbart sich in Untersuchungen, die danach streben, im Remember-Know-Paradigma erhaltene Untersuchungsergebnisse mit alternativen Verarbeitungsmodellen in Einklang zu bringen. In diesem Bereich lassen sich zwei Klassen von alternativen Theorien ausmachen. Zum einen werden Zwei-Prozess-Theorien benutzt, um Remembering auf den einen Prozess und Knowing auf den anderen zurückzuführen. Daneben zeichnen sich Bemühungen ab, Ergebnisse von Remember-Know-Untersuchungen durch einen einheitlichen Prozess abzubilden. So motiviert sich dann auch eine dritte wichtige Forschungsfrage, die den Zusammenhang zwischen Remember- und Know-Antworten auf der einen Seite und der Konfidenz bei Rekognitionsurteilen auf der anderen Seite thematisierten.

Einen Überblick über eine Reihe von Untersuchungen, die zusammengenommen Dissoziationen zwischen Remember- und Know-Antworten aufzeigten, boten Gardiner und Java (1993). Alles in allem deuteten die in dieser Überblicksarbeit dargestellten Ergebnisse darauf hin, dass sich Remember- und Know-Antworten doppelt dissoziieren lassen: Einerseits konnten Manipulationen gefunden werden, die ausschließlich Remember-Antworten beeinflussten aber keinen Einfluss auf Know-Antworten hatten. Zu diesen Manipulationen zählten: die Tiefe der Verarbeitung (Gardiner, 1988; vgl. Graf & Mandler, 1984; Jacoby & Dallas, 1981), das Retentionsintervall (Gardiner, 1988; Gardiner & Java, 1991; vgl. Sloman, Hayman, Ohta, Law & al., 1988) sowie geteilte versus ungeteilte Aufmerksamkeit (Gardiner & Parkin, 1990; vgl. Parkin, Reid & Russo, 1990). Andererseits konnte gezeigt werden, dass eine andere Gruppe von Manipulationen einen Einfluss auf Know-Antworten, nicht aber auf Remember-Antworten hatten. Zu diesen Variablen gehörte u. a. das maskierte Wiederholungspriming. Das maskierte Wiederholungspriming ist eine Manipulation, die auf Jacoby und Whitehouse (1989) zurückgeht. Die Autoren konnten zeigen, dass sowohl die Rate der Treffer als auch die der Falschen Alarme in einfachen Rekognitionsuntersuchungen anstieg, wenn in der Abrufphase der Präsentation eines Testwortes eine kurze und maskierte Präsentation des gleichen Wortes vorausging. Genau diesen Effekt hat Rajaram (1993, Exp. 3) mit der Remember-Know-Prozedur näher untersucht. Dabei hat sie feststellen können, dass sich das maskierte Wiederholungspriming auf die Häufigkeit von Know-Antworten, nicht aber die von Remember-Antworten auswirkte.

Weiterhin konnten Variablen identifiziert werden, die eine Dissoziation zwischen

Remember- und Know-Antworten produzieren, bei der die beiden Antwortmöglichkeiten in entgegengesetzter Richtung beeinflusst wurden. Zu diesen Variablen gehörten der Vergleich zwischen Wörtern und Nonwörtern (Gardiner & Java, 1990), das Alter der Probanden (Parkin & Walter, 1992) und die Präsentation von Wörtern respektive Bildern in der Lernphase (Rajaram, 1993, Exp. 2).

Das Gesamtbild dieser Befunde, dass sich Remember- und Know-Antworten voneinander getrennt beeinflussen lassen, harmoniert gut mit Tulvings SPI-Modell im Allgemeinen und der in diesem Rahmen formulierten Unabhängigkeitsannahme im Besonderen. Entsprechend ist ein möglicher Schluss, der an dieser Stelle gezogen werden kann, dass die Häufigkeiten von Remember- und Know-Antworten von zumindest zwei unterscheidbaren und unabhängigen Entitäten abhängen, bei denen es sich dem SPI-Modell zufolge um das semantische und das episodische Gedächtnis handeln soll.

Allerdings ist dies nicht die einzige mögliche Sichtweise. Ein alternatives Rahmenmodell, das der Tradition einer prozesstheoretischen Betrachtung des Gedächtnisses folgt, stellte Rajaram (1993, 1996) vor. In ihren Arbeiten ging die Autorin anfänglich davon aus, dass konzeptuelle Verarbeitungsprozesse Remembering beeinflussten und perzeptuelle Prozesse für Knowing von entscheidender Bedeutung seien. Sie bemerkte aber bald, dass dieses theoretische Fundament sich empirisch nicht bestätigen ließ. Problematisch für diese Perzeptuell-Konzeptuell-Theorie sind Untersuchungen, in denen sich zeigte, dass Remembering auch auf perzeptuelle Manipulationen reagierte sowie dass Knowing durch eine Veränderung der konzeptuellen Verarbeitungsprozesse gesteigert werden konnte. In diesem Sinne kritische Befunde wurden u. a. von Rajaram und Geraci (2000) vorgelegt, die zeigen konnten, dass semantisches Priming in der Abrufphase – eine Manipulation, die das Konzept betrifft – selektiv die Häufigkeit von Know-Antworten beeinflusste. Den Einfluss einer perzeptuellen Manipulation auf Remember-Antworten belegte der von Rajaram (1996) berichtete Effekt der perzeptuellen Kongruenz zwischen Lern- und Abrufphase auf die Häufigkeiten von Remember-Antworten.

In ihren späteren Arbeiten nahm die Autorin (Rajaram, 1996; Rajaram & Geraci, 2000) daher an, dass das Bewusstheitserleben einerseits und die Art der Verarbeitung – eher perzeptuell oder eher konzeptuell – andererseits unabhängig voneinander sind. Für die Erklärung des Bewusstheitserlebens erachtete sie dementsprechend die Perzeptuell-Konzeptuell-Dichotomie als nebensächlich und rückte an deren Stelle eine neue Dichotomie, in der sie Prozesse dahingehend unterschied, inwiefern diese auf der einen Seite Distinktheit respektive Verarbeitungsflüssigkeit auf der anderen Seite fördern. Demnach sollte sich erhöhte Verarbeitungsflüssigkeit steigernd auf Know-Antworten auswirken, während sich eine vermehrte Beschäftigung mit distinkten Merkmalen von Reizen in einem Mehr an Remembering niederschlagen sollte.

Veranschaulichen lässt sich diese Dichotomie an zwei Experimenten, die Dehurst und Anderson (1999) durchgeführt haben. Während in dem ersten dieser beiden Remember-Know-Experimente die zu lernenden Wörter in der Lernphase einmal, viermal oder achtmal wiederholt wurden (exakte Wortwiederholung), wurden in der Lernphase des anderen Experiments ein, vier oder acht Wörter, die der gleichen semantischen Kategorie angehörten, dargeboten (kategoriale Wiederholung). In der Zu-

sammenschau dieser beiden Experimente zeigte sich, dass eine exakte Wortwiederholung die Häufigkeit von Remember-Antworten erhöhte. Dieses Phänomen trat allerdings nur solange auf, wie zwischen den Wiederholungen der Wörter andere Wörter präsentiert wurden. In der Untersuchung von Dewhurst und Anderson lagen bei manchen Wörtern insgesamt fünf Durchgänge vor der nächsten Wortwiederholung ('lag 6'). Eine massierte Wiederholung, bei der die Reize in der Lernphase in direkter Folge präsentiert wurden ('lag 0'), führte hingegen zu weniger Remember-Antworten. Anders verhielt es sich bei der kategorialen Wiederholung. Sie erhöhte die Anzahl der Know-Antworten signifikant, während die Remember-Antworten nicht beeinflusst wurden. Die Art der Wiederholung ('lag 0' vs. 'lag 6') übte keinen signifikanten Einfluss auf die Anzahl der Know-Antworten aus.

Der Anstieg der Remember-Antworten infolge der exakten Wortwiederholung lässt sich nach der Distinktheits-Flüssigkeits-Hypothese von (Rajaram, 1996) mit der Anlage verschiedener (distinkter) episodischer Gedächtnisspuren erklären. Auf die gleiche Weise kann auch der Umstand erklärt werden, dass die Remember-Antworten in der 'lag 6'-Bedingung deutlich häufiger waren als in der 'lag 0'-Bedingung: Bei einem größeren Abstand zwischen den einzelnen Ereignissen steigt die Wahrscheinlichkeit dafür, dass diese in voneinander gut unterscheidbare episodische Repräsentationen eingebunden werden. Dies kann z. B. an dem deutlich verschiedenen Kontext liegen, der dadurch zustande kommt, dass zwischenzeitlich eine Reihe anderer Wörter präsentiert worden sind. Die Zunahme der Know-Antworten infolge kategorialer Wiederholung lässt sich hingegen mit einer erhöhten konzeptuellen Flüssigkeit begründen, die aus der häufigen Begegnung mit Wörtern aus einer semantischen Kategorie resultiert.

Bei dem Vergleich des Distinktheits-Flüssigkeits-Ansatzes mit dem SPI-Modell überwiegt die Zahl der Gemeinsamkeiten die Zahl der Unterschiede deutlich; der Distinktheits-Flüssigkeits-Ansatz wurde dementsprechend als Ergänzung zur Theorie der Gedächtnissysteme betrachtet (Gardiner, 2000; Rajaram, 1999). Zwar unterscheiden die genannten Ansätze sich darin, womit sie Remembering und Knowing erklären, gemeinsam ist ihnen jedoch, dass sie erstens beide darauf vertrauen, dass sich die Eigenheiten des Abruferlebens in einem Forschungsansatz offenbaren, der auf den subjektiven Berichten der Versuchspersonen aufbaut, und zweitens stimmen die Ansätze darin überein, dass zwei konzeptuell voneinander zu unterscheidende Komponenten angenommen werden, um Remembering und Knowing zu erklären.

Keiner dieser beiden Punkte traf auf ungeteilte Zustimmung. Im Verlaufe der vorliegenden Arbeit werden Modelle vorgestellt werden, die bewusst nicht den subjektiven Bericht der Probanden an sich in den Mittelpunkt des Forschungsinteresses rücken, sondern durch geeignete mathematische Methoden versuchen, die geistigen Abläufe, die hinter diesen Berichten verborgen liegen, zu beschreiben. Vorerst soll jedoch ein nicht weniger bedeutsamer Kritikpunkt diskutiert werden, der auf der Prämisse fußt, dass die Annahme zweier qualitativ unterschiedlicher theoretischer Konstrukte zur Erklärung von Remembering und Knowing unnötig sei.

Erstmals wurde dieser Kritikpunkt von Donaldson (1996) vorgebracht und unter dem Namen *detection model* in eine Form gegossen. Eine grafische Darstellung der Struktur des Donaldson-Modells, das auch unter dem Titel *Eindimensionales*

Kontinuum-Modell firmiert (Erdfelder, 2000), kann Abbildung 3.2 entnommen werden. Es beruht auf der Annahme, dass sich Gedächtnisinhalte – in Gedächtnisexperimenten die zu erinnernden Items – lediglich auf einem eindimensionalen Kontinuum hinsichtlich des Vertrautheitsgrades oder der Spurstärke unterscheiden. Weiter nahm Donaldson (1996) an, dass Remembering und Knowing sich auf diesem einen Kontinuum ausschließlich quantitativ unterscheiden; die Frage, ob eine Remember-Antwort oder eine Know-Antwort gegeben wird, ist demnach eine Frage der Spurstärke des zu beurteilenden Reizes. Ist die Spurstärke höher als ein bestimmtes Kriterium (C_R), so resultiert eine Remember-Antwort, ist sie hingegen niedriger als das Remember-Kriterium (C_R) aber höher als ein zweites Kriterium (C_{RK}), das niedriger als C_R liegt, so erfolgt eine Know-Antwort. Nur wenn das Kriterium C_{RK} nicht überschritten wird, geben die Probanden ein Neu-Urteil ab. Darüber hinaus wird angenommen, dass die mittlere Spurstärke gelernter Items um einen bestimmten Betrag größer ist als die mittlere Spurstärke der Distraktoren. Distraktoren sind solche Reize, die im Kontext der Untersuchung zum Zeitpunkt der Rekognitionsabfrage zum ersten Mal präsentiert werden.

Das Ausmaß dieses erfahrungsinduzierten Spurstärkezuwachses erfasst man in aller Regel durch das Maß d' ; als Sensitivitätsmaß kann aber auch A' verwendet werden, das den Beinamen ‘nonparametrisches’ Sensitivitätsmaß trägt (Grier, 1971). Während das Maß d' auf der Annahme beruht, dass sich die Spurstärken für Distraktoren und gelernte Reize im Gauß’schen Sinne normalverteilen und damit einer parametrischen Verteilung folgen, kommt die theoretische Basis von A' ohne eine solche Verteilungsannahme aus. Allerdings wiesen Macmillan und Creelman (1996) zu Recht darauf hin, dass auch die Schätzung von A' nicht vollkommen auf Verteilungsannahmen verzichtet. Vielmehr ist A' ein hybrides Maß, in dem zwei Sensitivitätsmaße miteinander verschmelzen, wovon das eine auf der Annahme einer Rechteckverteilung beruht und für das andere eine logistische Verteilung von Nöten ist. Dennoch wird A' häufig dann eingesetzt, wenn die Normalverteilungsannahme oder eine äquivalente Verteilungsannahme für ungerechtfertigt gehalten wird. Da Donaldson (1996) anders als Hirshman und Master (1997), die ein verwandtes Modell vorgeschlagen haben, bewusst darauf verzichtete, sein Modell durch eine Annahme über die Verteilungsform der Spurstärken einzuschränken, bedarf es der Kenntnis des Maßes A' , um die Argumente für sein Modell verstehen zu können. Die Formeln zur Berechnung von A' und eines verwandten Maßes (B''_D), das Donaldson (1992) vorgeschlagen hat, um Urteilstendenzen zu erfassen, können in den Gleichungen 3.1 und 3.2 gefunden werden. Die Berechnung dieser beiden Indizes beruht auf der empirisch ermittelten Treffer-Rate (H) und der Falsche-Alarm-Rate (F).

$$A' = \frac{1}{2} + \frac{1 + H - F}{4H(1 - F)} \quad (3.1)$$

$$B''_D = \frac{(1 - H)(1 - F) - HF}{(1 - H)(1 - F) + HF} \quad (3.2)$$

Um die Plausibilität seiner Annahmen zu unterstreichen, überprüfte Donaldson (1996) in einer Metaanalyse von 80 Remember-Know-Datensätzen, die aus 28 Ex-

perimenten stammten und in 17 verschiedenen Arbeiten gefunden wurden, zwei Vorhersagen, die direkt aus den Prämissen seines Modells folgten:

1. Wegen der Unidimensionalitätsannahme sollten Sensitivitätsmaße nicht davon abhängen, ob sie aus den Gesamt-Treffer- und Gesamt-Falsche-Alarm-Daten geschätzt werden oder ob zu ihrer Schätzung ausschließlich die Remember-Antworten verwendet werden;
2. Dem Modell zufolge sollte der Gedächtnisanteil an Know-Antworten von der Strenge des Antwortkriteriums abhängen – je strenger das Antwortkriterium ist, desto mehr Know-Treffer sollten pro Know-Falschen-Alarm beobachtet werden.

In Donaldsons (1996) Metaanalyse erwies sich der auf den Remember-Antworten beruhende Sensitivitätsindex A'_R verglichen mit dem Index A'_{RK} , der auf der Basis aller Alt-Antworten bestimmt wurde, mit $A'_R = .83$ im Vergleich zu $A'_{RK} = .86$ als nur minimal kleiner. In dieser Beobachtung sah Donaldson (1996) seine erste Hypothese bestätigt. Allerdings sicherte er den Befund nicht inferenzstatistisch ab.

Eine vollständig andere Interpretation dieser Ergebnisse lieferten Gardiner und Gregg (1997). Die Autoren untersuchten den vorgestellten A' -Unterschied näher und stellten fest, dass er zwar gering ausfiel, jedoch durchgängig in vielen Untersuchungen zu finden war: In 60 der 72 Bedingungen, in denen sich der Sensitivitätsindex für Remember-Antworten von der Gesamt-Sensitivität unterschied, war A'_R kleiner als A'_{RK} . Gardiner und Gregg sicherten diese Beobachtung über einen Vorzeichentest ab und konnten so ermitteln, dass der Effekt einem z -Wert von 5.54 entsprach.

Eine Schlussfolgerung aus Donaldsons Metaanalyse könnte demnach lauten, dass die Gesamt-Sensitivität überzufällig häufig größer ist als die Sensitivität auf der Basis der Remember-Antworten. Dies wäre konform mit Vorhersagen, die sich aus einem Zwei-Komponenten-Ansatz ableiten ließen, nicht aber mit denen des eindimensionalen Kontinuum-Modells. Für eine solche Deutung sprachen nach Gardiner und Gregg (1997) Beobachtungen, die nahe legten, dass die Differenz zwischen A'_{RK} und A'_R dann besonders groß war, wenn Manipulationen eingesetzt wurden, die es den Probanden erschwerten, Episoden abzulegen. Als Beispiel führten die Autoren eine Manipulation an, die auf Gregg und Gardiner (1994, Exp. 2) zurückgeht. Die Probanden bearbeiteten in der Lernphase dieser Untersuchung eine Liste von einzelnen Wörtern, die sehr schnell aufeinander folgten: Jedes Wort erschien für 300 ms auf einem Bildschirm, das Zeitintervall zwischen den einzelnen Wörtern betrug 200 ms. Zudem wurde die Bedeutung visueller Verarbeitung während der Enkodieraufgabe überbetont, indem die Probanden dazu angehalten wurden, fehlerhafte Buchstaben in den Wörtern zu entdecken, um später die bloße Anzahl fehlerhafter Wörter zu berichten. Tatsächlich befand sich kein beschädigtes Wort in der Liste, so dass sich in dieser Hinsicht keines der Wörter von den anderen abhob. Im Anschluss an diese inzidentelle Lernphase sahen die Probanden die Wörter ein zweites Mal in demselben visuellen Format, jedoch mit dem Unterschied, dass sie diesmal Rekognitions- und Remember-Know-Urteile abgeben sollten. Die genauen Ergebnisse dieses Experiments sowie die von drei weiteren Experimenten, in denen die Ergebnisse von Gregg und Gardiner (1994, Exp. 2)

Tabelle 3.2: *Ergebnisse der Studie von Gregg und Gardiner (1994, Experiment 2; 1) und von unveröffentlichten Experimenten (2-4), die von Gardiner und Gregg (1997) berichtet wurden. Angeführt werden zusammen mit den zugehörigen A' -Werten die relativen Häufigkeiten, mit denen bei alten bzw. neuen Reizen mit Remember vs. Alt (gesamt) geantwortet wurde.*

| Antworten | Stimuli | | A' |
|--------------|---------|-----|------|
| | Alt | Neu | |
| (1) Remember | .11 | .03 | .66 |
| gesamt | .63 | .22 | .80 |
| (2) Remember | .12 | .04 | .67 |
| gesamt | .75 | .21 | .84 |
| (3) Remember | .14 | .04 | .68 |
| gesamt | .65 | .22 | .80 |
| (4) Remember | .06 | .04 | .60 |
| gesamt | .58 | .32 | .71 |

repliziert werden konnten, zeigt Tabelle 3.2. In diesen vier Experimenten war A'_{RK} um einen Betrag, der in der Größenordnung zwischen .11 und .17 lag, größer als A'_R , womit die Differenz zwischen den beiden Sensitivitätswerten um ein Vielfaches größer war als die von Donaldson beobachtete mittlere Differenz von .03.

Da die von Gardiner und Gregg (1997) angeführten Untersuchungen bewusst darauf abzielten, episodisches Erinnern möglichst unwahrscheinlich zu machen, passten diese Ergebnisse – genauso wie die Ergebnisse ihrer Reanalyse der von Donaldson (1996) vorgestellten Metaanalyse – gut zu den Vorhersagen eines Zwei-Komponenten-Ansatzes. Zudem scheinen diese Resultate auf den ersten Blick gegen das eindimensionale Kontinuum-Modell zu sprechen. Bei einer genaueren Betrachtung erweist sich diese Vermutung allerdings als falsch: Zwar sind systematische Unterschiede in den Sensitivitäten abhängig von der Datenbasis (Remember-Antworten versus Reaktionsurteile) tatsächlich problematisch, wenn man die Gültigkeit des Kontinuum-Modells annehmen möchte, dies ist allerdings nur solange der Fall, wie die empirisch beobachteten Unterschiede zweifelsfrei auf tatsächliche Sensitivitätsunterschiede zurückgeführt werden können.

Begründete Zweifel, wonach der in dem paarweisen Vergleich von A' -Werten beobachtete Sensitivitätsunterschied ein Artefakt ist, das auf die Wahl des Sensitivitätsmaßes A' zurückzuführen ist, äußerte Dunn (2004). Dunn legte für insgesamt fünf ähnlich geartete Versuche, das eindimensionale Kontinuum-Modell zu falsifizieren, dar, weshalb die jeweiligen Beobachtungen nicht dazu geeignet waren, das in Frage stehende Modell zu falsifizieren. In seinen den Sensitivitätsunterschied betreffenden Analysen arbeitete Dunn heraus, dass das Maß A' – wenn die Spurstärken von gelernten Items und Distraktoren Normalverteilungen folgen, die die gleiche Streuung

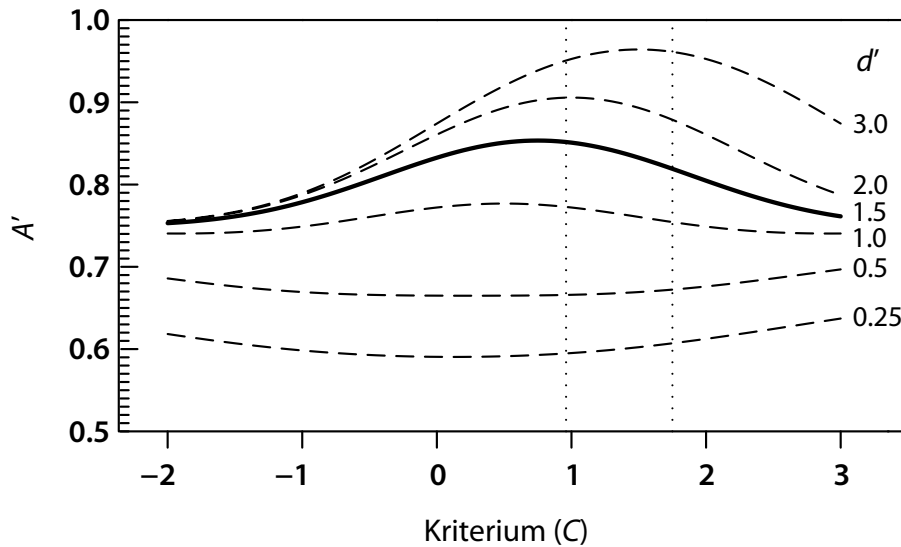


Abbildung 3.1: Dargestellt wird der Zusammenhang zwischen A' -Werten und d' sowie C nach Dunn (2004). Hervorgehoben wird die Kurve für $d' = 1.53$. Zudem finden sich gestrichelte Linien für die beiden Kriterien $C_{RK} = 0.96$ und $C_R = 1.75$

aufweisen und um d' zueinander verschoben sind, – nicht nur abhängig von der Sensitivität, in diesem Falle d' , ist, sondern zudem unerwünschterweise von dem angenommenen Entscheidungskriterium (c) abhängt. Wieder finden lassen sich diese Zusammenhänge in Abbildung 3.1, die ausschließlich horizontale Geraden enthalten würde, wenn A' auch unter den oben genannten Umständen ausschließlich die Sensitivität erfassen würde.

Um sein Argument weiter zu untermauern, bestimmte Dunn (2004) aus insgesamt 400 Remember-Know-Bedingungen die mittlere Sensitivität d' , den mittleren Kriteriumswert C_R , der sich aus der Anzahl der Remember-Antworten ergab, und den mittleren Wert des Kriteriums C_{RK} , das auf der mittleren Häufigkeit von Alt-Antworten basierte. Die entsprechenden Werte lagen bei $d' = 1.53$, $C_R = 1.75$ sowie $C_{RK} = 0.96$. Die durchgezogene Kurve in Abbildung 3.1 beruht auf $d' = 1.53$. An die Stelle der Kriteriumswerte für die Abgabe einer Rekognitionsentscheidung einerseits und für die einer Remember-Entscheidung andererseits sind in die Abbildung vertikale Linien übernommen worden. Wie man sieht, liegt der für $C_R = 1.75$ erwartete A' -Wert niedriger als der für $C_{RK} = 0.96$ erwartete. Demnach legt auch das eindimensionale Kontinuum-Modell – zumindest, wenn man von normalverteilten Spurstärken sowie von gleicher Varianz bei Distraktoren und gelernten Reizen ausgeht – nahe, dass für Werte, die im Bereich des Erwartungswertes für d' liegen, A'_{RK} größer sein sollte als A'_R .

Ebenso wie Donaldson (1996) verwendete Dunn (2004) nicht nur A' als Sensitivitätsindex, sondern auch d' . In letzterem Index verschwinden die durchgängig beobachteten Unterschiede zwischen der Sensitivität, die auf der Basis der Gesamtdaten ermittelt wurde, und derjenigen, für die nur Remember-Antworten herangezogen wur-

den: $d'_{RK} = 1.52$ unterschied sich kaum von $d'_R = 1.53$. Zudem war d'_R nur in 46.3 % der Bedingungen kleiner als d'_{RK} .

Alles in allem scheint der Schluss von den vorgestellten Daten auf die Gültigkeit eines der in Frage stehenden Modelle dadurch erschwert, dass die Art und Weise, wie das zentrale Maß – in obigem Falle die Sensitivität – erfasst wird, selbst auf Annahmen beruht, die falsch sein können. Leichter wird es, die Gültigkeit von Modellen zu überprüfen, wenn solche Annahmen nicht mehr als außerhalb des Modells stehend angenommen werden, sondern in die Modellannahmen übernommen werden. Dieses Vorgehen entspricht einer Weiterentwicklung der theoretischen Modelle zu Messmodellen, mit deren Hilfe sich die zentralen theoretischen Maße direkt erfassen lassen, ohne dass eine Notwendigkeit besteht, auf methodische Hilfsannahmen zurückzugreifen. Diese Thematik wird im kommenden Kapitel besprochen werden. An dieser Stelle wird deshalb zunächst auf die zweite von Donaldson (1996) aufgestellte Hypothese eingegangen.

In seiner Metaanalyse fand Donaldson (1996) eine nicht zu vernachlässigende Korrelation von $r = .57$ zwischen dem Urteilkriterium, erfasst durch B''_D , und dem Gedächtnisanteil an den Know-Antworten, der durch A'_K , dem auf der Basis von Know-Treffern und Know-Falschen-Alarmen berechneten A' , bestimmt wurde. Wie lässt sich diese Korrelation erklären? Mittlerweile gibt es auf diese Frage zwei Antworten. Donaldson zufolge reicht das eindimensionale Kontinuum-Modell, wie es in Abbildung 3.2 zu finden ist, als Erklärung aus: Da es annimmt, dass die Wahrscheinlichkeit von korrekten und falschen Know-Antworten anhand derjenigen Fläche vorhergesagt werden kann, die unter der jeweiligen Spurstärke-Verteilung in das Fenster zwischen Rekognitions-kriterium und Remember-Kriterium fällt, und weiter davon ausgeht, dass die Spurstärkeverteilung für Distraktoren im Vergleich zu der Spurstärkeverteilung für gelernte Items nach links verschoben ist, werden die Wahrscheinlichkeitsschätzungen für Know-Treffer und Know-Falsche-Alarme immer ungleicher je weiter links dieses Fenster zu finden ist. Berechnet man den Quotienten aus diesen beiden vorhergesagten Wahrscheinlichkeiten, so zeigt sich um den Schnittpunkt der Dichtefunktionen ein weitgehend ausgewogenes Verhältnis. Je weiter man sich aber von diesem entfernt, desto ungleicher wird das Verhältnis, wobei ein Verschieben nach links selbiges zu Gunsten der Falschen Alarme verändert.

Eine alternative Erklärung, weshalb in Remember-Know-Untersuchungen Urteilkriterium und Know-Sensitivität miteinander korrelieren, basiert auf der Annahme, dass Rateprozesse (*Guessing*) dazu führen könnten, dass die Antwortmöglichkeit Know nicht nur im Sinne der Remember-Know-Prozedur gebraucht wurde, sondern von den Probanden vielmehr als eine Art Restkategorie verstanden wurde, die sie immer dann wählten, wenn sie sich nicht im Sinne des Remembering an das Vorkommen eines Reizes erinnern konnten. Einen Ausweg sahen Gardiner, Java und Richardson-Klavehn (1996) in der Remember-Know-Guess-Prozedur (RKG-Prozedur). Dass insbesondere die Häufigkeit der Know-Antworten durch die Variation von Antworttendenzen systematisch beeinflusst werden kann, belegten einige Arbeiten wie z. B. die von Strack und Förster (1995).

In Untersuchungen, die der Remember-Know-Guess-Prozedur folgen, wird den

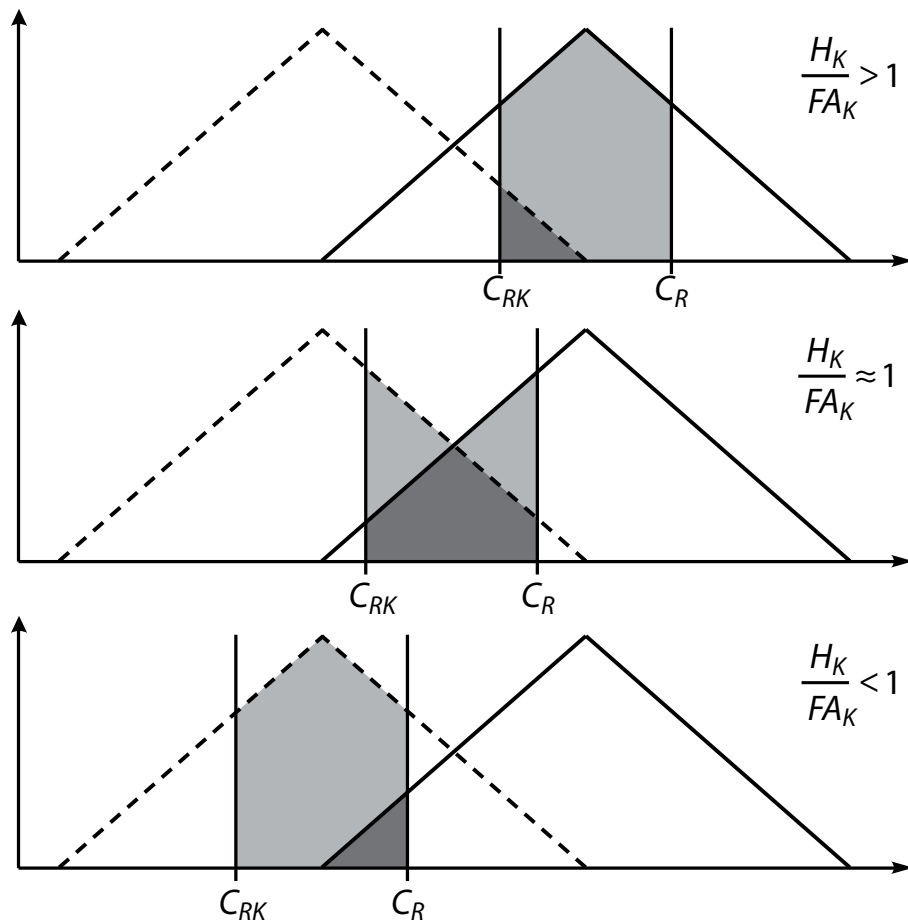


Abbildung 3.2: Dargestellt wird das Eindimensionale Kontinuum-Modell nach Donaldson (1996). Von oben nach unten verändert sich die Lage der Kriterien C_{RK} und C_R von relativ streng bis zu recht liberal. Damit geht eine Verschlechterung des Verhältnisses von Know-Treffern (H_K) zu Know-Falschen-Alarmen (FA_K) einher.

Probanden neben den bereits bekannten Optionen des Remembering und des Knowing eine dritte Antwortmöglichkeit zur Einschätzung ihres Abruferlebens angeboten: Sie erhalten die Möglichkeit, eine Guess-Antwort abzugeben, falls sie der Meinung sind, geraten zu haben. Dadurch, dass den Probanden nun eine extra für diesen Fall vorgesehene Antwortkategorie zur Verfügung steht, soll Gardiner, Java und Richardson-Klavehn (1996) zufolge der Einfluss von Ratenendenzen auf Know-Antworten eliminiert werden. Entsprechend wäre zu erwarten, dass die von Donaldson gefundene Korrelation zwischen Urteilkriterium und Know-Sensitivität verschwindet, sobald den Probanden eine solche zusätzliche Antwortkategorie zur Verfügung gestellt wird. In ein neues Gewand gekleidet sollte der Zusammenhang zwischen Antworttendenz und Rekognitionsurteil dann allerdings wieder auftauchen in einer Korrelation von Urteilstendenz B''_D und der Guess-Sensitivität A'_G .

Belege für diese Annahme fanden Gardiner, Ramponi und Richardson-Klavehn (2002) in einer Metaanalyse mit insgesamt 86 Experimentalbedingungen aus 23 unterschiedlichen Remember-Know-Guess-Arbeiten. Zum einen zeigte sich, wie von den Autoren erwartet, dass die substantielle Korrelation zwischen den Indizes B''_D und A'_K , die in Donaldsons (1996) Metaanalyse über Remember-Know-Datensätze gefunden wurde, nicht mehr beobachtet werden konnte. Sie betrug nur noch $r = .14$. Stattdessen trat wie erwartet eine signifikante Korrelation zwischen Urteilkriterium und Guess-Antworten ($r = .62$) auf. Zudem ist bemerkenswert, dass die Sensitivität für Guess-Urteile mit $A' = .498$, wie man es für Antworten auf dem Rateniveau erwartet, nicht über dem Zufallsniveau lag, das für $A' .5$ beträgt.

In Anbetracht dieser Befunde scheint die von Gardiner, Java und Richardson-Klavehn (1996) erstmals verwendete Remember-Know-Guess-Prozedur ein Fortschritt im Bereich der Erforschung des menschlichen Abruferlebens zu sein. Allerdings bleibt auch bei Anwendung dieses Verfahrens die Frage offen, ob sich das menschliche Abruferleben nicht doch sparsamer – und damit besser – durch das eindimensionale Kontinuum-Modell erklären lässt als durch ein komplexeres Zwei-Komponenten-Modell.

Die Ergebnisse von Rekognitionsuntersuchungen werden meistens anhand von rechnerisch bestimmten Sensitivitäts- und Antworttendenzindizes betrachtet. Dieses Vorgehen ist allerdings nicht die einzige Methode, die zum Zwecke der Überprüfung von Annahmen mit theoretischer Bedeutung angewandt werden kann. Eine alternative Betrachtungsweise von Gedächtnisleistungen und Antworttendenzen in Rekognitionsuntersuchungen stellt die Analyse sogenannter Receiver-Operating-Characteristics (ROCs) dar. ROCs sind Abbildungen, in denen Treffer- und Falsche-Alarm-Raten für verschieden strenge Antwortbedingungen übereinander abgetragen werden. Voraussetzung dafür, dass sich die so entstehenden Kurven sinnvoll interpretieren lassen, ist, dass allen Antwortbedingungen die gleiche Gedächtnisgüte zu Grunde liegt. Deshalb werden ROCs in der gedächtnispsychologischen Forschung auch als Isogedächtniskurven bezeichnet. Aus diesen Kurven können eine Reihe interessanter Information extrahiert werden, die selbige für die Gedächtnisforschung sehr nützlich machen.

Beispielsweise beinhaltet die Abweichung der Treffer-Rate von der Falsche-Alarm-Rate eine Information über die Sensitivität des Gedächtnisapparates. Gleichzei-

tig lassen sich für Gedächtnismodelle Vorhersagen über die Form der ROCs treffen, die bei Gültigkeit des jeweiligen Modells zu erwarten sind. Besonders gut zu überprüfende Vorhersagen lassen sich für eindimensionale Signalentdeckungsmodelle, die eine Normalverteilung der Vertrautheit annehmen, ableiten. Entsprechend werden ROCs, genauer gesagt z -ROC-Kurven, häufig benutzt, um die Plausibilität dieser Modelle zu untersuchen.¹

Wie Ratcliff, Sheu und Gronlund (1992) zeigten, lässt sich aus den Annahmen signalentdeckungstheoretischer Ein-Prozess-Modelle mit der zusätzlichen Annahme normalverteilter Spurstärken ableiten, dass alle Punkte ($z_{Hit}|z_{FA}$) auf einer Geraden liegen müssen, deren Steigung dem Verhältnis der beiden Varianzen σ_{Neu} und σ_{Alt} zueinander entspricht. Unter der Annahme gleicher Varianzen für Alt- und Neu-Verteilung betrüge die vorhergesagte Steigung im z -ROC-Raum also genau 1.

Für z -ROCs, die auf Rekognitionsexperimenten mit Konfidenzurteilen basieren, ermittelten Glanzer, Kim, Hilford und Adams (1999) sowie Ratcliff et al. (1992) in Metaanalysen die Steigung der bestpassenden Geraden. In beiden Arbeiten betrug die Steigung im Mittel 0.80. Demnach beläuft sich die Streuung der Vertrautheitsverteilung von Distraktoren nur auf 80% der Streuung gelernter Reize.

Ein Test der Vorhersage, dass die z -ROC-Kurve eine Gerade darstellt, ist allerdings nur dann möglich, wenn die Daten der Bedingung genügen, dass sie mehr als zwei Konfidenzbedingungen beinhalten, für die eine einheitliche Gedächtnisleistung erwartet werden kann. Ist diese Bedingung erfüllt, so lässt sich die zentrale Vorhersage aller eindimensionalen Gedächtnismodelle mit normalverteilterm Evidenzmaß im z -ROC-Raum überprüfen.

Remember-Know-Untersuchungen erfüllen diese Bedingung jedoch nicht, da mit Remember-Antworten und Know-Antworten nur zwei Antworttendenzbedingungen erhoben werden. Dennoch bieten ROC-Analysen auch für diesen Fall Informationen, anhand derer sich einzelne Annahmen des Ein-Prozess-Modells überprüfen lassen: Zum einen lässt sich die in Remember-Know-Untersuchungen zu Tage tretende Steigung der Zwei-Punkt- z -ROC-Geraden bestimmen. Diese Steigung lässt sich anschließend mit den von Glanzer et al. (1999) ermittelten Werten vergleichen. Zum anderen ist es möglich, die ROC-Kurven von Remember-Know-Untersuchungen direkt mit den anhand von Rekognitionsuntersuchungen mit Konfidenzurteilen bestimmten ROC-Kurven zu vergleichen.

In der Literatur lassen sich beide Ansätze zur Analyse von ROC-Kurven für Remember-Know-Daten finden. So untersuchten Rotello, Macmillan und Reeder (2004) in einer Metaanalyse, inwiefern sich das Varianzverhältnis, das Glanzer et al. (1999) in Rekognitionsuntersuchungen beobachteten, auf Remember-Know-Untersuchungen verallgemeinern lässt. Dazu ermittelten Rotello et al. die mittlere Steigung der z -ROCs, die sie anhand der Proportionen korrekter und falscher Alt-Antworten sowie der korrekten und falschen Remember-Antworten bestimmten. Die Analyse von Rotello et al. ergab, dass die mittlere Steigung solcher Zwei-Punkt-ROCs

¹Unter z -ROC-Kurven versteht man ROC-Kurven im z -transformierten Raum. In diesem Raum gilt, dass alle auf normalverteilten Ausgangsvariablen beruhenden Punkte einer ROC-Kurve auf einer Geraden liegen.

einen Wert von 1.01 annahm. Diese Steigung wich deutlich von der von Glanzer et al. veröffentlichten Schätzung ab. Rotello et al. betrachteten die Unterschiede in den Steigungen von konfidenzbasierten und abrufenerlebensbasierten z -ROCs als Argument gegen eine eindimensionale Interpretation von Remember-Know-Untersuchungen.

Nach Wixted und Stretch (2004) steht dieses Argument jedoch auf tönernen Füßen. Diese Autoren führten an, dass der Unterschied zwischen den von Glanzer et al. (1999) und Ratcliff et al. (1992) analysierten Experimenten und den Remember-Know-Experimenten auch in Besonderheiten der jeweiligen Untersuchungen liegen könnte, wie z. B. der Art des Reizmaterials oder den Eigenschaften der Probandenstichprobe.

Um die von Rotello et al. (2004) aufgestellte Vermutung eines systematischen Unterschieds zwischen den z -ROCs für Konfidenzurteile und denen für Abrufenerlebensberichte strenger zu prüfen, konzentrierten sich Wixted und Stretch (2004) auf Untersuchungen, in denen beide Informationen unter identischen Randbedingungen erhoben wurden. Einen Überblick über die Resultate ihrer Analyse liefert Tabelle 3.3. In den meisten dieser Untersuchungen ließ sich die Hypothese einheitlicher Steigungen verifizieren. Allerdings gab es – wie Wixted und Stretch selbst bemerkten – unter den berichteten Studien einige, die diesen Schluss als voreilig erscheinen ließen. Dabei handelte es sich um die Ergebnisse von Rotello et al. (2004), Rajaram, Hamilton und Bolton (2002) sowie Yonelinas (2001, Exp. 1 & 3). In diesen Untersuchungen erwies sich die Steigung der Remember-Know-Geraden durchweg als steiler als die Steigung der auf Konfidenzurteilen basierenden Geraden.

Leider erlauben ROCs auf der Basis von Remember-Know-Untersuchungen nur die Bestimmung von Steigungen, die anschließend mit einem externen Kriterium – wie z. B. der Steigung von z -ROCs für Konfidenzurteile – verglichen werden können. Ein Test der Vorhersagen des eindimensionalen Kontinuum-Modells ist anhand dieser Daten nicht möglich, da in Remember-Know-Untersuchungen mit Alt-Urteilen und Remember-Antworten nur zwei Konfidenzbedingungen erhoben werden und zwei Punkte immer auf einer Geraden liegen.

Anders stellt sich die Situation in Remember-Know-Guess-Untersuchungen dar. Für solche Untersuchungen lassen sich im z -ROC-Raum drei Punkte festlegen: Der erste ergibt sich, wenn nur Remember-Antworten berücksichtigt werden, der nächste resultiert bei Berücksichtigung von Remember- und Know-Antworten und der letzte Punkt beruht auf allen Alt-Urteilen. Diese drei Punkte können zwar auf einer Geraden im z -ROC-Raum liegen, müssen es aber nicht. Entsprechend ist ein Test der Behauptung, dass alle drei Punkte auf einer Geraden liegen, möglich.

Auf diesen Punkt wird in Kapitel 8 näher eingegangen werden. Vorher werden im nächsten Kapitel sechs mathematische Modelle vorgestellt, die dazu genutzt werden können, die Ergebnisse von Remember-Know- sowie Remember-Know-Guess-Untersuchungen zu erklären.

Tabelle 3.3: Dargestellt werden Steigungen von z -ROC-Geraden, die auf der Basis von Abruferlebensurteilen respektive Konfidenzurteilen gewonnen wurden (nach Wixted & Stretch, 2004).

| | Steigung der ROC-Kurve | |
|--|------------------------|-------------|
| | Konfidenzurteile | R/K Urteile |
| Gardiner & Java (1990) | | |
| Wörter (Experimente 2 und 3) | 0.94 | 0.60 |
| Nonwörter (Experimente 2 und 3) | 0.75 | 1.08 |
| Rajaram (1993) | | |
| wiederholte Primes (Experimente 3 und 4) | 0.74 | 0.68 |
| unspezifische Primes (Experimente 3 und 4) | 0.56 | 0.62 |
| Rajaram et al. (2002) | | |
| Experiment 1 | 0.55 | 0.97 |
| Experiment 2 | 0.66 | 1.20 |
| Rotello et al. (2004) | | |
| Experiment 1 | 0.81 | 0.85 |
| Experiment 2 | 0.91 | 1.03 |
| Stretch & Wixted (1998) | | |
| Experiment 1 | 0.74 | 0.76 |
| Experiment 2 | 0.68 | 0.65 |
| Experiment 3 | 0.59 | 0.57 |
| Experiment 4 | 0.72 | 0.66 |
| Yonelinas (2001) | | |
| Experiment 1 | | |
| volle Aufmerksamkeit | 0.63 | 0.73 |
| geteilte Aufmerksamkeit | 0.76 | 0.81 |
| Experiment 2A | | |
| volle Aufmerksamkeit | 0.64 | 0.60 |
| geteilte Aufmerksamkeit | 0.72 | 0.65 |
| Experiment 2B | | |
| volle Aufmerksamkeit | 0.79 | 0.73 |
| geteilte Aufmerksamkeit | 0.84 | 0.82 |
| Experiment 2C | | |
| volle Aufmerksamkeit | 0.74 | 0.70 |
| geteilte Aufmerksamkeit | 0.85 | 0.80 |
| Experiment 3 | | |
| tiefe Verarbeitung | 0.65 | 0.72 |
| oberflächliche Verarbeitung | 0.78 | 0.84 |

Kapitel 4

Messmodelle zur Remember-Know-(Guess-)Prozedur

Zu Beginn des letzten Kapitels wurden die Besonderheiten des episodischen Gedächtnisses dargestellt, die es von anderen Gedächtnissystemen, insbesondere dem semantischen Gedächtnis, unterscheiden. Eine dieser Eigenschaften besteht darin, dass es dem Menschen nur aufgrund des episodischen Gedächtnisses möglich sein soll, mentale Zeitreisen in die Vergangenheit zu vollführen. Das episodische Gedächtnis ermöglicht damit, einmal Erlebtes vergangener Tage erneut so zu durchleben als befände man sich wieder in der gleichen Situation. Wie dargelegt wurde, belegte Tulving (1983) diese Art des Abruferlebens mit dem Begriff der autonotischen Bewusstheit, die eines der zentralen Unterscheidungskriterien zwischen semantischem Abruf (Knowing) und episodischem Abruf (Remembering) darstellt.

Weiter wurde mit der Remember-Know-(Guess-)Prozedur eine Methode vorgestellt, die mit Hilfe einer einfachen, an die Probanden gerichteten Nachfrage Remembering und Knowing in Rekognitionsuntersuchungen erfassbar macht. Sie soll es auf diese Weise erlauben, die individuellen Beiträge von semantischem und episodischem Gedächtnis zum Rekognitionsgedächtnis zu quantifizieren. Um davon ausgehen zu können, dass Remembering und Knowing, wie sie in der Remember-Know-(Guess-)Prozedur erfasst werden, auch wirklich semantisches und episodisches Gedächtnis wiedergeben, braucht es empirische Evidenz. Wie erläutert wurde, lassen sich Remember- und Know-Antworten durch geeignete experimentelle Manipulationen unabhängig voneinander beeinflussen. Diese funktionale Unabhängigkeit kann einerseits als Indiz dafür interpretiert werden, dass man mit Remembering und Knowing die Leistungen zweier voneinander unabhängiger und autarker Komponenten erfasst. Unklar bleibt dabei jedoch, ob der entscheidende Unterschied zwischen Remembering und Knowing tatsächlich in übergeordneten Gedächtnissystemen gesucht werden muss, wie es nach Tulving (1985a) zu erwarten ist, oder ob die wesentliche Unterscheidung in der Art der ablaufenden Prozesse liegt, wie Rajaram (1998) im Rahmen des Distinktheits-Flüssigkeits-Ansatzes postulierte.

Andererseits existiert neben diesen beiden theoretischen Annäherungen an die Konzepte des Abruferlebens eine dritte Sichtweise, die im vorherigen Kapitel darge-

stellt wurde: Das eindimensionale Kontinuum-Modell von Donaldson (1996) kommt mit der Annahme einer einzigen Komponente aus und kann somit als sparsame Alternative zu komplexeren Zwei-Komponenten-Theorien betrachtet werden. Entsprechend zentral ist für die Fürsprecher der komplexeren Modelle die Auseinandersetzung mit dem eindimensionalen Kontinuum-Modell und das Bestreben, selbiges zu falsifizieren. Einige Versuche, eine Falsifikation des eindimensionalen Kontinuum-Modells herbeizuführen, wurden bereits näher betrachtet. Dabei stand eine Reihe von Metaanalysen im Mittelpunkt, die sich u. a. der Frage widmeten, inwiefern sich die Sensitivität der Remember-Antworten von der Sensitivität unterschied, die aus der Gesamtheit aller Rekognitionsurteile bestimmt wurde. Dies war deshalb von Interesse, da es dem eindimensionalen Kontinuum-Modell zufolge keinen Unterschied machen sollte, ob man ausschließlich Remember-Daten oder sowohl Remember-Daten als auch Know-Daten zur Sensitivitätsbestimmung heranzieht. Wie erläutert wurde, kann diese Frage trotz ihrer scheinbaren Einfachheit aus hauptsächlich methodischen Erwägungen heraus keineswegs als geklärt gelten. Dieser Umstand stellt, wie Dunn (2004) aufzeigte, keinen Einzelfall dar. Die argumentativen Schwierigkeiten bei der Falsifikation dieser einfachen Vorhersage sind vielmehr symptomatisch für eine ganze Reihe von Versuchen, das Donaldson-Modell zu falsifizieren.

In diesem Kapitel werden insgesamt sechs verschiedene mathematische Messmodelle zur Remember-Know-Guess-Prozedur vorgestellt, die allesamt einen Versuch darstellen, Remembering und Knowing theoretisch zu erklären. Bevor jedoch diese sehr konkreten Theorien dargelegt werden, sollen zunächst auf allgemeiner Ebene die Eigenheiten und Vorteile des Arbeitens mit mathematischen Modellen dargestellt werden.

Der Modellierungsansatz als solcher lässt sich am besten teleologisch als ein Weg charakterisieren, auf dem die Kluft zwischen theoretischen Annahmen und empirischen Argumenten überbrückt werden kann. Eine solche Überbrückung ist notwendig, um der Psychologie die Eigenschaft einer empirischen Wissenschaft zu verleihen (vgl. Erdfelder, 2000). Diese verschiedentlich auch als 'Theorie-Empirie-Überbrückungsproblem' bezeichnete Anforderung liegt in dem Umstand begründet, dass psychologische Theorien beobachtbares Verhalten in aller Regel durch latente, d. h. nicht direkt beobachtbare Konstrukte erklären, wie z. B. mentale Systeme, Verarbeitungsprozesse oder auch interne Zustände. Grundlegendes Kennzeichen einer empirischen Wissenschaft ist aber die Überprüfbarkeit theoretischer Annahmen anhand von Beobachtungen. Demnach wäre die Psychologie also nur in den Fällen als eine empirische Wissenschaft zu betrachten, in denen es Wissenschaftlern gelingt, „eine methodologisch tragfähige Brücke zwischen theoretischen Konstrukten einerseits und beobachtbaren Tatbeständen andererseits [...]“ herzustellen (Erdfelder, 2000). Betrachtet man die Geschichte psychologischer Forschung, so lässt sich eine Reihe von Versuchen ausmachen, das Theorie-Empirie-Überbrückungsproblem zu lösen. Während Erdfelder (2000) insgesamt vier solcher Ansätze aufführt, soll es für diese Arbeit genügen, den Modellierungsansatz mit dem klassischen 'Operationalisierungsansatz' zu vergleichen, dem die meisten Arbeiten der modernen psychologischen Forschung verpflichtet sind.

Eine im Sinne des Operationalisierungsansatzes wohl definierte Theorie besteht aus einer Reihe von Aussagen über die empirische Manifestation mentaler Zustände oder Prozesse unter klar umrissenen Beobachtungsbedingungen. Allerdings gelten diese Manifestationsregeln keineswegs als unveränderlich, sondern stellen vielmehr prinzipiell revidierbare Hypothesen dar. Man spricht von diesen Regeln deshalb auch als 'Hilfshypothesen' (Hager, 1992) oder 'Korrespondenzregeln' (Herrmann, 1979). Das Arbeiten mit dem Operationalisierungsansatz ist unproblematisch, solange Konsens darüber besteht, welche Korrespondenzregeln geeignet sind, um eine Theorie zu überprüfen. Besteht hierüber allerdings Uneinigkeit, so kann es geschehen, dass einem theoretischen Begriff mehrere empirische Konzepte zugeordnet werden.

Zudem stellt sich wegen der prinzipiellen Revidierbarkeit der Hilfshypothesen die Frage, in welchem Falle die Theorie durch widersprechende Befunde als falsifiziert gelten kann bzw. wann die Theorie unbeschadet fortbestehen kann und ausschließlich einzelne Hilfshypothesen verworfen werden müssen. Diese Problematik ordnet Erdfelder (2000) als 'Bedeutsamkeitsproblem' in eine Trias von Folgeproblemen ein, mit denen der Operationalisierungsansatz behaftet ist; bei den weiteren Bestandteilen dieser Problemtrias handelt es sich um das 'Dekomponierungsproblem' und das 'Fehlerproblem'. Da der zentrale Vorteil des Modellierens in der Minderung dieser drei Probleme liegt, wird im Folgenden auf alle näher eingegangen werden.

Betrachtet man das Bedeutsamkeitsproblem aus einer etwas anderen Perspektive, so fällt auf, dass es im Verlaufe dieser Arbeit bereits schon einmal präsent war: Die von Dunn (2004) geäußerten Zweifel, ob die Sensitivität in den Metaanalysen von Donaldson (1996) und der Reanalyse von Gardiner und Gregg (1997) korrekt durch A' erfasst wurde oder besser über d' hätte bestimmt werden sollen, sind nichts anderes als ein Ausdruck der Unsicherheit darüber, welches der beiden empirischen Konzepte das bessere ist, um das theoretische Konstrukt 'Sensitivität des Gedächtnisses' zu erfassen. Verschiedene Möglichkeiten zu haben, um ein und dasselbe theoretische Konstrukt zu operationalisieren, ist, wenn auch im Allgemeinen problembelastet (vgl. Erdfelder, 2000, S. 14), nicht unter allen Umständen unerwünscht; wenn man über verschiedene Situationen oder über verschiedene Maße hinweg konvergierende Evidenz erhält, also wenn die verschiedenen Indikatoren auf dieselbe theoretische Deutung hinweisen, kann ein solcher Operationalisierungspluralismus sogar positive Konsequenzen haben.

Bedauernswerterweise war erlaubt dies aber in der gegebenen Situation nicht der Fall; Dunn (2004) wies vielmehr auf widersprüchliche Evidenzen hin, da die Schlussfolgerungen aus den Ergebnissen einzelner Untersuchungen, wie der Metaanalyse von Donaldson (1996) von der Wahl des Sensitivitätsmaßes abhingen. Dieses Bedeutsamkeitsproblem kann nur dann einer befriedigenden Lösung zugeführt werden, wenn empirische Kriterien bereitgestellt werden, anhand derer objektiv zwischen adäquaten und inadäquaten Korrespondenzregeln unterschieden werden kann. Diese Kriterien würden dann durch die Elimination inadäquater Korrespondenzregeln die Willkür bei der Operationalisierung einschränken und so das Bedeutsamkeitsproblem im günstigsten Fall sogar gegenstandslos machen.

In der Praxis lässt sich dieses durch Messmodelle erreichen, die die Eigenschaft der T -Theorizität (vgl. Westermann, 1987; Westmeyer, 1989; 1992) erfüllen. T -

Theorizität lässt sich nach Stegmüller (1973, S. 47) mit den folgenden Worten umschreiben:

Theoretisch in Bezug auf eine Theorie T sind genau diejenigen Größen oder Funktionen, deren Werte sich nicht berechnen lassen, ohne auf diese Theorie T selbst (genauer auf die erfolgreich angewendete Theorie T) zurückzugreifen.

Bezogen auf das eindimensionale Kontinuum-Modell bestünde die notwendige Konsequenz, die aus dem von Dunn (2004) aufgezeigten Bedeutsamkeitsproblem zu ziehen ist, in der Entwicklung eines Messmodells, das t -theoretisch in o. g. Sinne ist. Die adäquate Lösung bestünde demnach darin, eine Annahme über die Form der Spurstärkeverteilung und damit über das Wesen des korrekten Sensitivitätsmaßes in den Kern der Theorie zu übernehmen und nicht in einer Hilfhypothese zu belassen. Um das eindimensionale Kontinuum-Modell zu einem Messmodell fortzuentwickeln, könnte also postuliert werden, dass sich die Spurstärken im Gauß'schen Sinne normalverteilen, weshalb die Sensitivität durch d' erfasst werden sollte. Formal betrachtet stellt ein solches Messmodell eine Abbildung f dar, die theoretische Konstrukte (T) auf empirische Argumente (E) projiziert. Für Messmodelle von besonderer Bedeutung sind zwei Eigenschaften dieser Abbildung $f : T \rightarrow E$, die man unter die Begriffe 'Nicht-Surjektivität' und 'Injektivität' fassen kann. Nicht-surjektiv ist eine Abbildung dann, wenn die Theorie nicht alle empirisch möglichen Beobachtungen zum Gegenstand hat, sondern einige prinzipiell denkbare Ereignisse gerade nicht vorher sagt, so dass Beobachtungen dieser Ereignisse dazu genutzt werden können, die Theorie selbst zu falsifizieren. Injektivität wiederum besagt, dass unterschiedliche theoretische Zustände in voneinander unterscheidbare empirische Ausgänge münden, so dass es – die Gültigkeit der Theorie vorausgesetzt – möglich wird, von den empirischen Ausgängen eindeutig auf die dahinter liegenden theoretischen Zustände zu schließen.

Auch diese beiden Eigenschaften von Messmodellen lassen sich am Beispiel des eindimensionalen Kontinuum-Modells veranschaulichen. Die empirische Seite dieses Modells besteht aus Antworthäufigkeiten, die anhand dreier latenter Konstrukte auf theoretischer Seite erklärt werden. Diese drei Konstrukte lassen sich grob in zwei Kategorien einteilen: die Gedächtnissensitivität d' einerseits und die Antworttendenz, erfasst durch die Antwortkriterien C_R und C_K , andererseits. Antworttendenz und Sensitivität sind in dem Modell beide unverzichtbare Bestandteile, da auf der Basis von nur einem dieser beiden Konstrukte spezifische Rekognitions- oder Antworthäufigkeiten weder verstanden noch erklärt werden können. Durch das entworfene Modell werden so also die theoretischen Annahmen über das Zusammenspiel von Antworttendenz und Gedächtnisleistung auf die empirisch zu erhaltenden Antworthäufigkeiten abgebildet. In der Realität kann sich dann zeigen, ob sich die Antworthäufigkeiten wie vom Modell vorhergesagt gestalten oder ob sich bedeutsame Abweichungen von der Modellvorhersage offenbaren.

Diese enge Verzahnung von Antworttendenz und Sensitivität ist keineswegs nur typisch für das Kontinuum-Modell, vielmehr wird damit einem Problem Tribut gezollt, das jedes Modell des Rekognitionsgedächtnisses lösen muss. Ein anderer Ansatz,

dieses Problem zu lösen, besteht darin, die Annahme zu treffen, dass korrekte Alt-Antworten in Rekognitionstests generell auf einer von zwei möglichen Ursachen beruhen: Entweder erinnert die Versuchsperson sich an das Vorkommen eines Reizes oder sie errät, dass der fragliche Stimulus alt ist. Dieses theoretische Konzept des Zusammenwirkens von Gedächtnis und Antworttendenz ist typisch für die Klasse der multinomialen Verarbeitungsbaum-Modelle, die auch unter dem Namen Schwellenmodelle bekannt geworden sind. Genauso wenig wie mit dem veränderten eindimensionalen Kontinuum-Modell kann mit einem Schwellenmodell geklärt werden, welcher dieser beiden Prozesse – Erinnern oder Raten – ursächlich ist für das Zustandekommen einer einzelnen Alt-Antwort. Beide Modelle stellen ausschließlich Mittel zur Verfügung, um Antworttendenzen und Gedächtnisleistung in ganzen Rekognitionstests voneinander zu separieren. Dadurch, dass diese Modelle identifizierbare Modellparameter zur Verfügung stellen, die die beiden Prozesse repräsentieren, lösen sie ein für alle Rekognitionsexperimente typisches Dekomponierungsproblem.

Allgemein besteht ein Dekomponierungsproblem dann, wenn mehrere theoretische Konstrukte in derselben Weise auf einen empirischen Indikator einwirken. In diesem Falle liegt die beste Lösung darin, den empirischen Indikator so in seine Bestandteile zu untergliedern, dass am Ende einzelne Maße für alle beteiligten Prozesse resultieren.

Auch das Wissen um die Möglichkeiten zur Lösung des Dekomponierungsproblems hilft bei einem im vorherigen Kapitel angetroffenen Problem weiter. Wie berichtet wurde, wies Donaldson (1996) auf eine substantielle Korrelation zwischen der Antworttendenz und der Know-Sensitivität hin. Erklären ließ sich dieses Phänomen unter anderem durch einen Rateprozess, der die Häufigkeit von Know-Antworten bedeutsam beeinflusste. In diesem Fall würden zumindest die Häufigkeiten der Know-Antworten von zwei theoretischen Konstrukten profitieren: Neben dem semantischen Gedächtnis nehmen Rateprozesse Einfluss auf die Anzahl selbiger. Eine solche Kontamination mit Rateprozessen muss allerdings keineswegs auf die Häufigkeiten der Know-Antworten beschränkt bleiben, vielmehr können auch Remember-Antworten Rateeinflüssen unterliegen. Entsprechend besteht bei Remember- und Know-Antworten eine zentrale Dekomponierungsaufgabe darin, zwischen Rateeinflüssen auf der einen Seite und Gedächtniseinflüssen auf der anderen Seite zu trennen.

Einen ersten Versuch, solch eine Separierung herbeizuführen, wurde bereits dargestellt: Gardiner, Java und Richardson-Klavehn (1996) fügten der Abfrage des Abruferlebens eine Guess-Kategorie hinzu, die von den Probanden immer dann gewählt werden sollte, wenn sie geraten haben. Damit könnte das Separierungsproblem gelöst werden, wenn Probanden durchgängig in der Lage wären, anhand ihres subjektiven Erlebens zwischen einem Gedächtnisabruf und Raten zu unterscheiden. Remember-Know-Guess-Experimente, in denen die Antworttendenz alleine manipuliert wurde, ohne dass sich die Lern- und Abrufumstände zwischen den Bedingungen veränderten, sprechen jedoch dagegen, dass das Dekomponierungsproblem auf diese Weise gelöst werden kann. In einer Reihe solcher Untersuchungen konnte nämlich gezeigt werden, dass experimentell induzierte Antworttendenzen auch dann die Häufigkeiten der Remember- und Know-Urteile systematisch beeinflussten, wenn den Probanden die Guess-Kategorie als Antwortmöglichkeit zur Verfügung stand.

Erdfelder und Bayen (2002) legten beispielsweise Ergebnisse von zwei Experimenten vor, in denen die Antworttendenz über die Auszahlungsbedingungen manipuliert wurde. Hierbei wurden in einer Bedingung richtige Alt-Antworten stärker belohnt als richtige Neu-Antworten und falsche Neu-Antworten stärker bestraft als falsche Alt-Antworten, während in einer anderen Bedingung das Gegenteil der Fall war. In diesen beiden Untersuchungen zeigte sich ein Einfluss der Antworttendenzmanipulation auf alle Arten von Alt-Antworten. In die gleiche Richtung weist die Arbeit von Xu und Bellezza (2001). Diese Autoren haben die relative Anzahl alter Stimuli in der Testphase variiert und dabei feststellen können, dass sich die Häufigkeiten korrekter sowie inkorrekt Remember-Antworten mit der Anzahl aus der Lernphase bekannter Stimuli im Test erhöhte.

Diese Befunde legen nahe, dass das Dekomponierungsproblem durch die Aufnahme der Guess-Kategorie nicht vollends gelöst wurde, da weiterhin in allen Antwortbedingungen Rateinflüsse festgestellt werden konnten. Entsprechend bleibt die Frage bestehen, auf welche Weise die einzelnen Antworthäufigkeiten so dekomponiert werden können, dass letztendlich Maße resultieren, die Gedächtnisprozesse und Antworttendenzprozesse getrennt erfassen. Eine vorstellbare Lösungsmöglichkeit liegt sicher in einem Modell, das neben Gedächtniseinflüssen, wie zum Beispiel dem semantischen bzw. dem episodischen Gedächtnis, für jede Antwortkategorie spezifische Rateprozesse annimmt. Für den Erfolg dieses Vorgehens von entscheidender Bedeutung ist die Annahme, dass, wie in Abbildung 4.1 dargestellt, zwar korrekte Remember- und Know-Antworten durch einen Abruf aus dem Gedächtnis verursacht sein können, dies aber nicht für die falschen Remember- und Know-Antworten zutrifft. Wenn man annimmt, dass die letztgenannten nur auf Rateprozessen beruhen, so rückt eine Separierung der Einflüsse der beiden Prozessarten in greifbare Nähe: Anhand der Remember- und Know-Falschen-Alarme ließe sich das für diese Antwortkategorien spezifische Guessing erfassen, welches dann aus den Remember- respektive Know-Treffern herausgerechnet werden kann. Am Ende würden, wie erwünscht, Indizes resultieren, die Antworttendenz einerseits und Rateprozesse andererseits getrennt voneinander wiedergeben. Wie diese Indizes konkret bestimmt werden können, wird aufgezeigt werden, sobald die einzelnen Modelle zur Remember-Know-Guess-Prozedur vorgestellt werden. An dieser Stelle soll deshalb die grobe Beschreibung des Vorgehens genügen und auf das nächste Problem eingegangen werden.

Dazu empfiehlt sich noch einmal ein Blick auf Abbildung 4.1, in der dargestellt ist, wie man im Rahmen eines Zwei-Komponenten-Modells Remembering, Knowing und Guessing konzipieren kann. Zuerst einmal lässt sich der Kern der Zwei-Komponenten-Theorien ausmachen, der darin besteht, dass zwei unterschiedliche Gedächtniskomponenten, im Beispiel das episodische und das semantische Gedächtnis, als Ursachen für korrekte Remember-Antworten einerseits und korrekte Know-Antworten andererseits postuliert werden. Neben diesen Gedächtniseinflüssen müssen zusätzlich die in gestrichelten Ellipsen dargestellten Antworttendenzinflüsse angenommen werden, die durch eigene Indizes erfasst werden. Weiterhin findet sich mit dem Fehler in der Abbildung eine dritte, bisher vernachlässigte Einflussquelle. Neben den systematischen Einflüssen auf die beobachteten Häufigkeiten, die in Modellen durch eigene Modellpa-

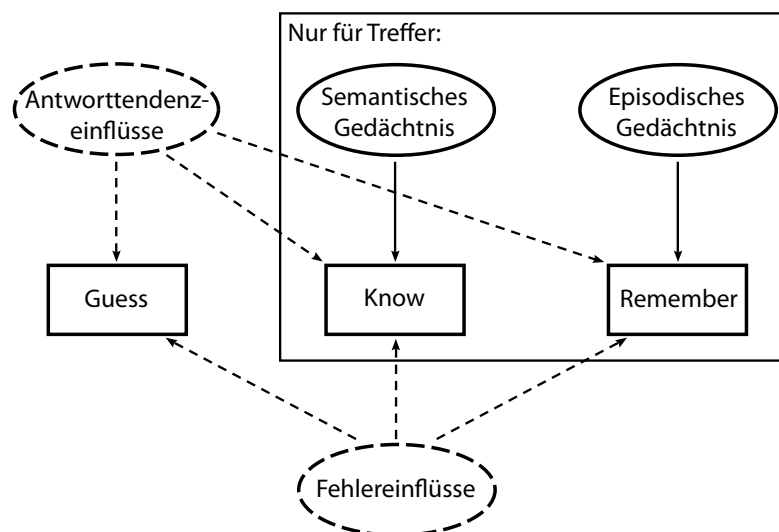


Abbildung 4.1: Dargestellt werden mögliche Faktoren, die auf die Häufigkeiten, mit denen Remember, Know oder Guess geantwortet wird, Einfluss nehmen. In diesem Konzept wird davon ausgegangen, dass das semantische Gedächtnis ausschließlich Know-Treffer und das episodische Gedächtnis ausschließlich Remember-Treffer beeinflusst. Für alle Antwortarten werden zudem Antworttendenz-einflüsse und Fehlereinflüsse angenommen.

parameter repräsentiert werden, sind die beobachtbaren Häufigkeiten auch unsystematischen Fehlereinflüssen unterworfen, die in den Modellen nicht durch eigene Parameter erfasst werden.

Diese Fehlereinflüsse sind es, denen das letzte von Erdfelder (2000) benannte Problem seinen Namen verdankt. Sie sind insofern problematisch, als durch ihre Präsenz die Widerlegung einer Theorie oder eines Modells deutlich erschwert wird, da Ereignisse, die einem Modell zufolge nicht zugelassen sind, aufgrund von Fehlereinflüssen zufällig beobachtet werden können. In diesem Falle sollte die Theorie nicht verworfen werden, sondern erkannt werden, dass die widersprüchlichen Beobachtungen Zufallsergebnisse und damit vernachlässigbar sind. Die Lösung des Fehlerproblems besteht entsprechend in geeigneten statistischen Konzepten, mit deren Hilfe festgestellt werden kann, ob ein Ergebnis dem Zufall geschuldet ist oder als im statistischen Sinne bedeutsam erachtet werden muss.

Bei der mathematischen Modellierung kann die statistische Absicherung von Ergebnissen durch Modellgütetests vorgenommen werden, bei denen die Modellannahmen mit der Nullhypothese eines statistischen Hypothesentests verknüpft werden. Bei einem derartigen statistischen Test bewährt sich ein Modell dann, wenn der herangezogene Datenanpassungsindex unter einem kritischen Vergleichswert bleibt. Solche Datenanpassungsindizes basieren auf der Berechnung von Abweichungen zwischen den Modellvorhersagen und den tatsächlich beobachteten Antworthäufigkeiten. Wie diese statistische Absicherung konkret geschieht, wird im nächsten Kapitel vorgestellt wer-

den.

Zunächst sollen insgesamt sechs Modelle erörtert werden, anhand derer man die Ergebnisse von Remember-Know-Guess-Untersuchungen erklären kann. Bei allen diesen Modellen handelt es sich um multinomiale Modelle, die eine Reihe der bereits bekannten Eigenschaften aufweisen: Alle Modelle liefern anhand ihrer Modellgleichungen Vorhersagen für die Häufigkeiten, mit denen Probanden auf gelernte Reize bzw. Distraktoren mit Remember-, Know-, Guess- und Neu-Antworten reagieren. Dazu werden mit Hilfe von Modellgleichungen und den darin enthaltenen Modellparametern die beobachteten Häufigkeiten in Indizes dekomponiert, die spezifische latente Konstrukte repräsentieren. Diese theoretischen Konstrukte lassen sich in zwei Klassen einteilen abhängig davon, ob sie Gedächtnisleistung erfassen oder Antworttendenzen abbilden. Jedes der vorzustellenden Modelle löst also das im vorangegangenen Abschnitt diskutierte Dekomponierungsproblem.

Allgemein lassen sich die zu beschreibenden Modelle anhand von zwei zentralen Eigenschaften in drei Klassen einteilen. Die zentralen Klassifikationsmerkmale sind erstens, ob die Modelle ein oder zwei Komponenten annehmen, um Remembering und Knowing zu erklären, und zweitens, ob sie die angenommenen Gedächtnisprozesse signalentdeckungstheoretisch oder als Hochschwelle konzeptualisieren. Fünf der sechs Modelle nehmen zwei Gedächtniskomponenten an. Von diesen vier Modellen ist eines ein reines Hochschwellenmodell, bei zweien handelt es sich um rein signalentdeckungstheoretische Modelle und das vierte ist ein hybrides Modell, das einen Prozess als Hochschwelle konzeptualisiert und den anderen signalentdeckungstheoretisch auffasst. Das einzige eindimensionale Erklärungsmodell beruht in weiten Teilen auf den Annahmen des eindimensionalen Kontinuum-Modells von Donaldson (1996), die um Zusätze, wie z. B. die bereits angesprochene Normalverteilungsannahme ergänzt und so zum Signalentdeckungsmodell von Hirshman und Master (1997) ausgebaut wurden. Dieses bereits gut bekannte Exemplar wird nun als erstes der sechs vorgeschlagenen Modelle vorgestellt werden. Im Anschluss daran wird das Hochschwellenmodell von Erdfelder, Cüpper, Auer und Undorf (Im Druck) behandelt. An dritter und vierter Stelle folgen dann die signalentdeckungstheoretisch konzipierten Zwei-Prozess-Modelle, bei denen es sich um das Source-of-Activation-Confusion-Modell sowie das Sum-Difference-Theory-Of-Remembering-and-Knowing genannte Modell handelt. Das zuerst genannte Modell geht auf Arbeiten von Reder et al. (2000) zurück, während letzteres von Rotello et al. (2004) vorgeschlagen wurde. Abgeschlossen wird dieses Kapitel durch die Darstellung des hybriden Hochschwellen-Signalentdeckungsmodells von Yonelinas und Jacoby (1995). Im Verlaufe dieses Kapitels werden zwei Varianten des Hochschwellen-Signalentdeckungsmodells vorgestellt werden. Während die erste Variante direkt auf den Arbeiten von Yonelinas und Mitarbeitern beruht, behandelt der letzte Abschnitt dieses Kapitels eine Erweiterung dieses Modells, innerhalb dessen für den signalentdeckungstheoretischen Zweig Varianzunterschiede zwischen Distraktoren und alten Reizen zugelassen werden.

4.1 Das eindimensionale Signalentdeckungsmodell

Unabhängig von Donaldson (1996) schlugen auch Hirshman und Master (1997) ein eindimensionales Signalentdeckungsmodell (SDM) zur Remember-Know-Prozedur vor. Dieses kann als eine Variante des klassischen Signalentdeckungsmodells für das Rekognitionsgedächtnis (Glanzer, Adams, Iverson & Kim, 1993; Green & Swets, 1966) betrachtet werden und basiert auf den folgenden drei Annahmen:

1. Alle Reize lassen sich auf einer Vertrautheits- oder Spurstärkedimension anordnen. Bei der Vertrautheit handelt es sich um eine kontinuierliche Variable, die, wie Gillund und Shiffrin (1984) vorgeschlagen haben, der Evidenz dafür entspricht, dass ein Stimulus vorher schon einmal präsentiert worden ist. Als Basis für Remember- und Know-Antworten wird demnach ein eindimensionales und kein zweidimensionales Gedächtnismodell angenommen;
2. Die sich überschneidenden Vertrautheitsverteilungen für alte und neue Items werden als normalverteilt angenommen. Dabei zeichnet sich jede der beiden Verteilungen durch einen eigenen Mittelwert und eine eigene Varianz aus. Allerdings wird der Einfachheit halber der Mittelwert der Vertrautheitsverteilung für die Distraktoren auf 0 gesetzt und deren Streuung auf 1 normiert. Der Mittelwert alter Reize wird als μ_{Alt} und deren Varianz als σ_{Alt} bezeichnet;
3. Die Probanden setzen entlang der Vertrautheitsachse zwei Kriterien. Das erste Kriterium (C_O) dient dazu, Alt-Neu-Urteile zu treffen; wenn die Spurstärke eines Items selbiges übersteigt, wird eine Alt-Antwort gegeben. Das strengere Kriterium (C_R) hingegen wird benutzt, um das im Anschluss daran anstehende Remember-Know-Urteil zu treffen: Liegt die Spurstärke eines Items oberhalb dieses Kriteriums, gibt die Versuchsperson eine Remember-Antwort, bleibt die Vertrautheit hingegen in dem Bereich zwischen C_O und C_R , so resultiert eine Know-Antwort.

Um mit diesem Modell nicht nur Ergebnisse von Remember-Know- sondern auch solche von Remember-Know-Guess-Untersuchungen erklären zu können, bedarf es einer minimalen Veränderung der unter Punkt 3 skizzierten Annahme. Durch ein zusätzliches Kriterium C_K , das zwischen C_O und C_R angesiedelt ist, lässt sich die Proportion der Guess-Antworten erfassen. Die Spurstärke von Reizen, die mit Know-Antworten belegt werden, liegt dann nicht nur über C_O sondern zusätzlich über C_K . Wird nur C_O , nicht aber C_K überschritten, so resultieren Guess-Antworten, deren Wahrscheinlichkeit sich aus der Fläche ergibt, die zwischen den Kriterien C_O und C_K eingeschlossen wird. Diese Zusatzannahme entspricht der von Gardiner et al. (2002) getroffenen, die aus einem Remember-Know-Signalentdeckungsmodell mit gleichen Varianzen ein entsprechendes Remember-Know-Guess-Modell entwickelten. In Abbildung 4.2 wird die Erweiterung des von Hirshman und Master (1997) vorgeschlagenen Modells dargestellt, das, wie oben geschildert, unterschiedliche Varianzen für alte und neue Reize annahm.

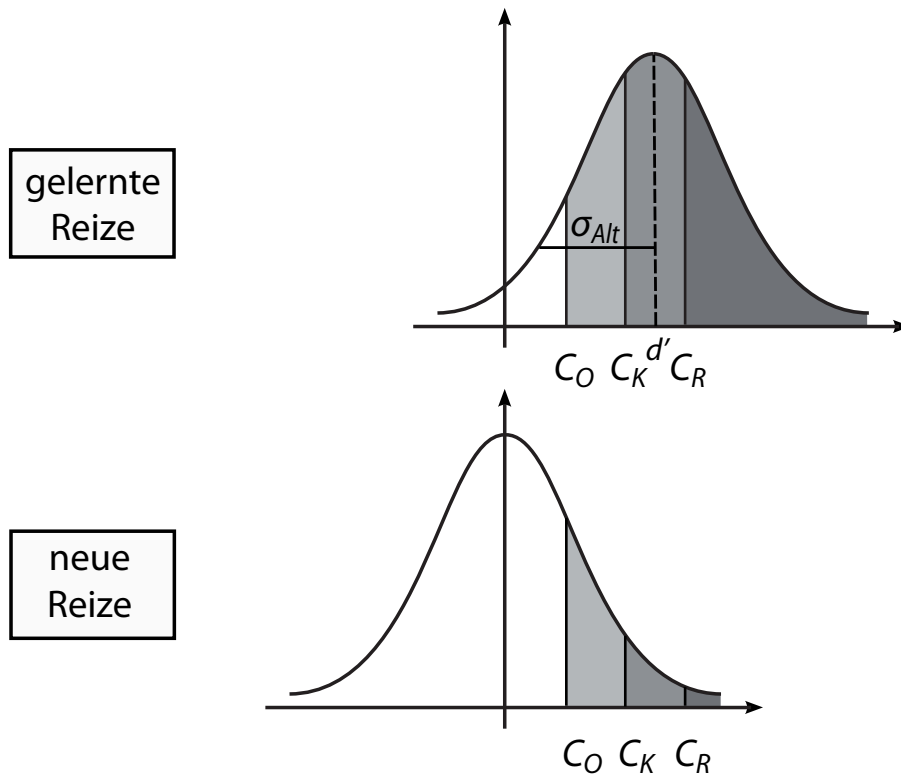


Abbildung 4.2: Dargestellt wird das Eindimensionale Signalentdeckungsmodell nach Hirshman und Master (1997), erweitert um die Annahme eines Kriteriums C_K zur Unterscheidung zwischen Know- und Guess-Antworten. Daneben enthält das Modell den Sensitivitätsparameter d' , den Streuungsparameter σ_{Alt} und zwei weitere Antwortkriterien: C_O zur Unterscheidung alter Reize von neuen Reizen sowie C_R zur Unterscheidung zwischen Remember- und Know-Antworten.

Wie bereits erwähnt, lassen sich anhand der Kriterien und der spezifischen Vertrauensverteilungen Wahrscheinlichkeiten für alle beobachtbaren Ereignisse ableiten. Die Wahrscheinlichkeit, mit der auf einen alten Reiz eine Remember-Antwort folgt, entspricht der Fläche unter der Vertrauensverteilung für alte Reize, die über C_R liegt. Die Wahrscheinlichkeit für Know-Treffer lässt sich analog dazu bestimmen durch die Fläche zwischen C_K und C_R unter derselben Verteilung. Die unter dieser Verteilung zwischen C_O und C_K eingeschlossene Fläche repräsentiert die Häufigkeiten von Guess-Treffern, während der Bereich links von C_O die Häufigkeiten von Neu-Antworten wiedergibt. Die Häufigkeiten von Remember-, Know-, und Guess-Falschen-Alarmen sowie die der korrekten Zurückweisungen lassen sich in analoger Weise unter der Vertrauensverteilung für die Distraktoren in den Bereichen über C_R , zwischen C_R und C_K bzw. C_K und C_O sowie unterhalb von C_O finden.

Anstatt als Fläche unter einer Spurstärkeverteilung können die Antwortwahrscheinlichkeiten mathematisch auch als das Integral der Normalverteilungsdichtefunktion in den oben genannten Grenzen bezeichnet werden. Die Wahrscheinlichkeitsfunktion der Standardnormalverteilung wird im Verlaufe des Textes mit Φ abgekürzt werden, während die dazu gehörige Dichtefunktion ϕ genannt wird. Die exakten Formeln können den Gleichungen 4.1 und 4.2 entnommen werden.

$$\Phi(x) = \int_{-\infty}^x \phi(t) dt \quad (4.1)$$

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad (4.2)$$

Die acht Modellgleichungen, auf denen das an dieser Stelle zu erläuternde Modell beruht, sind in den Gleichungen 4.3 bis 4.10 niedergelegt. Sie beinhalten für eine Untersuchung, die aus einer einzigen Lern- und Abrufbedingung besteht, d. h. in der eine einzige Klasse von Reizen gelernt wurde, zu der in der Testphase genau eine Distraktorklasse hinzukommt, genau fünf Modellparameter: die drei Kriteriumsparameter C_R , C_K und C_O sowie die Verteilungsparameter μ_{Alt} und σ_{Alt} . Anhand dieser Modellparameter können auf Basis der Modellannahmen die Antwortwahrscheinlichkeiten für Remember-Treffer ($p(R|Alt)$) Know-Treffer ($p(K|Alt)$) und Guess-Treffer ($p(G|Alt)$), sowie der Falschen Zurückweisungen ($p(N|Alt)$) ebenso beschrieben werden wie die entsprechenden Wahrscheinlichkeiten für die Distraktoren ($p(R|Neu)$, $p(K|Neu)$, $p(G|Neu)$, $p(N|Neu)$). Das vorgestellte Gleichungssystem lässt sich ohne Schwierigkeit auch auf komplexere Untersuchungsdesigns anwenden; dabei werden für jede hinzukommende Reizklasse zusätzliche Gleichungen aufgenommen und, falls nötig, die Menge der Modellparameter erweitert.

$$p(R|Alt) = \Phi((d' - C_R)/\sigma_{Alt}) \quad (4.3)$$

$$p(K|Alt) = \Phi((d' - C_K)/\sigma_{Alt}) - \Phi((d' - C_R)/\sigma_{Alt}) \quad (4.4)$$

$$p(G|Alt) = \Phi((d' - C_G)/\sigma_{Alt}) - \Phi((d' - C_K)/\sigma_{Alt}) \quad (4.5)$$

$$p(N|Alt) = \Phi((C_G - d')/\sigma_{Alt}) \quad (4.6)$$

$$p(R|Neu) = \Phi(-C_R) \quad (4.7)$$

$$p(K|Neu) = \Phi(-C_K) - \Phi(-C_R) \quad (4.8)$$

$$p(G|Neu) = \Phi(-C_G) - \Phi(-C_K) \quad (4.9)$$

$$p(N|Neu) = \Phi(C_G) \quad (4.10)$$

4.2 Das Vier-Abrufzustände-Modell

Das Vier-Abrufzustände-Modell (*4-States-Modell*) von Erdfelder und Mitarbeitern (Erdfelder, 2000; Erdfelder & Bayen, 2002; Erdfelder et al., Im Druck) gehört zu der allgemeinen Klasse der multinomialen Verarbeitungsbaummodelle und reiht sich dort in die Familie der Zwei-Hochschwellen-Modelle des Rekognitionsgedächtnisses ein. Die Basis des Modells erinnert deshalb an das Zwei-Hochschwellen-Modell, dem zufolge sich die Sensitivität von Rekognitionsurteilen durch $P_r = H - F$ und die Antworttendenz durch $B_r = F/(1 - P_r)$ bestimmen lässt (Snodgrass & Corwin, 1988). Zudem ist das 4-States-Modell eng verwandt mit Modellen, die sich in anderen Paradigmen der Gedächtnisforschung bewährt haben, wie z. B. dem Modell zum Prozess-Dissoziations-Paradigma von Erdfelder und Buchner (1995) oder den Modellen zum Quellengedächtnis von Bayen, Murnane und Erdfelder (1996), von Klauer und Wegener (1998) sowie von Meiser und Bröder (2002).

Die zentrale Idee des 4-States-Modells besteht darin, dass bei Gedächtnisaufgaben vier Zustände unterschieden werden müssen, bei denen es sich erstens um rekollektives Erinnern, zweitens um Vertrautheit, drittens um Unsicherheit und viertens um Distraktorentdeckung handelt. Die genuin gedächtnisbezogenen Zustände der Vertrautheit und der rekollektiven Erinnerung resultieren aus zwei unterschiedlichen Gedächtnisprozessen und bilden die Basis für Remember- und Know-Treffer. Das rekollektive Erinnern fördert dabei Remember-Antworten, während Know-Antworten von Vertrautheit profitieren.

Remember- und Know-Antworten können gemäß dem 4-States-Modell jedoch nicht nur durch die beiden erläuterten Abrufprozesse zustande kommen. Zusätzlich wird davon ausgegangen, dass die Probanden aus einem Zustand der Unsicherheit heraus nicht nur Neu- und Guess-Antworten, sondern auch Remember- sowie Know-Urteile abgeben können. Zur Modellierung der Antworten auf alte Reize werden im 4-States-Modell also drei der vier Zustände benötigt: der Zustand des rekollektiven Erinnerns, der unweigerlich zu Remember-Antworten führt, der Zustand der Vertrautheit, der in Abwesenheit von rekollektiver Erinnerung zu Know-Antworten führt, und der Unsicherheitszustand, der in Guess-Antworten münden kann, aber auch zur Nennung einer der anderen Antwortkategorien führen kann. Der vierte Zustand, die Distraktorentdeckung, ist nur bei neuen Reizen, deren Neuheit dem Modell zufolge von Probanden entdeckt werden kann, von Bedeutung. Die Neuheitsentdeckung führt dann unweigerlich zu sicheren Neu-Urteilen. Die Annahme eines solchen Distraktorentdeckungsmechanismus ist definitorischer Bestandteil eines jeden

Zwei-Hochschwellen-Modells. Bei ausbleibender Neuheitsentdeckung kann der Proband auch bei neuen Reizen unsicher darüber sein, ob der Reiz gelernt worden ist oder nicht. In diesem Falle kann Raten zu allen Arten von Alt-Antworten führen. Ansonsten resultieren aus dem Unsicherheitszustand korrekte Neu-Urteile.

$$p(R|Alt) = r + (1 - r) \cdot (1 - k) \cdot g_R \quad (4.11)$$

$$p(K|Alt) = (1 - r) \cdot k + (1 - r) \cdot (1 - k) \cdot (1 - g_R) \cdot g_K \quad (4.12)$$

$$p(G|Alt) = (1 - r) \cdot (1 - k) \cdot (1 - g_R) \cdot (1 - g_K) \cdot g_G \quad (4.13)$$

$$p(N|Alt) = (1 - r) \cdot (1 - k) \cdot (1 - g_R) \cdot (1 - g_K) \cdot (1 - g_G) \quad (4.14)$$

$$p(R|Neu) = (1 - d) \cdot g_R \quad (4.15)$$

$$p(K|Neu) = (1 - d) \cdot (1 - g_R) \cdot g_K \quad (4.16)$$

$$p(G|Neu) = (1 - d) \cdot (1 - g_R) \cdot (1 - g_K) \cdot g_G \quad (4.17)$$

$$p(N|Neu) = d + (1 - d) \cdot (1 - g_R) \cdot (1 - g_K) \cdot (1 - g_G) \quad (4.18)$$

Diese Annahmen finden sich sowohl in den Modellgleichungen 4.11 – 4.18 als auch in der Abbildung 4.3 wieder, in der das 4-States-Modell als Entscheidungsbaum dargestellt ist. Für ein einfaches Einbedingungsexperiment, in dem eine einzige Reizklasse gelernt worden ist und in der Abrufphase eine Distraktorklasse hinzukommt, besteht diese Darstellung wie abgebildet aus zwei Bäumen; jeweils einer wird benötigt, um die Reaktionen auf alte Reize bzw. Distraktoren wiederzugeben. Die Stimulusart (*Alt* bzw. *Neu*) ist links an der Wurzel eines jeden Unterbaumes in einem rechteckigen Kasten notiert. Von der Wurzel ausgehend werden die Wahrscheinlichkeiten, einen bestimmten Zustand zu erreichen oder zu verfehlen, in binärer Weise abgebildet. Die Parameter, die sich an den Wegen zwischen je zwei Zuständen befinden, können als die Wahrscheinlichkeiten verstanden werden, mit denen das nachfolgende Ereignis eintritt, gegeben alle vorausgehende Ereignisse sind bereits eingetreten. Jeder Ast des Baumes mündet am rechten Rand in eine Beobachtungskategorie, bei der es sich um Remember-, Know-, Guess- oder Neu-Antworten handeln kann. Die vorhergesagte Antwortwahrscheinlichkeit für eine bestimmte Antwortkategorie ergibt sich als Summe der Astwahrscheinlichkeiten, die zu der entsprechenden Beobachtungskategorie führen. Die Astwahrscheinlichkeiten ihrerseits lassen sich berechnen als Produkte aller entlang des jeweiligen Astes anzutreffenden Übergangswahrscheinlichkeiten.

Die Wahrscheinlichkeit einer Remember-Antwort bei alten Reizen ergibt sich bei Anwendung dieser Regel als die Summe der Wahrscheinlichkeit r , mit der ein Item rekollektiv erinnert wird, und dem Produkt $(1 - r) \cdot (1 - k) \cdot g_R$. Dieser zweite Summand gibt die Wahrscheinlichkeit wieder, mit der aus dem mit der Wahrscheinlichkeit $(1 - r) \cdot (1 - k)$ erreichten Unsicherheitszustand heraus Remember geraten wird (g_R). Know-Antworten ereignen sich mit einer Wahrscheinlichkeit, die der Summe zweier Produkte entspricht. Diese Produkte sind zum einen $(1 - r) \cdot k$, das für Vertrautheit (k) in Abwesenheit von rekollektiver Erinnerung ($1 - r$) steht, und zum anderen $(1 - r) \cdot (1 - k) \cdot (1 - g_R) \cdot g_K$, das die Wahrscheinlichkeit dafür ausdrückt, dass aus Unsicherheit heraus und in der Abwesenheit von Remember-Raten ($1 - g_R$) Know geraten wird (g_K). Da in die Guess-Kategorie nur ein Ast mündet, lässt sich

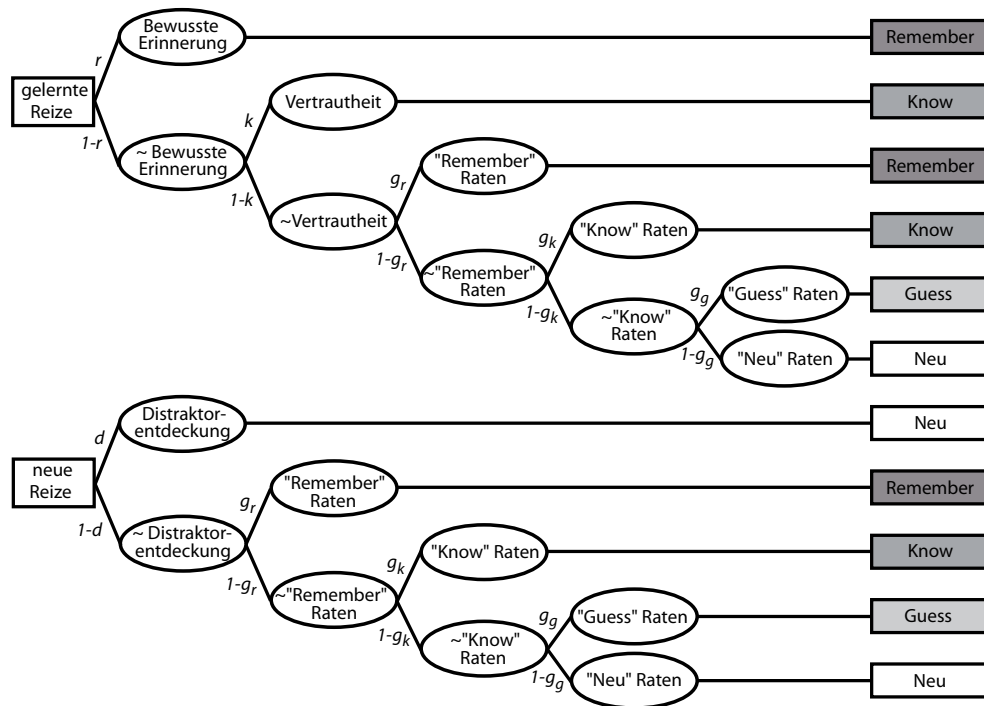


Abbildung 4.3: Dargestellt wird das Four-States-Modell nach Erdfelder et al. (2007). Die beiden Entscheidungsbäume veranschaulichen die Annahmen darüber, wie für alte bzw. neue Reize Remember-, Know-, Guess- und Neu-Antworten zustandekommen. Durch Ovale werden latente kognitive Zustände gekennzeichnet. Die an den Ästen vermerkten Parameter stehen für die bedingten Übergangswahrscheinlichkeiten (r = Wahrscheinlichkeit, bewusster Erinnerung, k = Wahrscheinlichkeit, eines Vertrautheitsabrufes im Falle fehlender bewusster Erinnerung, g_R = Wahrscheinlichkeit, aus einem Zustand der Unsicherheit R zu raten, g_K = Wahrscheinlichkeit, aus einem Zustand der Unsicherheit K zu raten, wenn vorher nicht R geraten wurden, g_G = Wahrscheinlichkeit, aus einem Zustand der Unsicherheit G zu raten, wenn vorher weder R noch K geraten wurde, d = Wahrscheinlichkeit, einen Distraktor zu entdecken).

die Wahrscheinlichkeit einer Guess-Antwort, wie in Gleichung 4.15 zu sehen ist, in einem einzigen Produkt wiedergeben; dieses entspricht der Wahrscheinlichkeit dafür, dass sich Probanden im Zustand der Unsicherheit befinden und weder Remember noch Know raten, aber Guessing berichten $((1 - r) \cdot (1 - k) \cdot (1 - g_R) \cdot (1 - g_K) \cdot g_G)$. Die Wahrscheinlichkeit dafür, dass auf einen alten Reiz mit einer Neu-Antwort reagiert wird, gibt der letzte Ast wieder. Dies geschieht demnach mit der Wahrscheinlichkeit $(1 - r) \cdot (1 - k) \cdot (1 - g_R) \cdot (1 - g_K) \cdot (1 - g_G)$, also der Gegenwahrscheinlichkeit zur Summe aller vorgenannten Astwahrscheinlichkeiten.

Die Antworthäufigkeiten für Distraktoren können den Gleichungen 4.15 bis 4.18 entnommen werden. Distraktoren werden mit der Wahrscheinlichkeit d sicher als neu erkannt. Geschieht dies nicht, befinden sich die Probanden in einem Zustand der Unsicherheit $(1 - d)$. Aus diesem Zustand der Unsicherheit heraus können sie wie bei alten Reizen Remember, Know oder Guess raten. Geschieht auch dies nicht, wird eine Neu-Antwort gegeben.

Mit Hilfe des 4-States-Modells lassen sich die acht Antwortwahrscheinlichkeiten für diesen einfachen Fall anhand von 6 Zustandsparametern erklären, wovon drei spezifische Antworttendenzen abbilden – g_R , g_K und g_G –, während die anderen drei – r , k sowie d – dazu benutzt werden, die Sensitivität des Gedächtnisapparates zu erfassen. Um komplexere Datenmuster analysieren zu können, muss dieses Ursprungsmodell – ebenso wie das SDM-Modell – lediglich um die dazu nötigen Bäume und gegebenenfalls Modellparameter erweitert werden.

4.3 Sum-difference-Theory-of-Remembering-and-Knowing

Mit der Sum-Difference-Theory-of-Remembering-and-Knowing (STREAK) haben Rotello et al. (2004) ein zweidimensionales Signalentdeckungsmodell zur Erklärung von Ergebnissen aus Remember-Know-Experimenten vorgeschlagen. Um diese Theorie auch auf Remember-Know-Guess-Daten anwenden zu können, müssen die Annahmen dieses Modells noch eine entsprechende Ergänzung erfahren. Diese Anpassung des Modells auf die Remember-Know-Guess-Prozedur wird nach der Darstellung des STREAK-Modells für die einfache Remember-Know-Prozedur erläutert werden.

Die formale Struktur des STREAK-Modells erinnert stark an mehrdimensionale Klassifikationsmodelle, wie den ‘decision bound’-Ansatz von Ashby (1992) sowie Ashby und Townsend (1986). Die grundlegenden Annahmen, auf denen das STREAK-Modell beruht, lassen sich wie folgt zusammenfassen:

1. Die Spurstärke eines Items variiert auf zwei Dimensionen. Diese beiden Dimensionen sind erstens die globale Vertrautheit des Testitems und zweitens die spezifische Spurstärke von Details, die mit den Testitems in Beziehung stehen. Die globale Vertrautheit eines Reizes entspricht in etwa der im Rahmen des SDM-Modells vorgestellten Vertrautheit. Das Konzept der spezifischen Spurstärke hingegen hat anderen Ursprung. Hierbei wird angenommen, dass

auch die Verfügbarkeit spezifischer Information auf einer kontinuierlichen Dimension abgebildet werden kann und nicht nach dem Alles-oder-Nichts-Prinzip entweder vorliegt oder abwesend ist. Diese Annahme stimmt mit der Idee differenzierter Information überein, wie sie für das Quellenüberwachungsparadigma formuliert worden ist (Johnson, Hashtroudi & Lindsay, 1993; Mitchell & Johnson, 2000). Demnach beinhalten Erinnerungen Informationen über qualitative Eigenschaften der Enkodierungssituation, wie z. B. die persönliche Stimmung, den sozialen Kontext, in dem ein bestimmtes Ereignis erlebt wurde, oder aber auch die Schriftgröße oder die Schriftart, mit der ein Wort in einem Rekognitionsexperiment präsentiert worden ist. Die Verfügbarkeit solcher spezifischer Informationen kann ebenso wie die Vertrautheit des Reizes von Item zu Item variieren. Selbst innerhalb der Klasse der Distraktoren kann der eine Reiz enger mit manchen qualitativen Eigenschaften der Lernphase in Zusammenhang stehen als ein anderer.

2. Gelernte Items und Distraktoren unterscheiden sich sowohl auf der Dimension der globalen Vertrautheit als auch auf der Dimension der spezifischen Spurstärke: Gelernte Reize sind in der Regel vertrauter als neue Reize und aktivieren zudem mit höherer Wahrscheinlichkeit spezifische Informationen. Zudem variieren sowohl globale Vertrautheit als auch spezifische Spurstärke sowohl innerhalb der gelernten Items als auch innerhalb der Distraktoren. Rotello et al. (2004) gehen davon aus, dass sich diese Variation am besten durch eine Normalverteilung beschreiben lässt. Demnach ist die Entscheidungssituation durch die Annahme zweier bivariater Normalverteilungen charakterisiert. Die Verteilung, die die Spurstärke der alten Stimuli darstellt, hat den Schwerpunkt in dem Punkt $(d_x|d_y)$, während die Verteilung für Distraktoren um den Nullpunkt des kartesischen Koordinatensystems $(0|0)$ liegt. Die Werte auf der x-Achse dieses zweidimensionalen Koordinatensystems, zu denen auch d_x gehört, repräsentieren die globale Spurstärke, während die spezifische Spurstärke auf der y-Achse abgetragen wird. Dadurch, dass der Erwartungswert der Distraktorverteilung auf den Koordinatenursprung gelegt wird, lassen sich d_x und d_y auch interpretieren als den relativen Vorteil, den gelernte Reize auf der jeweiligen Dimension gegenüber den neuen Stimuli haben.
3. Die beiden Normalverteilungen sollen sich jedoch nicht nur in der Lage unterscheiden: Zusätzlich nehmen die Autoren an, dass sich die Varianzen der beiden Verteilungen insofern deutlich voneinander unterscheiden, als die Varianz der Distraktorverteilung kleiner ist als die Varianz der Verteilung für gelernte Reize. Aus Gründen der Praktikabilität fixierten die Autoren beide Varianzen, wobei sie die Varianz für gelernte Stimuli auf $\sigma_{Alt} = 1$ normierten und für die Spurstärkeverteilung der Distraktoren den Wert $\sigma_{Neu} = 0.8$ annahmen.
4. In Remember-Know-Untersuchungen setzen die Probanden zwei Kriterien, C_O und C_R . Das Kriterium C_O dient dazu, neue Reize von gelernten Reizen zu unterscheiden, während C_R benutzt wird, um für die mutmaßlich alten Stimuli zwi-

schen Remembering und Knowing zu differenzieren. Aus einer Reihe möglicher Alternativen entschieden sich Rotello et al. (2004) für lineare Entscheidungskriterien, die sich zueinander orthogonal verhalten.

Das C_O -Kriterium wird als eine Gerade mit der Steigung $-d_y/d_x$ und einem variablen y-Achsenabschnitt, der die Konservativität bzw. Liberalität des angelegten Kriteriums wiedergibt, konzipiert. Alle Items, die oberhalb dieser Geraden liegen, werden mit Alt-Antworten belegt, während auf Items, deren Spurstärken diese Gerade nicht überschreiten, mit Neu reagiert wird. Dieses Konzept zeichnet sich durch zwei Eigenschaften aus, die man unter den Schlagworten der 'Kompensation' und der 'Abhängigkeit vom Erwartungswert' fassen kann.

Unter Kompensation ist zu verstehen, dass die relative Stärke der globalen bzw. spezifischen Spurstärke das Fehlen von Evidenz auf der jeweils anderen Spurstärkedimension ausgleichen kann. Dies wird erreicht, indem die Evidenz beider Dimensionen aufsummiert wird, bevor ein Vergleich mit dem Alt-Neu-Kriterium vorgenommen wird. Die zweite der genannten Eigenschaften besteht darin, dass die relative Position des C_O -Kriteriums auf den Spurstärkedimensionen von dem Erwartungswert der Alt-Verteilung auf diesen Dimensionen abhängt: Je höher die mittlere Spurstärke auf einer der beiden Achsen relativ zu der mittleren Spurstärke auf der anderen Achse liegt, desto höher fällt auch der Spurstärkebetrag aus, der auf dieser Dimension erreicht werden muss, damit Alt-Urteile gefällt werden.

Für das Remember-Know-Kriterium C_R gilt die eben angesprochene kompensatorische Verknüpfung nicht. Dadurch, dass dieses Kriterium senkrecht auf dem C_O -Kriterium steht, ist es vielmehr sensitiv für die Differenzen zwischen spezifischer und globaler Spurstärke. Das Kriterium C_R unterteilt die Items dadurch in solche, die sich durch eine relativ hohe spezifische Spurstärke auszeichnen, und solche, die sich durch eine relativ hohe globale Spurstärke auszeichnen. Auf diesem Kontrast beruht nach Rotello et al. (2004) die Unterteilung alter Reize in Items, an deren Enkodierumstände man sich bewusst erinnert und Items, von denen man nur weiß, dass sie in der Lernphase präsentiert worden sind. Eine zentrale Annahme des Modells lautet, dass Remember-Antworten umso wahrscheinlicher werden, je höher dieser Kontrast zugunsten der spezifischen Spurstärke ausfällt. Know-Antworten resultieren hingegen, wenn die spezifische Spurstärke relativ zur globalen niedrig ausfällt. Aus diesen Annahmen folgt, dass ein Urteil über das Abruferleben umso schwerer fällt, je ausgeglichener das Verhältnis dieser beiden Spurstärkeindizes für ein Item ist.

Nachdem nun das von Rotello et al. (2004) für die Remember-Know-Prozedur vorgeschlagene Modell vorgestellt worden ist, soll dessen Anpassung auf das Remember-Know-Guess-Paradigma erörtert werden. Dazu ist es erforderlich, die Annahmen bezüglich der von den Probanden gesetzten Kriterien zu erweitern. Durch die Aufnahme einer dritten Kriteriumsgeraden, die zwischen solchen Alt-Antworten differenziert, die als Remembering bzw. Knowing erlebt wurden, und solchen, die die Versuchspersonen als geraten berichten, gelingt die erwünschte Anpassung an das Remember-

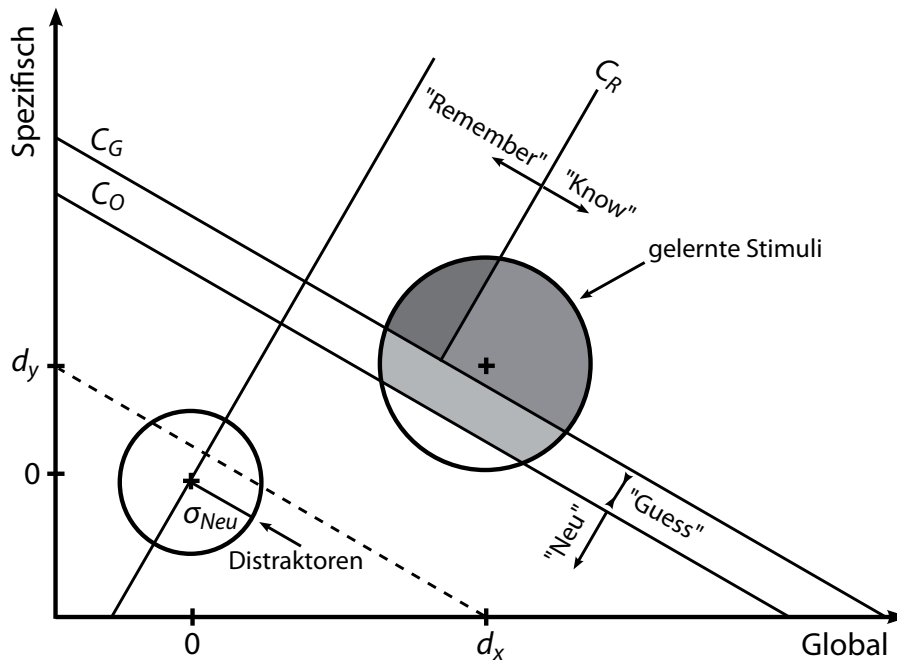


Abbildung 4.4: Dargestellt wird die an die Remember-Know-Guess-Prozedur angepasste Version des STREAK-Modells. Gelernte Reize und Distraktoren unterscheiden sich in ihrer globalen Aktivierung um d_x , im Grad ihrer spezifischen Aktivierung um d_y sowie in ihrer Varianz, wobei die Varianz der Spurstärkeverteilung für Distraktoren σ_{Neu} beträgt. Zwischen Reizen, die für Remember-, Know-, Guess- bzw. Neu-Antworten in Frage kommen, wird anhand dreier Entscheidungskriterien entschieden: Wenn C_O überschritten wird, resultieren Alt-Antworten. Liegt die Aktivierung über C_O , bleibt aber unter C_G , so werden Guess-Antworten gegeben. Wird C_G hingegen überschritten und liegt die Spurstärke links von C_R , so wird Remember geantwortet, liegt die Spurstärke rechts von C_R , so resultieren Know-Antworten.

Know-Guess-Paradigma. Das einzuführende Kriterium scheint sinnvoll konzipiert als eine Gerade, die oberhalb des C_O -Kriteriums und parallel zu selbigem verläuft, so dass Guess-Antworten als Alt-Antworten zu verstehen sind, deren summierte Spurstärke besonders niedrig ausfällt. Eine solche Annahme würde den Befunden von Gardiner et al. (2002) entgegenkommen, denen zufolge die Sensitivität von Guess-Antworten im Bereich der Ratewahrscheinlichkeit liegt.

Mit der Entwicklung eines auf dem STREAK-Modell basierenden Remember-Know-Guess-Modells geht eine Veränderung der Datensituation einher, die es erlaubt, die willkürliche Festsetzung der Varianz neuer Reize auf 0.8 aufzuheben und an deren Stelle einen zu schätzenden Parameter aufzunehmen. So enthält die Abbildung 4.4 neben den bereits besprochenen Parametern d_x , d_y , C_O , C_R und C_G den Parameter σ_{Neu} . Dieser Parameter setzt die Varianz der Distraktorverteilung in Relation zu der Varianz der Alt-Verteilung, die weiterhin auf 1 normiert bleibt.

Anhand von Abbildung 4.4 lassen sich die in den Gleichungen 4.19 bis 4.26 dargelegten Modellgleichungen besser verstehen, mit denen im STREAK-Ansatz Antworthäufigkeiten bei Distraktoren sowie gelernten Reizen für die Kategorien, Remember, Know, Guess und Neu modelliert werden. In der grafischen Darstellung werden die Spurstärkeverteilungen für Distraktoren und gelernte Reize durch Kreise angedeutet. Diese Verteilungen enden selbstverständlich nicht an den Grenzen der jeweiligen Kreise, sondern erfassen prinzipiell die gesamte Ebene. Die Mittelpunkte der Kreise bilden die Erwartungswerte der Verteilungen $(0|0)$ respektive $(d_x|d_y)$. Durch den Radius der Kreise wird die Varianz der Spurstärkeverteilungen dargestellt: Je geringer die Varianz ausfällt, desto kleiner ist der Kreis, der um den Erwartungswert gezogen wird. Während die Varianz der Alt-Verteilung konstant ist, wird – wie oben angedeutet – für die Distraktorverteilung angenommen, dass ihre Varianz σ_{Neu} von Datensatz zu Datensatz variieren kann.

Neben den Spurstärkeverteilungen enthält die Abbildung 4.4 insgesamt drei Geraden, die die Urteilskriterien für die Unterscheidung zwischen Alt und Neu (C_O), Raten respektive Erinnern (C_G) sowie Remember versus Know (C_R) darstellen. Die Lage des Kriteriums C_O wurde wie auch die Lage der übrigen Kriterien in der Abbildung 4.4 willkürlich gewählt. Abhängig von der Strenge des Kriteriums kann die Urteilsgerade parallel zu der eingezeichneten nach oben oder unten verschoben werden. Unabhängig von der exakten Lage dieser Urteilsgeraden mit der Steigung $-d_y/d_x$ entspricht das Volumen, das unter beiden Spurstärkeverteilungen links dieses Kriteriums liegt, der Wahrscheinlichkeit für Neu-Antworten. Die Fläche, die oberhalb des Alt-Neu-Kriteriums liegt, aber unterhalb des C_G -Kriteriums zu finden ist, repräsentiert die Wahrscheinlichkeit von Guess-Antworten. Der Bereich oberhalb des Guess-Kriterium wird durch das C_R -Kriterium in zwei Teile zerlegt. Links dieser Entscheidungsgeraden mit der Steigung d_x/d_y finden sich jene Items wieder, die verglichen mit ihrer globalen Spurstärke eine höhere spezifische Spurstärke aufweisen. Dieser Teil der bivariaten Spurstärkeverteilungen entspricht dementsprechend der Wahrscheinlichkeit von Remember-Antworten, während der Teil, der rechts unter diesem Kriterium liegt, die Wahrscheinlichkeit von Know-Antworten wiedergibt.

Um aus diesem Gerüst die Modellgleichungen herzuleiten, ist es hilfreich, Pro-

jektionen vorzunehmen, bei denen die Spurstärkeverteilungen – also entweder die Distraktorverteilung oder die Alt-Verteilung – auf einer Entscheidungsachse abgebildet werden, auf der sich die Entscheidungskriterien linear verschieben lassen. Für die Alt-Neu-Entscheidung und das parallele Kriterium C_G wäre die ideale Projektionsachse die Ursprungsgerade mit der Steigung d_x/d_y , auf der sowohl C_G als auch C_O lotgerecht stehen. Die Distraktorverteilung liegt bereits auf dieser Achse, für die Alt-Verteilung muss hingegen eine Transformation vorgenommen werden. Die Projektion des Mittelpunktes der Alt-Verteilung auf diese Entscheidungsachse entspricht dem Wert $2d_x d_y / \sqrt{d_x^2 + d_y^2}$ (vgl. Rotello et al., 2004, App. B). Demnach kann die Wahrscheinlichkeit aller korrekten Alt-Antworten modelliert werden durch $p(Alt|Alt) = \Phi(2d_x d_y / \sqrt{d_x^2 + d_y^2} - C_O)$, während sich die Wahrscheinlichkeit für Falsche Alarmer durch die Gleichung $p(Alt|Neu) = \Phi(-C_O / \sigma_{Neu})$ vorhersagen lässt.

$$p(R|Alt) = \Phi\left(\frac{2d_x d_y}{\sqrt{d_x^2 + d_y^2}} - C_G\right) \Phi(C_R) \quad (4.19)$$

$$p(K|Alt) = \Phi\left(\frac{2d_x d_y}{\sqrt{d_x^2 + d_y^2}} - C_G\right) \Phi(-C_R) \quad (4.20)$$

$$p(G|Alt) = \Phi\left(\frac{2d_x d_y}{\sqrt{d_x^2 + d_y^2}} - C_O\right) - \Phi\left(\frac{2d_x d_y}{\sqrt{d_x^2 + d_y^2}} - C_G\right) \quad (4.21)$$

$$p(N|Alt) = \Phi\left(C_O - \frac{2d_x d_y}{\sqrt{d_x^2 + d_y^2}}\right) \quad (4.22)$$

$$p(R|Neu) = \Phi(-C_G / \sigma_{Neu}) \Phi\left(\frac{1}{\sigma_{Neu}} \cdot \left(\sqrt{\frac{d_x^2 - d_y^2}{d_x^2 + d_y^2}} + C_R\right)\right) \quad (4.23)$$

$$p(K|Neu) = \Phi(-C_G / \sigma_{Neu}) \left(1 - \Phi\left(\frac{1}{\sigma_{Neu}} \cdot \left(\sqrt{\frac{d_x^2 - d_y^2}{d_x^2 + d_y^2}} + C_R\right)\right)\right) \quad (4.24)$$

$$p(G|Neu) = \Phi(-C_O / \sigma_{Neu}) - \Phi(-C_G / \sigma_{Neu}) \quad (4.25)$$

$$p(N|Neu) = \Phi(C_O / \sigma_{Neu}) \quad (4.26)$$

Während eine Projektion der Alt-Verteilung bei der Alt-Neu- sowie Erinnert-Geraten-Entscheidung notwendig ist, muss die Distraktorverteilung für das Remember-Know-Urteil auf die entsprechende Achse projiziert werden. Die für dieses Urteil relevante Entscheidungsachse verläuft mit der Steigung $-\frac{d_y}{d_x}$ durch den Punkt $(d_x|d_y)$, also den Mittelpunkt der Alt-Verteilung. Die Projektion des Mittelwerts der Distraktorverteilung auf diese Entscheidungsachse entspricht dem Wert $(d_x^2 - d_y^2) / \sqrt{d_x^2 + d_y^2}$ (vgl. Rotello et al., 2004, App. B). Da Remember-Treffer ebenso wie Remember-Falsche Alarmer in dieser Modellfassung als Reize konzipiert werden, deren Spurstärke das Guess-Kriterium überschreitet und die zudem links des Remember-Kriteriums zu finden sind, lässt sich ihre Wahrscheinlichkeit als das Produkt aus zwei Einzelwahrscheinlichkeiten wiedergeben. Multipliziert wird dazu, wie man den Gleichungen 4.19 und 4.23 entnehmen kann, die Wahrscheinlichkeit eines Überschreitens des Guess-Kriteriums mit der Wahrscheinlichkeit dafür, dass Distraktoren oder gelernte Reize

links des Remember-Kriteriums wieder gefunden werden können. In analoger Weise entspricht die Wahrscheinlichkeit von Know-Treffern sowie Know-Falschen-Alarmen dem Produkt der einzelnen Wahrscheinlichkeiten dafür, dass das Kriterium C_G überschritten wird, sowie dafür, dass die Spurstärke unterhalb des Remember-Kriteriums liegt. Guess-Treffer und Guess-Falsche-Alarme ergeben sich als die Wahrscheinlichkeit dafür, dass zwar das Alt-Kriterium nicht aber das C_G -Kriterium überschritten wird, während sich Neu-Antworten ergeben als die Wahrscheinlichkeit dafür, dass die Spurstärke unterhalb des C_O -Kriteriums bleibt.

Es lässt sich festhalten, dass das STREAK-Modell Ergebnisse von Remember-Know-Guess-Untersuchungen anhand zweier Gedächtniskomponenten erklärt: Die erste Komponente ist die globale Vertrautheit von Reizen und die zweite beruht auf der spezifischen Spurstärke von Detailinformation. Während für die Alt-Neu-Entscheidung die Summe dieser beiden Spurstärken ausschlaggebend ist, wird für die Remember-Know-Entscheidung die spezifische Spurstärke mit der globalen Vertrautheit durch das Bilden einer Differenz kontrastiert: Überwiegt die spezifische Spurstärke, so resultieren Remember-Antworten, fällt der Kontrast hingegen zugunsten der globalen Spurstärke aus, dann geben die Probanden Know-Urteile ab.

4.4 Das Source-of-Activation-Confusion-Modell

Das nächste Modell, das für diese Arbeit von Bedeutung ist, haben Reder und Mitarbeiter (Diana, Peterson & Reder, 2004; Reder, Angstadt, Cary, Erickson & Ayers, 2002; Reder et al., 2000) vorgeschlagen und Source-Of-Activation-Confusion-(SAC-)Modell genannt. Der Ursprung dieses Rahmenmodells liegt in Andersons *ACT-R*-Theorie der menschlichen Kognition (Anderson, Bothell, Lebiere & Matessa, 1998; Anderson & Lebiere, 1998). Reder et al. (2000) beziehen sich auf die mit *ACT-R* verbundene Gedächtnistheorie und bemühen sich, selbige weiter zu konkretisieren.

So wie *ACT-R* eine Umgebung ist, mit deren Hilfe sich Ergebnisse von kognitionspsychologischen Untersuchungen simulieren lassen, wurde das SAC-Modell von Reder und ihren Mitarbeitern zur Simulation von Daten und deren Vergleich mit tatsächlich erhobenen empirischen Daten verwendet. Die theoretische Basis dieser Simulationen liegt in der Annahme eines Netzwerkmodells des Gedächtnisses, in dem Wissen in Informationsknoten abgelegt wird. Diese Knoten sind hierarchisch gegliedert: Es werden basale Knoten angenommen, wie z. B. Buchstabenknoten, die in komplexeren Knoten, wie Wortknoten, zusammenkommen, welche wiederum in noch komplexere Knoten – wie z. B. Satzknöten eingebettet sein können. Für das episodische Gedächtnis besonders relevant ist die Annahme, dass Informationen, wie beispielsweise Items eines Gedächtnisexperimentes, in dem angenommenen Netzwerk nicht nur in einem eigenen Itemknoten repräsentiert sind, sondern zudem gemeinsam mit ihrem Kontext einen Episodenknoten formen.

Zur Erklärung von Remember-Know-Ergebnissen gehen Diana et al. (2004) davon aus, dass Remember-Antworten dann gegeben werden, wenn die Enkodierepisode abgerufen werden kann, d. h. die Aktivierung der Episodenknoten einen Mini-

malwert überschreitet. Für Know-Antworten nehmen sie hingegen an, dass diese eine überschwellige Aktivierung des Itemknotens widerspiegeln. Evidenz für diese Annahmen fanden beispielsweise Cary und Reder (2003). Die Autorinnen zeigten, dass sich mit dem SAC-Modell der Mirror-Effekt, also das Phänomen gegenläufiger Reaktionen von Treffer- und Falscher-Alarm-Rate, nachbilden ließ, der in Remember-Know-Untersuchungen in Folge von Variationen der Listenlänge und Itemstärke beobachtet wurde.

Erst jüngst zeigten Rotello et al. (2004) auf, dass sich die Annahmen Reders und ihrer Mitarbeiter so in Modellannahmen übersetzen lassen, dass das entstehende Modell als Schätz- und Analysewerkzeug genutzt werden kann. Dazu schlagen sie vor, das SAC-Modell als zweidimensionales Signalentdeckungsmodell zu begreifen, das in vielen Bereichen dem STREAK-Modell ähnlich ist. So wird auch im Rahmen des SAC-Modells angenommen, dass sich alte und neue Reize auf zwei Dimensionen unterscheiden. Zudem wird ebenso wie im STREAK-Modell davon ausgegangen, dass die auf diesen Dimensionen erreichten Werte innerhalb der Itemklassen im Sinne einer Normalverteilung variieren. Ein Unterschied zwischen diesen beiden zweidimensionalen Signalentdeckungsmodellen kann darin gesehen werden, dass die Achsen des in Abbildung 4.5 dargestellten Koordinatensystems nicht globale und spezifische Spurstärke wiedergeben, sondern, wie von Reder et al. (2000) vorgeschlagen, die Aktivierung des Episodenknotens auf der y -Achse über der Aktivierung des Itemknotens auf der x -Achse abgetragen wird. Im Modell bleibt es trotz dieses theoretischen Unterschieds bei der Annahme zweier bivariater Normalverteilungen für Distraktoren einerseits und für alte Reize andererseits, die sich in Varianz und Erwartungswert unterscheiden. Wie bei dem zuvor erläuterten STREAK-Modell wird die Varianz der Distraktorverteilung durch σ_{Neu} geschätzt, während die der Alt-Verteilung auf 1 gesetzt wird. Der Mittelpunkt der Distraktorverteilung liegt ebenso wie beim STREAK-Modell im Koordinatenursprung, während der Mittelpunkt der Alt-Verteilung durch die Koordinaten $(d_x|d_y)$ geschätzt wird.

Der bedeutendste Unterschied zwischen dem STREAK-Modell von Rotello et al. (2004) und dem SAC-Modell besteht in der Art der Kriterien, die angenommen werden, um die einzelnen Antworttypen voneinander zu unterscheiden. Während das STREAK-Modell Kriterien enthielt, bei denen es sich um Spurstärken handelte, die kompensatorisch oder im Sinne eines Kontrastes miteinander verrechnet wurden, lassen sich die Kriterien im SAC-Modell am besten als prozessrein beschreiben (vgl. Rotello et al., 2004). Aus Reders Darstellung des SAC-Modells lässt sich für den einfachen Remember-Know-Fall die Annahme ableiten, dass Probanden zwei Kriterien setzen, um Remembering von Knowing zu unterscheiden. Das erste dieser beiden Kriterien (C_R) orientiert sich entlang der y -Achse. Es stellt einen Minimalwert für die Episodenaktivierung dar, der überschritten werden muss, damit sich die Versuchspersonen bewusst an die Enkodierepisode erinnern. Die Probanden geben in diesem Falle Alt-Urteile ab, die sie als Remembering qualifizieren. Liegt die Aktivierung des Episodenknotens zwar nicht über C_R , die Itemaktivierung jedoch über dem Kriterium C_K , so erfolgt eine Alt-Antwort, die als Knowing berichtet wird. In dem Falle, dass beide Aktivierungsarten unter dem jeweils relevanten Kriteriumswert bleiben, antworten die

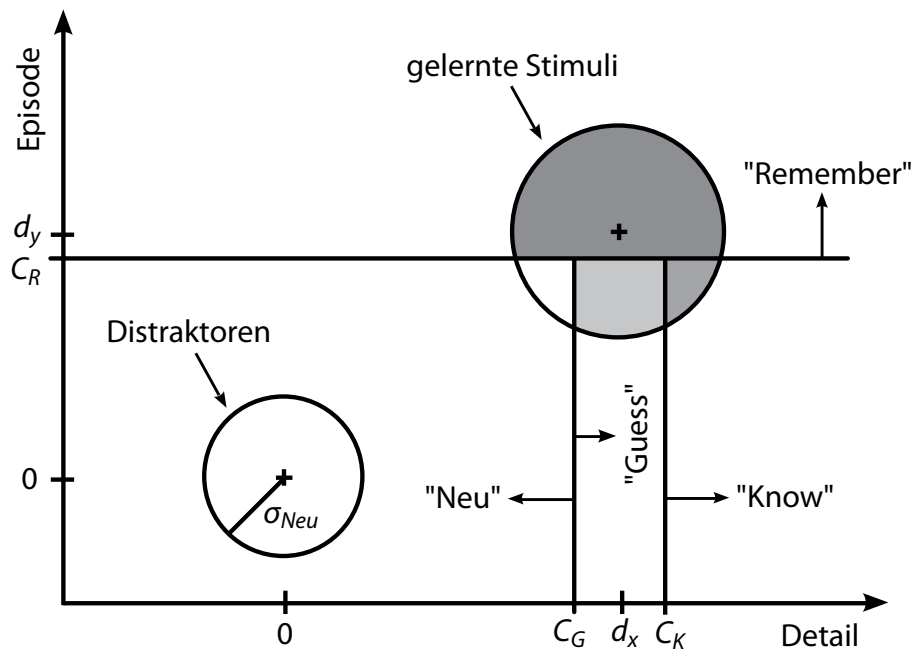


Abbildung 4.5: Dargestellt wird die auf die Remember-Know-Guess-Prozedur abgestimmte Version des SAC-Modells. Gelernte Reize und Distraktoren unterscheiden sich in ihrer Detailaktivierung um d_x , im Grad der Episodenaktivierung um d_y und in der Varianz, wobei die Varianz der Spurstärkeverteilung für Distraktoren σ_{Neu} beträgt. Zwischen Reizen, die für Remember-, Know-, Guess- bzw. Neu-Antworten in Frage kommen, wird anhand dreier Entscheidungskriterien entschieden: Wenn die Detailaktivierung größer als C_G , nicht aber größer als C_K ist und die Episodenaktivierung gleichzeitig unter C_R bleibt, so resultieren Guess-Antworten. Bleibt die Episodenaktivierung unter C_R , während die Detailaktivierung C_K überschreitet, so resultieren Know-Antworten, liegt die Episodenaktivierung über C_R , so folgen Remember-Antworten. Neu-Antworten ergeben sich, wenn weder die Detailaktivierung über C_G liegt noch die Episodenaktivierung größer als C_R ausfällt.

Probanden in Remember-Know-Untersuchungen mit ‘Neu’.

Um im Rahmen des SAC-Modells Ergebnisse von Remember-Know-Guess-Experimenten behandeln zu können, empfiehlt es sich, ein weiteres Kriterium anzunehmen. Dieses Kriterium C_G steht wie C_K senkrecht auf der x-Achse und ist links von letzterem zu finden. Ein Item, dessen Aktivierung dieses Kriterium überschreitet, nicht aber das höhere Kriterium (C_K), wird, wenn die Aktivierung der zugehörigen Enkodierepisode C_R nicht übersteigt, mit einer Guess-Antwort versehen.

Dementsprechend lassen sich die relativen Häufigkeiten von Remember-, Know-, Guess- und Neu-Antworten darstellen als diejenigen Bereiche unter den Verteilungen, die die Bedingungen erfüllen, um mit einem entsprechenden Urteil belegt werden zu können. Die Wahrscheinlichkeit einer Remember-Antwort lässt sich also erfassen durch den Abschnitt der Verteilungen, der oberhalb des Kriteriums C_R liegt. Know-Antworten setzen voraus, dass das Kriterium C_K , nicht aber das Kriterium C_R überschritten wird, so dass sich als relevanter Abschnitt der Verteilungen jener ergibt, der rechts des Know-Kriteriums, aber unterhalb des Remember-Kriteriums liegt. Die Wahrscheinlichkeit von Guess-Antworten wiederum kann in dem Areal unter den Verteilungen wieder gefunden werden, das unter dem C_R -Kriterium und zwischen den beiden anderen Kriterien liegt. Zuletzt resultieren Neu-Antworten dann, wenn weder C_R noch C_G überschritten werden. Dieser Wahrscheinlichkeit entspricht der Bereich unter den Verteilungen, der links des Guess-Kriteriums und unterhalb des Remember-Kriteriums zu finden ist.

Die mathematische Struktur des Modells kann den Gleichungen 4.27 bis 4.34 entnommen werden. Wie auch bei dem zuvor vorgestellten Modell werden zum Vergleich der Evidenzen mit den zugehörigen Kriterien die Verteilungen auf diejenigen Entscheidungsachsen projiziert, auf denen sich die Kriterien bewegen. Bei diesen Achsen handelt es sich – wie bereits erläutert – für die Itemaktivierung um die x-Achse und für die Episodenaktivierung um die y-Achse, so dass nur die Alt-Verteilung auf die jeweils relevante Entscheidungsachse projiziert werden muss. Beim SAC-Modell stellen die Koordinaten des Mittelpunktes der bivariaten Alt-Verteilung die geeigneten Projektionspunkte auf die Entscheidungsachsen dar. Ist es wie im Falle der Know-Antworten erforderlich, die Lage der Verteilungen zu Kriterien auf beiden Dimensionen in Beziehung zu setzen, so geschieht dies durch die multiplikative Verknüpfung der Einzelwahrscheinlichkeiten. Am Beispiel der Know-Antworten handelt es sich hierbei, wie in Gleichung 4.28 und 4.32 dargelegt, erstens um die Wahrscheinlichkeit dafür, dass die Episoden-Aktivierung unter dem für eine Remember-Antwort notwendigen Kriterium C_R bleibt, und zweitens um die Wahrscheinlichkeit dafür, dass die Item-Aktivierung das Kriterium C_K überschreitet. Die Wahrscheinlichkeit von Guess-Antworten lässt sich ermitteln, indem die erste dieser beiden Wahrscheinlichkeiten mit der Wahrscheinlichkeit dafür multipliziert wird, dass zwar das Guess-Kriterium, nicht aber das Know-Kriterium überschritten wird, wie den Gleichungen 4.29 bzw. 4.33 entnommen werden kann. Für die Ermittlung des Remember-Anteils an den Antworten für alte und neue Reize ist, wie in den Gleichungen 4.27 und 4.31 dargestellt, lediglich die Wahrscheinlichkeit entscheidend, mit der die Episoden-Aktivierung der Items das C_R -Kriterium überschreitet. Die Wahrscheinlichkeit von Neu-Antworten ergibt sich aus der Wahr-

scheinlichkeit, mit der keines der Kriterien überschritten wird (vgl. Gleichungen 4.30 und 4.34).

Ebenso wie das STREAK-Modell ist das SAC-Modell ein Signalentdeckungsmodell, das auf der Annahme fußt, dass für Remembering und Knowing zwei verschiedene Evidenzarten entscheidend sind. Anders als Rotello et al. (2004), die von Kompensation und Kontrast der Spurstärken ausgehen, wird im Rahmen des SAC-Modells jedoch angenommen, dass Remembering und Knowing den Grad der Episoden-Aktivierung bzw. die Stärke der Item-Aktivierung prozessrein erfassen.

$$p(R|Alt) = \Phi(d_y - C_R) \quad (4.27)$$

$$p(K|Alt) = \Phi(d_x - C_K) \cdot \Phi(C_R - d_y) \quad (4.28)$$

$$p(G|Alt) = (\Phi(d_x - C_G) - \Phi(d_x - C_K)) \cdot \Phi(C_R - d_y) \quad (4.29)$$

$$p(N|Alt) = \Phi(C_G - d_x) \cdot \Phi(C_R - d_y) \quad (4.30)$$

$$p(R|Neu) = \Phi\left(-\frac{C_R}{\sigma_{Neu}}\right) \quad (4.31)$$

$$p(K|Neu) = \Phi\left(-\frac{C_K}{\sigma_{Neu}}\right) \cdot \Phi\left(\frac{C_R}{\sigma_{Neu}}\right) \quad (4.32)$$

$$p(G|Neu) = \left(\Phi\left(-\frac{C_G}{\sigma_{Neu}}\right) - \Phi\left(-\frac{C_K}{\sigma_{Neu}}\right)\right) \cdot \Phi\left(\frac{C_R}{\sigma_{Neu}}\right) \quad (4.33)$$

$$p(N|Neu) = \Phi\left(\frac{C_G}{\sigma_{Neu}}\right) \cdot \Phi\left(\frac{C_R}{\sigma_{Neu}}\right) \quad (4.34)$$

4.5 Das Hochschwellen-Signalentdeckungsmodell

Auf der Basis des von Yonelinas, Dobbins, Szymanski, Dhaliwal und King (1996) vorgeschlagenen *dual-process model* (DPM) kann ein hybrides Modell zur Remember-Know-Guess-Prozedur entwickelt werden, das eine der Gedächtniskomponenten signalentdeckungstheoretisch auffasst und einen zweiten Gedächtnisprozess annimmt, der erfolgreich ist, wenn alte Items eine Schwelle überschreiten. Das DPM von Yonelinas et al. (1996) zeichnet sich dadurch aus, dass es dem Versuch entstammt, verwandte Forschungsfelder mit demselben Modell zu erklären. So versuchten Yonelinas und Jacoby (1994) mit Hilfe des DPMs sowohl Remember-Know-Befunde als auch Befunde, die im Paradigma der Prozessdissoziation (Jacoby, 1991) erfasst worden sind, in einem einheitlichen Rahmen verständlich zu machen. Um eine solche gemeinsame Erklärung leisten zu können, wurde angenommen, dass sich Remember- und Know-Antworten auf die Aktivität kontrollierter Gedächtnisprozesse einerseits und automatischer Verarbeitungsprozesse andererseits zurückführen lassen. Darüber hinaus gingen Yonelinas et al. (1996) davon aus, dass sich kontrollierte Abrufprozesse am besten durch eine Schwellenannahme beschreiben lassen, während automatische Prozesse als kontinuierliches Konstrukt, über dem sich neue wie alte Reize normalverteilen, beschrieben

werden sollten. Hinter diesem Vorschlag steckt die Idee, dass die Ergebnisse automatischer Prozesse, wie sie im DPM-Modell angenommen werden, eng verwandt sind mit dem Konstrukt der globalen Vertrautheit, wie sie im Rahmen des SDM-Modells angenommen wird.

Die Struktur des DPM-Modells lässt sich beschreiben als ein Hochschwellenmodell, innerhalb dessen davon ausgegangen wird, dass Probanden informiert raten, wenn sie sich nicht bewusst an das Vorkommen eines Reizes erinnern können. Für das informierte Raten von essentieller Bedeutung ist ein Vertrautheitskontinuum, auf dem sich alle Reize anordnen und miteinander vergleichen lassen, indem ein Kriterium angelegt wird, das hilft, zwischen alten Reizen und Distraktoren zu unterscheiden. Demnach gibt es zwei Wege, wie Probanden zu korrekten Alt-Antworten gelangen können: Erstens kann es sein, dass sie sich mit der Wahrscheinlichkeit r an Items rekollktiv erinnern, und zweitens ist es möglich, dass Items im Falle des Ausbleibens einer bewussten Erinnerung $(1 - r)$ eine Vertrautheit aufweisen, die über dem Alt-Kriterium C_O liegt. Entsprechend ergibt sich die Treffer-Wahrscheinlichkeit $p(Alt|Alt)$ als Summe der Wahrscheinlichkeiten der beiden genannten Ereignisse, d. h. $p(Alt|Alt) = r + (1 - r) \cdot \Phi(d' - C_O)$. Die Wahrscheinlichkeit eines Falschen Alarms lässt sich hingegen mit der Gleichung $p(Alt|Neu) = \Phi(-C_O)$ als die Wahrscheinlichkeit beschreiben, mit der ein Distraktor das Vertrautheitskriterium für Alt-Antworten überschreitet. Im Rahmen des DPM-Modells wird also weder davon ausgegangen, dass es eine Distraktorentdeckungsschwelle gibt, wie sie im 4-States-Modell angenommen wird, noch beinhaltet das Modell einen Parameter, anhand dessen sich Unterschiede in den Varianzen der Vertrautheitsverteilungen für alte Reize und Distraktoren erfassen ließen, wie es im SDM-, STREAK- oder SAC-Modell der Fall ist.

Als Modell für die Remember-Know-Prozedur ist das DPM-Modell insofern interessant, als Yonelinas et al. (1996) belegten, dass sich anhand der Häufigkeiten von Remember- und Know-Antworten Indizes für rekollktive Erinnerung und globale Vertrautheit berechnen ließen, die dazu genutzt werden konnten, Ergebnisse anderer Untersuchungen vorherzusagen. Ein Messmodell, das mit Hilfe von Modellgleichungen die Antworthäufigkeiten in Remember-Know- oder Remember-Know-Guess-Untersuchungen vorherzusagen vermag, liegt bis dato nicht vor, kann auf den Annahmen des DPM-Modells aufbauend aber leicht entwickelt werden. Dieses Modell soll nun unter dem Namen Hochschwellen-Signalentdeckungsmodell (HTSD-Modell) vorgestellt werden.

Im HTSD-Modell sollen die Kernannahmen des DPM-Modells von Yonelinas et al. (1996) dazu verwendet werden, die Antworthäufigkeiten in Remember-Know- und Remember-Know-Guess-Untersuchungen vorherzusagen und zu erklären. Dazu werden zwei zentrale Prozesse angenommen: Zum einen rekollktive Erinnerung, die als Schwellenprozess konzipiert wird, der mit der Wahrscheinlichkeit r erfolgreich ist, und zum anderen Vertrautheit, die signalentdeckungstheoretisch konzipiert wird. Für die Vertrautheit wird davon ausgegangen, dass sich diejenige alter Reize normal um den Mittelwert d' verteilt, während sich die Vertrautheit der Distraktoren um den Nullpunkt normalverteilt. Zudem wird angenommen, dass Probanden bei den Entscheidungen, für die sie auf die Vertrautheit der Reize zurückgreifen, drei Kriterien berücksichtigen:

1. Das Kriterium C_O muss von der Vertrautheit der Items überschritten werden, damit Alt-Antworten gegeben werden. Bleibt die Vertrautheit hinter dem Wert dieses Kriteriums zurück, so antworten die Probanden mit Neu;
2. Die Vertrautheit der Items, die mit Guess-Antworten versehen werden, liegt dem HTSD-Modell zufolge zwar über dem Alt-Kriterium (C_O), unterschreitet jedoch das Know-Kriterium (C_K);
3. Überschreitet die Vertrautheit des Items das Know-Kriterium (C_K), ist entscheidend, wie sich die vorliegende Vertrautheit im Vergleich zu dem dritten Kriterium (C_R) verhält: Liegt die Vertrautheit über C_K , nicht aber über C_R , dann resultieren Know-Antworten, werden hingegen beide Kriterien überschritten, geben die Probanden Remember-Antworten.

Die Aufnahme des letztgenannten Kriteriums (C_R) bringt zwar eine Abkehr von der prozessreinen Erfassung rekollektiver Prozesse durch Remember-Häufigkeiten, wie sie Yonelinas et al. (1996) vorgesehen haben, mit sich, ermöglicht dafür aber eine adäquate Modellierung des Zustandekommens von Remember-Falschen-Alarmen. Im Rahmen des DPM-Modells werden falsche Remember-Antworten nur insofern thematisiert, als Yonelinas und Mitarbeiter davon ausgehen, dass sie Ergebnis uninformierten Ratens sind. Diese Annahme scheint, wie Dunn (2004) herausgearbeitet hat, allerdings irreführend zu sein. Dunn (2004) argumentiert, dass Remember-Falsche-Alarme als trotz ihrer Neuheit hoch vertraute Distraktoren betrachtet werden sollten. Dieser Maxime folgt das HTSD-Modell, wie es in Abbildung 4.6 dargestellt wird.

Im unteren Segment der Abbildung 4.6 wird das Zustandekommen von Remember-, Know-, Guess- und Neu-Antworten für Distraktoren dargestellt, während im oberen das Zustandekommen selbiger Antworten für gelernte Reize skizziert wird. Für die gelernten Reize ist aus der Abbildung ersichtlich, dass eine signalentdeckungstheoretische Komponente in den Rahmen eines Schwellen-Modells eingebettet wird: Rekollektive Erinnerung führt hier mit der Wahrscheinlichkeit r direkt zu einer Remember-Antwort, während beim Ausbleiben einer rekollektiven Erinnerung die Vertrautheit der Items entscheidend dafür ist, ob Remember-, Know-, Guess- oder Neu-Antworten gegeben werden. Der untere Teil der Abbildung zeigt, dass die Wahrscheinlichkeit, mit der für die Distraktoren eine der vier Antwortalternativen gewählt wird, ausschließlich signalentdeckungstheoretisch modelliert wird. Maßgeblich hierfür ist die Größe der Fläche unter der Distraktorverteilung, die über, zwischen bzw. unter den Kriterien liegt.

Zudem lässt sich das HTSD-Modell über die Gleichungen 4.35 bis 4.42 beschreiben, in die die zuvor dargelegten Annahmen einfließen. Remember-Treffer resultieren – wie Gleichung 4.35 entnommen werden kann – entweder mit der Wahrscheinlichkeit r aufgrund einer rekollektiven Erinnerung oder, falls ein rekollektiver Abruf ausbleibt ($1 - r$), mit einer Wahrscheinlichkeit, die von der erwarteten Vertrautheit alter Stimuli d' und dem Entscheidungskriterium C_R abhängt ($\Phi(d' - C_R)$). Die Wahrscheinlichkeit von Know-Treffern lässt sich durch Gleichung 4.36 modellieren. Ihr liegt die Annahme zugrunde, dass Know-Treffer dann zu Stande kommen, wenn rekollektive

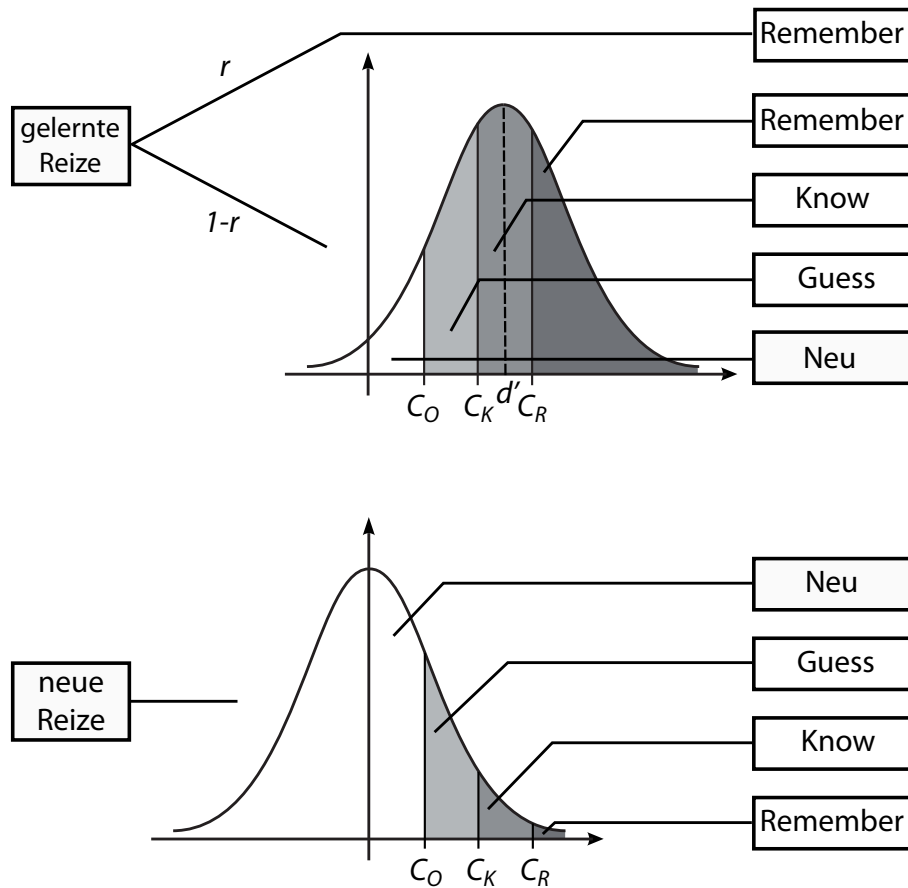


Abbildung 4.6: Abgebildet ist das HTSD-Modell zur Remember-Know-Guess-Prozedur. Mit Wahrscheinlichkeit r führen für gelernte Reize rekollektive Prozesse zu Remember-Antworten. Ist dies nicht der Fall, wird die Vertrautheit der gelernten Reize genau wie bei neuen Reizen mit den Kriterien C_O , C_K und C_R verglichen: Bleibt die Vertrautheit unter C_O , so wird Neu geantwortet, fällt sie zwischen C_O und C_K oder C_K und C_R , so folgen Guess- bzw. Know-Antworten. Übertrifft die Vertrautheit C_R , so resultieren Remember-Antworten. Die Vertrautheit der Reize verteilt sich dabei für beide Reizklassen normal, wobei die Vertrautheit der gelernten Reize im Mittel um d' größer ist als die der Distraktoren.

Prozesse nicht zum Ziel führen ($1 - r$), die Vertrautheit des Items jedoch zwischen den Kriterien C_K und C_R liegt. Letzteres entspricht der Differenz zwischen den beiden Wahrscheinlichkeiten $\Phi(C_R - d')$ und $\Phi(C_K - d')$. Die Wahrscheinlichkeit von Guess-Treffern ergibt sich, wie in Gleichung 4.37 dargestellt, aus dem Produkt von $(1 - r)$ und der Größe der Fläche unter der Alt-Verteilung, die zwischen den Kriterien C_O und C_K liegt. Neu-Antworten auf alte Reize erfolgen dann, wenn weder rekollektive Abrufprozesse erfolgreich sind ($1 - r$), noch die Vertrautheit der Reize über dem Kriterium C_O liegt ($1 - \Phi(d' - C_O)$), wie es in Gleichung 4.38 festgehalten ist. Die Antworthäufigkeiten für neue Reize lassen sich entsprechend anhand der Gleichungen 4.39 bis 4.42 verstehen als die Größe des Bereichs der Distraktorverteilung, der über dem C_R -Kriterium (Remember, Gleichung 4.39), zwischen dem C_K - und dem C_R -Kriterium (Know, Gleichung 4.40), zwischen dem C_O - und dem C_K -Kriterium (Guess, Gleichung 4.41) bzw. unter dem Alt-Kriterium C_O liegt (Neu, Gleichung 4.42).

$$p(R|Alt) = r + (1 - r) \cdot \Phi(d' - C_R) \quad (4.35)$$

$$p(K|Alt) = (1 - r) \cdot (\Phi(d' - C_K) - \Phi(d' - C_R)) \quad (4.36)$$

$$p(G|Alt) = (1 - r) \cdot (\Phi(d' - C_O) - \Phi(d' - C_K)) \quad (4.37)$$

$$p(N|Alt) = (1 - r) \cdot (1 - \Phi(d' - C_O)) \quad (4.38)$$

$$p(R|Neu) = \Phi(-C_R) \quad (4.39)$$

$$p(K|Neu) = \Phi(-C_K) - \Phi(-C_R) \quad (4.40)$$

$$p(G|Neu) = \Phi(-C_O) - \Phi(-C_K) \quad (4.41)$$

$$p(N|Neu) = 1 - \Phi(-C_O) \quad (4.42)$$

4.6 Das Hochschwellen-Signalentdeckungsmodell ungleicher Varianzen

Das letzte zu diskutierende Modell beruht auf einer Erweiterung des HTSD-Modells um die Annahme von Varianzunterschieden zwischen der Vertrautheitsverteilung für alte und für neue Reize. Dieses in Abbildung 4.7 dargestellte Hochschwellen-Signalentdeckungsmodell mit ungleichen Varianzen (HTUVSD) stellt ein gemeinsames Obermodell für das HTSD-Modell und das in Abschnitt 4.1 beschriebene SDM-Modell dar. Dementsprechend enthält das Hochschwellen-Signalentdeckungsmodell mit ungleichen Varianzen neben den Parametern des HTSD-Modells einen Streuungsparameter σ_{Alt} , der die Varianz der Vertrautheitsverteilung für gelernte Reize in Beziehung setzt zu der auf 1 normierten Varianz der Distraktorverteilung.

Da die Annahmen dieses Modells weitgehend mit den Annahmen des vorher dargestellten HTSD-Modell übereinstimmen, kann an dieser Stelle auf die Erläuterung der in den Gleichungen 4.43 bis 4.50 festgehaltenen Modellgleichungen verzichtet werden.

$$p(R|Alt) = r + (1 - r) \cdot \Phi\left(\frac{d' - C_R}{\sigma_{Alt}}\right) \quad (4.43)$$

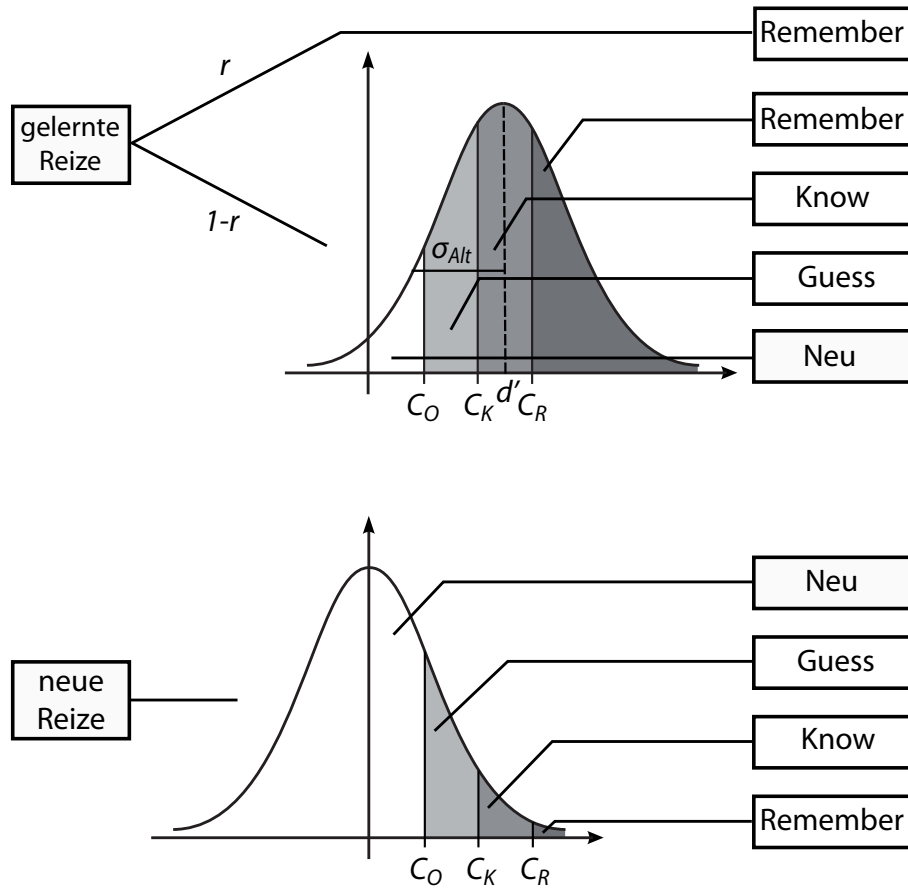


Abbildung 4.7: Dargestellt wird das HTUVSD-Modell zur Remember-Know-Guess-Prozedur. Mit Wahrscheinlichkeit r führen rekollektive Prozesse für gelernte Stimuli zu Remember-Antworten. Wenn dies nicht der Fall ist, wird die Vertraulichkeit der alten Reize genau wie bei neuen Reizen mit den Kriterien C_O , C_K und C_R verglichen: Bleibt die Vertraulichkeit unter C_O , so wird Neu geantwortet, fällt sie zwischen C_O und C_K oder C_K und C_R , folgen Guess- bzw. Know-Antworten. Übertrifft die Vertraulichkeit C_R , so resultieren vertraulichkeitsbasierte Remember-Antworten. Die Vertraulichkeit der Reize verteilt sich dabei für beide Reizklassen normal, wobei die Vertraulichkeit der gelernten Reize im Mittel um d' größer ist als die der Distraktoren. Die Streuung der Vertraulichkeitsverteilung für alte Reize beträgt σ_{Alt} .

$$p(K|Alt) = (1 - r) \cdot (\Phi(\frac{d' - C_K}{\sigma_{Alt}}) - \Phi(\frac{d' - C_R}{\sigma_{Alt}})) \quad (4.44)$$

$$p(G|Alt) = (1 - r) \cdot (\Phi(\frac{d' - C_O}{\sigma_{Alt}}) - \Phi(\frac{d' - C_K}{\sigma_{Alt}})) \quad (4.45)$$

$$p(N|Alt) = (1 - r) \cdot (1 - \Phi(\frac{d' - C_O}{\sigma_{Alt}})) \quad (4.46)$$

$$p(R|Neu) = \Phi(-C_R) \quad (4.47)$$

$$p(K|Neu) = \Phi(-C_K) - \Phi(-C_R) \quad (4.48)$$

$$p(G|Neu) = \Phi(-C_O) - \Phi(-C_K) \quad (4.49)$$

$$p(N|Neu) = 1 - \Phi(-C_O) \quad (4.50)$$

An dieser Stelle lässt sich sehr gut erkennen, dass sowohl das HTSD-Modell als auch das SDM-Modell Untermodelle des HTUVSD-Modells darstellen. Dies wird daraus ersichtlich, dass sich das HTUVSD-Modell durch geeignete Parameterrestriktionen in die beiden anderen Modelle überführen lässt: Um das HTSD-Modell zu erstellen, ist es notwendig, die Einschränkung $\sigma_{Alt} = 1$ vorzunehmen, während sich das HTUVSD-Modell durch die Restriktion $r = 0$ in das SDM-Modell überführen lässt. Dieser Umstand ermöglicht es, mit einfachen Methoden einen direkten Vergleich der beiden Untermodelle mit ihrem Obermodell vorzunehmen. Für den Vergleich der beiden Untermodelle untereinander sowie für den Vergleich aller übrigen Modelle gilt dies nicht. Hierfür sind, wie im nächsten Kapitel gezeigt wird, deutlich komplexere methodische Ansätze erforderlich.

Kapitel 5

Methoden der Modellselektion

Im letzten Kapitel wurden sechs quantitative Modelle betrachtet, die Kernaussagen verschiedener Theorien nutzen, um Antworthäufigkeiten in Remember-Know-Guess-Untersuchungen zu erklären. Dabei wurde stets auf die für die jeweilige Theorie relevanten Konstrukte zurückgegriffen. Wie aufgezeigt wurde, erlauben es die vorgestellten Messmodelle, drei gewichtigen Problemen der empirischen Forschung in der kognitiven Psychologie beizukommen, die einer Entscheidung zwischen zu vergleichenden Theorien anhand des Operationalisierungsansatzes im Wege stehen. Zwei dieser drei Probleme – das Bedeutsamkeitsproblem und das Dekomponierungsproblem – wurden im Rahmen des Modellierungsansatzes gelöst, indem die Kernannahmen der Theorien verschmolzen wurden mit den Annahmen darüber, wie aus empirischen Daten Maße für die bedeutsamen Konstrukte hergeleitet werden können. Die Lösung des Fehlerproblems hingegen lag nicht in bestimmten Eigenschaften der Messmodelle an sich. Eine solche konnte vielmehr in geeigneten statistischen Methoden gefunden werden, die es erlauben, zwischen zufälligen Modellverstößen, die auf Fehlern beruhen, und ernstzunehmenden Datenanpassungsproblemen, die auf unbrauchbare Modellannahmen schließen lassen, zu unterscheiden. Die Frage danach, wie derartige Methoden aussehen, ist bislang genauso offen geblieben wie die weiterführende Frage nach Möglichkeiten, die Güte verschiedener Messmodelle zu demselben Gegenstandsbe- reich objektiv zu bewerten, um darauf aufbauend einen Vergleich zwischen den Modellen vornehmen zu können. Diese beiden Fragen werden in diesem Kapitel nachge- gangen werden.

5.1 Modellgütekriterien

Eine Reihe wichtiger Modellgütekriterien geht auf einen häufig zitierten Modellver- gleich von Jacobs und Grainger (1994) zurück, in dem mehrere Modelle zur visuellen Worterkennung hinsichtlich ihrer Güte evaluiert wurden. Da sich in der Arbeit von Ja- cobs und Grainger keine genaue Auflistung der verwendeten Kriterien findet, variiert die Anzahl der Kriterien in der auf diese Arbeit Bezug nehmenden Literatur. Während Pitt, Myung und Zhang (2002) insgesamt sechs Kriterien erwähnten, unterschied Ma-

cho (2004) nur vier Gütekriterien. Zwischen diesen beiden Arbeiten besteht Einigkeit über drei Kriterien, die, wie Pitt et al. (2002) betonten, für den Vergleich mathematischer Modelle besonders wichtig sind. Bei dem ersten dieser drei Kriterien handelte es sich um die deskriptive Adäquatheit, die sich mit der Frage befasst, ob das Modell eine gute Beschreibung der beobachteten Daten abgibt. Einigkeit besteht zudem bezüglich der Generalisierbarkeit, also der Fähigkeit des Modells, zukünftige Ergebnisse vorherzusagen, sowie bezüglich der Modellkomplexität, worunter zu verstehen ist, inwieweit das Modell den Untersuchungsgegenstand auf die einfachste Art und Weise erfasst.

Die Verfahren zur Erfassung dieser Modellgütekriterien ordneten Pitt, Kim, Navarro und Myung (2006) in ein Klassifikationsschema ein, in dem auf der ersten Dimension quantitative von qualitativen Verfahren unterschieden werden. Quantitative Modellselektionsverfahren zeichnen sich dadurch aus, dass bei der Modellevaluation auf die Fähigkeit der Modelle zurückgegriffen wird, quantitative Aussagen über den Ausgang von Untersuchungen zu machen. Da alle Modelle, die in dieser Arbeit im Mittelpunkt stehen, quantitative Vorhersagen für Antworthäufigkeiten erlauben, interessieren an dieser Stelle vorwiegend die quantitativen Verfahren. Pitt et al. (2006) zählten zu diesen quantitativen Verfahren Datenanpassungsindizes wie die Likelihood-Quotienten-Statistik (G^2) und Informationskriterien, die neben der Datenanpassungsleistung auch die Modellkomplexität berücksichtigen, wie das Akaike-Informationskriterium (AIC , Akaike, 1973) oder das Bayesianische Informationskriterium (BIC , Schwartz, 1978). Auch der Kreuzvalidierungsindex (CV , Stone, 1974) zur Erfassung der Generalisierbarkeit sowie das Landscaping-Verfahren (Navarro, Pitt & Myung, 2004) wurden zu diesen quantitativen Verfahren gezählt.

Auf der zweiten Entscheidungsachse unterteilten Pitt et al. (2006) diese Zugänge in lokale und globale Verfahren. Mit Ausnahme des Landscapings fielen alle erwähnten Methoden in die Kategorie der lokalen Verfahren, da sie das Verhalten der Modelle nur vor dem Hintergrund der aktuell vorhandenen empirischen Daten reflektieren. Diese Verfahren blicken also lokal auf einen spezifischen Punkt des Parameterraums anstatt das Modellverhalten global über weite Teile des Parameterraumes hinweg zu beleuchten (vgl. Pitt et al., 2006). Eine derartige Erfassung des Modellverhaltens in seiner Gänze ist das zentrale Anliegen der globalen Verfahren.

Während in den Arbeiten zur Modellselektion Einigkeit über diese drei quantitativ erfassbaren Kriterien besteht, lassen sich bezüglich der Eigenschaften, die diese Gütekriterien flankieren, deutliche Unterschiede ausmachen. Nimmt man die Arbeiten von Pitt et al. (2002) und von Macho (2004) zur Hand, so stellt man fest, dass Pitt et al. (2002) drei weitere Kriterien unterschieden, die sich allesamt nur qualitativ erfassen lassen. Bei diesen drei zusätzlichen Kriterien handelte es sich erstens um die Plausibilität, die danach fragt, wie plausibel die Modellannahmen vor dem Hintergrund wissenschaftlicher Erkenntnisse sind; zweitens um die explanatorische Adäquatheit, die sich damit beschäftigt, ob das Modell eine schlüssige Erklärung liefert, die sich in bereits vorhandenes Wissen einbetten lässt, und drittens um die Interpretierbarkeit, die dann vorliegt, wenn sich das Modell im Ganzen und die einzelnen Modellparameter sinnvoll interpretieren lassen.

Im Unterschied dazu griff Macho (2004) ausschließlich das Konzept der expla-

natorischen Adäquatheit auf, unter der er die Plausibilität der dem Modell zugrunde liegenden Annahmen versteht. Damit erinnert das Konstrukt der explanatorischen Adäquatheit nach Macho an das von Pitt et al. (2002) angeführte Kriterium der Plausibilität. Durch den zusätzlichen Vorschlag einer Erfassungsmethode, die darin besteht, die Plausibilität von Parameterschätzungen zu untersuchen, erweitert Macho die explanatorische Adäquatheit um Aspekte, die am ehesten dem von Pitt et al. (2002) vorgeschlagenen Konzept der Interpretierbarkeit entsprechen. Bei der Analyse der explanatorischen Adäquatheit konzentrierte sich Macho (2004) auf qualitative Betrachtungen von Parameterschätzungen für experimentell erlangte Datensätze. Dabei interessierten ihn sowohl die Werte verschiedener Parameter innerhalb einer experimentellen Bedingung, als auch die Unterschiede von Parameterwerten zwischen verschiedenen Bedingungen eines Experiments. Ein explanatorisch adäquates Modell zeichnet sich nach Macho dadurch aus, dass sich seine Parameter in Experimenten genauso verhalten, wie es aufgrund ihres theoretischen Hintergrundes zu erwarten ist.

In diesem Punkt erinnert das Konzept der explanatorischen Adäquatheit an die von Erdfelder (2000) vorgestellte Methode der Modellvalidierung, die eine systematische Überprüfung der Gültigkeit eines Modells und seiner Teile erlaubt. Das Vorgehen bei der Validierung eines Messmodells entspricht einer Konstruktvalidierung im Sinne von Cronbach und Meehl (1955), mit dem Unterschied, dass statt einer korrelativen Erfassung, wie sie bei der diagnostischen Testkonstruktion vorgenommen wird, eine experimentelle Untersuchung der Konstrukte erfolgt. Dabei wird darauf abgezielt, für jeden im Modell enthaltenen Parameter Manipulationen zu finden, von denen aufgrund des theoretischen Hintergrundes des Parameters angenommen werden kann, dass sie einen spezifischen Effekt auf diesen ausüben. Können derartige Effekte auf die Parameter nachgewiesen werden, so belegt dies die Konstruktvalidität dieser Parameter im Sinne von konvergenter Validität. Erdfelder (2000) sprach in diesem Zusammenhang von einem schwachen Nachweis der Konstruktvalidität. Von dieser abheben ließ sich ein starker Nachweis der Konstruktvalidität, der durch die zusätzliche Beobachtung diskriminanter Validität erbracht werden kann (vgl. Klauer & Wegener, 1998). Ein starker Nachweis der Konstruktvalidität liegt dann vor, wenn die oben angesprochenen experimentellen Manipulationen nicht nur zu dem angenommenen Effekt auf einen der Parameter führen, sondern zugleich keinen der übrigen Modellparameter affizieren.

Da die Anzahl der möglichen Manipulationen, die sich in spezifischer Weise auf einen Parameter auswirken, nicht nach oben begrenzt ist, ist es zwar falsch, einem Modell endgültig die Eigenschaft der Konstruktvalidität zuzuschreiben, wiederholte Belege der Konstruktvalidität eines Modells sprechen jedoch im Sinne einer empirischen Bewährung für die Gültigkeit der Modellannahmen. Leider liefert weder das Verfahren der Modellvalidierung noch die von Macho (2004) vorgeschlagene Methode zur Bestimmung der explanatorischen Adäquatheit einen exakten Index, der es erlaubt, verschiedene Modelle direkt miteinander zu vergleichen. Wie in Kapitel 10 zu sehen sein wird, ist es allerdings möglich, sich einer quantitativen Erfassung der Konstruktvalidität anzunähern.

Dazu werden die Modelle in einer hypothesenkonformen Variante und einer hypotheseninkonformen Variante an Daten verschiedener Experimente angepasst. In der

theoriekonformen Variante ist das Modell so formuliert, dass sich zwischen den Experimentalbedingungen die Werte der Parameter verändern können, für die ein Effekt theoretisch vorhergesagt wird, während die Parameter, für die keine Veränderung erwartet wird, zwischen den Bedingungen gleichgesetzt werden. In der hypotheseninkonformen Modellvariante werden hingegen nur die Parameter frei gelassen, für die man theoretisch keinen Effekt erwartet, wohingegen die Parameter, die auf die Manipulation mit einer Veränderung reagieren sollten, gleichgesetzt werden. In dieser Art von Test sollten sich konstruktvalide Modelle dadurch auszeichnen, dass sie bei hypotheseninkonformer Anpassung deutlich häufiger zurückgewiesen werden als bei theoriekonformer Anpassung. In Kapitel 10 wird diese Methode genauer beschrieben werden.

Grundlegend für alle vorgestellten Verfahren zur Erfassung der Modellgüte sind sogenannte Maximum-Likelihood-Schätzverfahren. Anhand dieser lassen sich zum einen für Datensätze Parameterwerte bestimmen, die, die Modellgültigkeit vorausgesetzt, die Daten erzeugt haben könnten, und zum anderen ergibt sich dabei mit der maximalen Likelihood ein quantitativer Wert, der wichtige Informationen über die Datenanpassungsgüte liefert. Anhand dieses Wertes lassen sich, wie im Anschluss dargelegt werden wird, Maße für die deskriptive Adäquatheit eines Modells bestimmen. Zudem kann die maximale Likelihood so erweitert werden, dass der resultierende Index neben der deskriptiven Adäquatheit auch die Modellkomplexität berücksichtigt bzw. es erlaubt, die Flexibilität des Modells und die Generalisierbarkeit der Modellvorhersagen in jeweils eigenen Maßen zu erfassen.

5.2 Maximum-Likelihood-Parameterschätzung

Das statistische Konzept der maximalen Likelihood geht auf Arbeiten von Ronald A. Fisher zurück, der schon mit 22 Jahren argumentiert haben soll, dass der beste Schätzer für einen unbekannt Parameter stets derjenige ist, der am wahrscheinlichsten ist und damit eine maximale Likelihood aufweist (vgl. Burnham & Anderson, 2003). Um für kategoriale Daten die Likelihood eines einzelnen Parameters oder eines Parametervektors zu ermitteln, bedient man sich der Likelihood-Funktion, die in Gleichung 5.1 dargestellt ist. Die Likelihood-Funktion (\mathcal{L}) entspricht dem Produkt von Potenzen, in denen die erwarteten Kategorienwahrscheinlichkeiten (π_i) für jede Kategorie i von insgesamt K unterschiedlichen Kategorien mit den zugehörigen empirischen Kategorienhäufigkeiten (y_i) potenziert werden. Diese Funktion lässt sich einfach auf verbundene multinomiale Modelle erweitern (vgl. Erdfelder et al., 2006). Effiziente Maximum-Likelihood-Schätzer, also Werte, die, wenn sie für π_i eingesetzt werden, zu einer maximalen Likelihood führen, sind die relativen Kategorienhäufigkeiten $\xi_i = y_i / \sum_{i=1}^K y_i$. Setzt man diese Kategorienhäufigkeiten an Stelle von π_i in die Likelihood-Funktion ein, so erhält man die in Anbetracht der Daten maximal erreichbare Likelihood. Dieser Fall, in dem sich die Daten selbst erklären, wird auch als Datenanpassung des Null-Modells bezeichnet. Bei der Datenanpassung des Null-Modells bleibt der Informationsgehalt der Daten in vollem Umfang erhalten, so dass der sich er-

gebende Anpassungswert als obere Grenze der Likelihood betrachtet werden kann, die ein theoretisches Modell erreichen kann. Dieser Punkt wird in Kürze benötigt, sobald Indizes behandelt werden, die es ermöglichen, die Datenanpassungsgüte zu erfassen. Zuvor soll eine andere Frage beleuchtet werden.

$$\mathcal{L}(\pi|y) = \frac{(\sum_{i=1}^K y_i)!}{\prod_{i=1}^K y_i!} \cdot \prod_{i=1}^K \pi_i^{y_i} \quad (5.1)$$

Bei der Arbeit mit theoretisch fundierten Modellen interessiert in aller Regel weniger, wie gut die Daten sich selbst erklären. Vielmehr zielen solche Modelle darauf ab, eine eigene Erklärung für die beobachteten Daten zu liefern. Dies lässt sich dadurch bewerkstelligen, dass in der Likelihood-Funktion an die Stelle der erwartenden Kategorienwahrscheinlichkeiten (π_i) keine konkreten Werte geschrieben werden, sondern die Vorhersagegleichungen des jeweiligen Modells für die korrespondierenden Datenkategorien eingesetzt werden. Diese Vorhersagegleichungen entsprechen formal Funktionen der Modellparameter θ , wobei θ für den Parametervektor steht, der alle Modellparameter umfasst, die für das jeweilige Modell relevant sind. Wenn die Modellgleichungen $p_i(\theta)$ an die Stelle der erwarteten Häufigkeiten in die Likelihood-Funktion gesetzt worden sind, ist die Likelihood-Funktion eine Funktion der Modellparameter θ auf der einen und der beobachteten Kategorienhäufigkeiten auf der anderen Seite.

Das nächste Ziel kann nun darin gesehen werden, genau jene Werte für θ zu ermitteln, die die Likelihood der Daten maximieren. Solche Maximum-Likelihood-Schätzer $\hat{\theta}$ lassen sich für die jeweils im Zentrum des Modells stehenden Parameter nach Gleichung 5.2 ermitteln. Technisch bedeutet dies, dass $\hat{\theta}$ dadurch bestimmt wird, dass der maximale Wert der Likelihood-Funktion $\mathcal{L}(\theta)$ ermittelt wird. Die an der Stelle $\max \mathcal{L}(\theta)$ beobachteten Parameterwerte entsprechen dann den Maximum-Likelihood-Schätzern $\hat{\theta}$. Der erreichte Likelihood-Funktionswert stellt die maximale Likelihood (*ML*) der Modellparameter vor dem Hintergrund der beobachteten Daten sowie der Modellannahmen dar. Diese Maximierung ist nur in den seltensten Fällen auf analytischem Wege möglich, weshalb man sich dazu in der Regel iterativer Verfahren bedient, die eine effiziente Bestimmung der Optimalwerte ermöglichen.

$$\mathcal{L}(\hat{\theta}) = \max \mathcal{L}(\theta) \quad (5.2)$$

Da der Betrag der Likelihood mit wachsendem Umfang des Datensatzes sehr schnell gegen Null konvergiert, empfiehlt es sich in der Praxis, nicht mit der Likelihood selbst, sondern stattdessen mit dem Logarithmus Naturalis der Likelihood zu arbeiten. Zusätzlich wird der Multinomialkoeffizient $\left(\frac{(\sum_{i=1}^K y_i)!}{\prod_{i=1}^K (y_i)!}\right)$, der ausschließlich von Art und Umfang des Datensatzes abhängt, und deshalb für einen gegebenen Datensatz – unabhängig vom zu betrachtenden Modell – konstant ist, in der Regel vernachlässigt. Entsprechend wird das Maximum der Likelihood-Funktion anhand der Log-Likelihood-Funktion (\mathcal{LL}) bestimmt, die in Gleichung 5.3 niedergelegt ist.

$$\mathcal{LL}(\theta|y) = \sum_{i=1}^K y_i \cdot \ln p_i(\theta) \quad (5.3)$$

Der bekannteste Algorithmus, der zum Auffinden des globalen Maximums dieser Funktion eingesetzt werden kann, ist der EM-Algorithmus (Dempster, Laird & Rubin, 1977). Eben dieser wird in Programmen wie MBT von Hu (1999), Appletree von Rothkegel (1999) oder HMMTree von Stahl und Klauer (Im Druck), verwendet (vgl. Erdfelder et al., 2006). Diese Programme sind speziell dazu entwickelt worden, um Parameterschätzungen und Datenanpassungen für multinomiale Verarbeitungsbaummodelle vorzunehmen. Alternativ können sogenannte Allzweck-Optimierer eingesetzt werden, um das Maximum der \mathcal{LL} -Funktion zu finden. So empfehlen beispielsweise Macho (2002) sowie Dodson, Prinzmetal und Shimamura (1998) zur Datenanpassung an multinomiale Modelle allgemeiner Art das Produkt Solver, bei dem es sich um den in Microsoft-Excel verfügbaren Allzweck-Optimierer handelt.

In der vorliegenden Arbeit wurde zur Maximierung der \mathcal{LL} -Funktion auf den BFGS-Algorithmus zurückgegriffen. Dieser beruht auf einer gradientenbasierten Quasi-Newton-Methode, die im Jahre 1970 von Broyden und Fletcher sowie Goldfarb und Shanno vorgestellt worden ist. Eine genaue Bewertung des BFGS-Verfahrens kann Nocedal und Wright (2005) entnommen werden.

Wie bei allen hier angesprochenen Algorithmen kann auch bei der BFGS-Methode das Problem auftreten, dass statt des globalen Maximums ein lokales Maximum der zu optimierenden Funktion gefunden wird, was dazu führt, dass die genannten Algorithmen die Suche fälschlicherweise mit einer Erfolgsmeldung abbrechen. Um dieses Problem zu umgehen, wurde in der vorliegenden Arbeit jede Maximierung zehnmal mit unterschiedlichen Startwerten durchgeführt. Verwendet wurde anschließend die jeweils beste dieser zehn Lösungen.

Neben den Maximum-Likelihood-Schätzern für die angenommenen Modellparameter gibt das verwendete Minimierungsprogramm nach zehn erfolgreichen Optimierungsdurchgängen auch den Maximum-Likelihood-Wert¹ des Modells zurück. Dieser Likelihood-Wert stellt ein Maß dafür dar, wie wahrscheinlich das Modell vor dem Hintergrund der gegebenen empirischen Daten ist. Daher greift eine ganze Reihe von Datenanpassungsindizes auf diesen Wert zurück, um die Abweichung zwischen den aufgrund des Modells zu erwartenden und den tatsächlich beobachteten Daten zu quantifizieren. Einige dieser Indizes werden der Familie der Power-Divergenz-Statistiken PD^λ zugeordnet (Read & Cressie, 1988).

$$PD^\lambda(\theta|y) = \frac{2}{\lambda(\lambda + 1)} \cdot \sum_{i=1}^K y_i \cdot \left[\left(\frac{y_i}{\sum_{j=1}^K y_j \cdot p_j(\theta)} \right)^\lambda - 1 \right], \text{ für } \lambda \in \mathbb{R} \setminus [-1; 0] \quad (5.4)$$

¹Mit *Likelihood* wird von nun an der nach Formel 5.1 logarithmierte Likelihood-Wert bezeichnet.

Die zu dieser Familie gehörigen Anpassungsmaße lassen sich aus der in Gleichung 5.4 dargelegten Formel entwickeln, indem für den Scharparameter λ ein für das jeweilige Maß charakteristischer Wert eingesetzt wird. Bei $\lambda \mapsto 0$ ergibt sich beispielsweise die im Rahmen der multinomialen Modellierung häufig verwendete Likelihood-Quotienten-Statistik G^2 .

Die Statistik G^2 entspricht, wie in Gleichung 5.5 ersichtlich, dem doppelten Wert der Differenz zwischen der logarithmierten Likelihood des theoretisch fundierten Modells $\mathcal{L}\mathcal{L}(\hat{\theta}|y)$ und der logarithmierte Likelihood des Null-Modells ($\mathcal{L}\mathcal{L}(\xi|y)$).

$$G^2 = 2 \cdot (\mathcal{L}\mathcal{L}(\hat{\theta}|y) - \mathcal{L}\mathcal{L}(\xi|y)) \quad (5.5)$$

Wie bereits dargelegt worden ist, entspricht die Likelihood des infrage stehenden Messmodells im besten Falle der Likelihood des Null-Modells, so dass die G^2 -Statistik mindestens einen Wert von 0 haben muss. Weiter hat die positiv reellwertige G^2 -Statistik die Eigenschaft, bei Gültigkeit des zugrunde liegenden Messmodells einer $\chi^2_{(df)}$ -Verteilung zu folgen, deren Freiheitsgrade (df) bis auf in definierten Ausnahmefällen (Bamber & Van Santen, 1985) durch die Zahl der Modellparameter, die das Null-Modell mehr enthält als das infrage stehende Messmodell, erfasst werden kann. Dadurch wird es möglich zu überprüfen, ob die durch das Modell eingeführte Informationsreduktion verglichen mit dem Null-Modell zu einer statistisch bedeutsamen Verschlechterung der Datenanpassung geführt hat. Eine signifikante Verschlechterung der Datenanpassung ist dabei nicht angestrebt oder erwünscht, vielmehr geht man, wenn der Erklärungswert eines Modells allzu oft hinter dem des Null-Modells zurückbleibt, davon aus, dass das Modell entweder fehlspezifiziert ist oder zumindest wichtige Einflussgrößen nicht thematisiert.

Solche Likelihood-Quotienten-Tests können jedoch nicht nur eingesetzt werden, um ein Messmodell mit einem Null-Modell zu vergleichen, sondern erlauben zudem den Vergleich von ineinander genesteten Modellen. Unter genesteten Modellen sind Messmodelle zu verstehen, die sich durch geeignete Parameterrestriktionen ineinander überführen lassen. Die Datenanpassungsleistungen dieser Modelle lassen sich anhand der Differenz zwischen dem G^2 -Wert des weniger stark restringierten Modells und dem G^2 -Wert des stärker restringierten Tochtermodells miteinander vergleichen. Statistisch kann dieser Vergleich über einen $\chi^2_{(df)}$ -Referenzwert abgesichert werden, wobei die Freiheitsgrade (df) dieses Tests der Anzahl von freien Parametern entspricht, die das weniger restringierte Obermodell mehr enthält als das Tochtermodell.

Wie bereits deutlich geworden ist, stellen die Likelihood-Quotienten-Statistik G^2 und verwandte Indizes wichtige Modellgütekriterien dar. Es ist jedoch ebenfalls angeklungen, dass diese den direkten Vergleich verschiedener Modelle an einem Datensatz nur für ineinander genestete Modelle erlauben (Myung & Pitt, 2002). Dies ist vor allem darauf zurückzuführen, dass für den Vergleich nicht genesteter Modelle neben der Datenanpassungsleistung (im Englischen *Goodness-of-Fit*) die Komplexität der einzelnen Modelle berücksichtigt werden muss. Letztere kann die Datenanpassungsleistung eines Modells positiv beeinflussen, ohne einen Erkenntnisfortschritt bereit zu halten.

Wie Roberts und Pashler (2000) betonten, führt eine ausschließliche Orientierung an der Datenanpassungsgüte bei der Modellselektion in den meisten Fällen nicht zur

Auswahl des Modells, das am besten dazu geeignet ist, die Gesetzmäßigkeiten zu erfassen, die sich in den beobachteten Daten verbergen. Vielmehr führt eine solche Orientierung gerade vor dem Hintergrund fehlerbelasteter Daten dazu, dass generell komplexere Modelle weniger komplexen Modellen vorgezogen werden (Pitt et al., 2002). Modellkomplexität machten die Autoren dabei an zwei verschiedenen Eigenschaften fest: Erstens können Modelle eine unterschiedliche Anzahl latenter Konstrukte annehmen und dadurch von der Parameterzahl her unterschiedlich komplex sein, zweitens ist es möglich, dass Unterschiede in der mathematischen Struktur der Modelle dazu führen, dass ein Modell eine höhere Komplexität aufweist als ein anderes.

Unabhängig davon, worin die hohe Komplexität eines Modells begründet liegt, führt diese zu einem besseren Datenanpassungsverhalten, was darauf beruht, dass komplexere Modelle nicht nur regelhafte Einflüsse auf die Daten modellieren, sondern auch Fehlereinflüsse nachzubilden vermögen. Um zu vermeiden, dass bei der Modellselektion durchgängig weniger korrekte, komplexere Modelle sparsameren Modellen vorgezogen werden, zeigten Pitt et al. (2002) zwei Lösungswege auf. Einer dieser Lösungswege besteht darin, zum Vergleich von Modellen Maße zu verwenden, die sowohl die Datenanpassungsleistung als auch die Komplexität der Modelle berücksichtigen. Als zweiten Zugang schlugen die Autoren vor, den Fokus weniger auf die deskriptive Adäquatheit der Modelle, wie sie durch Goodness-of-Fit-Maße erfasst wird, zu richten und stattdessen das Hauptaugenmerk auf die Generalisierbarkeit zu legen. Anders als auf die deskriptive Adäquatheit wirkt sich eine Steigerung der Modellkomplexität auf die Generalisierbarkeit nicht durchgängig positiv aus. Vielmehr steigt die Generalisierbarkeit eines Modells nur solange mit der Komplexität an, wie durch die hinzukommenden Komponenten Gesetzmäßigkeiten abgebildet werden. Nimmt die Modellkomplexität über diesen Punkt hinaus weiter zu, so kann das zwar weiterhin zu einer Verbesserung der Datenanpassungsleistung führen, die Generalisierbarkeit sinkt unter diesen Umständen hingegen, da der Komplexitätsüberschuss dann zwangsläufig dazu genutzt wird, temporäre Fehlervarianz zu erfassen (Myung & Pitt, 2002).

Verbreitete Modellselektionskriterien, die sowohl den Goodness-Of-Fit als auch die Modellkomplexität berücksichtigen, können Tabelle 5.1 entnommen werden. Zu diesen Kriterien gehören unter anderem das Bayesianische Informationskriterium (*BIC*), das Akaike-Informationskriterium (*AIC*) sowie der Kreuzvalidierungsindex (*CV*). Sowohl *AIC* als auch *BIC* berücksichtigen die Modellkomplexität insofern, als sie neben der Datenanpassungsleistung ($\ln \mathcal{L}(y|\hat{\theta})$) einen als *Penalty-Term* bezeichneten Strafterm aufnehmen, der die Komplexität erfasst und von der Datenanpassungsleistung abgezogen wird. Dieser Strafterm wird in beiden Fällen hauptsächlich durch die Anzahl der in dem Modell enthaltenen Parameter (k) bestimmt. *BIC* berücksichtigt zudem den Umfang des Datensatzes (n) und gewichtet die Anzahl der Modellparameter in solchen Fällen stärker, in denen n , also die Anzahl der Beobachtungen, groß ist. Allerdings beziehen weder *AIC* noch *BIC* die formale Komplexität des Modells aufgrund seiner mathematischen Struktur mit ein.

Die Methode der Kreuzvalidierung dagegen beinhaltet auch diese Art der Komplexität und unterscheidet sich somit von den zuvor erläuterten Methoden. Bei der Kreuz-

Tabelle 5.1: Aufgelistet werden geläufige Modellselektionsindizes sowie Vorschriften zu ihrer Berechnung.

| Datenanpassungsmaß | Berechnungsvorschrift |
|--|---|
| Akaike-Informationskriterium ^a (<i>AIC</i>) | $AIC = -2 \cdot \mathcal{L}\mathcal{L}(\hat{\theta} y) + 2 \cdot k$ |
| <i>AIC</i> für kleine Datensätze ^b (<i>AIC_c</i>) | $AIC_c = AIC + \frac{2k(k+1)}{n-k-1}$ |
| Kreuzvalidierung ^c (<i>CV</i>) | $CV = -\mathcal{L}\mathcal{L}(\hat{\theta}_{Kal} y_{Val})$ |
| Bayesianisches Informationskriterium ^d (<i>BIC</i>) | $BIC = -2 \cdot \mathcal{L}\mathcal{L}(\hat{\theta} y) \cdot \ln n$ |

Anmerkungen. y = Stichprobe der Größe n ; $\hat{\theta}$ = Parametervektor, der die Likelihood-Funktion $\mathcal{L}(\theta|y)$ maximiert; k = Anzahl der Parameter; n = Stichprobenumfang; y_{Val} = Validierungsstichprobe; $\hat{\theta}_{Kal}$ = Parameterschätzungen für die Kalibrierungsstichprobe

^aAkaike (1973)

^bHurvich und Tsai (1989)

^cAllen (1974) und Stone (1974)

^dSchwartz (1978)

validierung werden die beobachteten Daten zunächst in m Teile unterteilt. Je $m - 1$ dieser Teile werden bei der Bestimmung des CV-Indexes als sogenannte Kalibrierungsstichprobe verwendet, während der übrig bleibende Teil als Validierungsstichprobe eingesetzt wird. In dem dann anstehenden Schritt werden zuerst die Parameterwerte bestimmt, die an die Kalibrierungsstichproben am besten passen. Anschließend wird der Kreuzvalidierungsindex berechnet, indem die Datenanpassung der in den m Kalibrierungsdurchgängen erhaltenen Maximum-Likelihood-Schätzwerte an die m Validierungsstichproben ermittelt wird. Dieses auf Stone (1974) und Allen (1974) zurückgehende Vorgehen stellt eine in vielen Fällen einfach anzuwendende Methode dar, um die Generalisierbarkeit von Modellen heuristisch zu erfassen. Zwar liefert die Kreuzvalidierung kein explizites Maß für die Komplexität eines Modells, aufgrund des stark auf die Erfassung der Generalisierbarkeit abstellenden Vorgehens kann jedoch angenommen werden, dass die Komplexität implizit berücksichtigt wird.

Neben diesen drei vorgestellten Verfahren existieren noch einige weitere Methoden, die einen Modellvergleich auf der Basis empirischer Daten ermöglichen. Diese erfordern teilweise einen mathematischen Schritt der Integration der Likelihood-Funktion, wie sie für die in dieser Arbeit zu vergleichenden Modelle nicht geleistet werden kann. Auf die Vorstellung der entsprechenden Verfahren – das sind die Bayesianische Modellselektion (Kass & Raftery, 1995) und die Minimale Beschreibungslänge (Grünwald, 2000; Rissanen, 1996) – wird aus diesem Grund verzichtet. Weitere Methoden bedienen sich sogenannter Resampling-Verfahren, die unter dem Namen *Bootstrapping* bekannt geworden sind. Diese können im Rahmen der Modellselektion zu verschiedenen Zwecken eingesetzt werden. Wie im folgenden Abschnitt gezeigt werden wird, erlaubt es das Bootstrapping nicht nur, Vertrauensintervalle für Parameterschätzungen zu ermitteln, sondern kann darüber hinaus dazu eingesetzt wer-

den, um die Generalisierbarkeit von Modellvorhersagen zu ermitteln.

5.3 Bootstrapping und Modellselektion

Obwohl schon vor der ersten Beschreibung des Bootstrapping durch Efron (1979) Methoden verwendet wurden, die dessen Logik folgten, wurden Name und Technik des Bootstrapping Efron (1979) zugeschrieben, der sie zum ersten Mal systematisch behandelt hat. Dabei leitet sich der Name 'Bootstrapping' von den berühmten Stiefelschlaufen des Baron Münchhausen ab, an denen dieser sich eigenhändig samt Pferd aus einem Sumpfloch rettete, in dem beide zu versinken drohten. Zudem folgt der Namen dem Vorbild Tukeys (1958), der die von ihm vorgeschlagene Resampling-Technik 'Jackknifing' nannte, weil sie seiner Meinung nach für Statistiker das darstellen sollte, was für den Pfadfinder ein Klappaschenmesser ist. Die Theorie des Bootstraps findet sich gut zusammengefasst in der Arbeit von Efron und Tibshirani (1993) und der Arbeit von Manly (1997) wieder. Einen sehr umfassenden Überblick findet die interessierte Leserin bzw. der interessierte Leser in dem Buch von Davison und Hinkley (1997).

Die grundlegende Idee des Bootstrappings besteht in der Annahme, dass eine Zufallsstichprobe der Größe n aus einer Population der beste Schätzer für die Populationsverteilung ist, wenn außer dieser Zufallsstichprobe selbst keine Informationen zur Verfügung stehen. Deshalb ist es sinnvoll, wiederholt Stichproben aus der Zufallsstichprobe zu ziehen, um sich approximativ der Antwort auf die Frage anzunähern, was passieren würde, wenn man wiederholt Stichproben aus der Population zöge. Ziel des Bootstrappings ist es also, durch wiederholtes Ziehen von Stichproben aus der beobachteten Zufallsstichprobe Rückschlüsse auf die unbekannt reale Population zu ziehen. Dazu wird von einer unendlichen Population ausgegangen, die aus den n beobachteten Ereignissen besteht. Aus dieser Population werden dann mit Zurücklegen B Stichproben der Größe n gezogen. Dabei wird allen Ereignissen die gleiche Sampling-Wahrscheinlichkeit $1/n$ zugewiesen. Können die Ereignisse zu Klassen der Stärke j zusammengefasst werden, entspricht die Sampling-Wahrscheinlichkeit von Ereignissen einer jeden Klasse dem Quotienten aus Klassenstärke und der absoluten Anzahl von Beobachtungen (j/n).

Eines der wichtigsten Themen der Bootstrapping-Forschung war in den letzten zwanzig Jahren die Entwicklung von Methoden, anhand derer sich verlässliche Konfidenzintervalle für Populationsparameter bestimmen lassen. Daneben wurden in neuerer Zeit einige Methoden vorgeschlagen, die den Bootstrap in der Modellselektion nutzbar machten.

Werden Bootstrapping-Methoden eingesetzt, um Konfidenzintervalle zu ermitteln oder den Standardfehler eines Merkmals zu schätzen, so steht dabei jeweils die Genauigkeit einzelner Parameterschätzungen im Mittelpunkt der Betrachtungen. Wie im Folgenden dargestellt wird, kann das Bootstrapping auch eingesetzt werden, um die Messgenauigkeit eines ganzen Modells zu erfassen. Die Erfassung der Genauigkeit ganzer Modelle mit Hilfe des Bootstrappings ist das Ziel von Methoden, die unter dem Na-

men Bootstrap-Modellselektion (BMS) bekannt geworden sind. Im Mittelpunkt dieser Verfahren steht der sogenannte Vorhersagefehler, anhand dessen quantifiziert werden kann, wie genau ein Modell zukünftige Ereignisse vorhersagt. Der Vorhersagefehler stellt also ein Maß für die Generalisierbarkeit eines Modells dar. Je geringer der Vorhersagefehler ausfällt, desto besser eignet sich ein Modell zur Vorhersage zukünftiger Ereignisse. Entsprechend werden Techniken zur Erfassung dieser Fehlerart häufig zur Modellselektion eingesetzt, wobei sich das Modell am besten bewährt, das den geringsten Vorhersagefehler aufweist.

Eine Standardmethode zur Bestimmung des Vorhersagefehlers wurde im vorherigen Abschnitt unter dem Stichwort der Kreuzvalidierung bereits behandelt. An dieser Stelle soll ein weiteres Verfahren besprochen werden, das unter dem Namen Bootstrap-Modellselektion (BMS) von Efron (1983) und Shao (1996) vorgeschlagen wurde und darauf abzielt, den Vorhersagefehler zu erfassen. Die BMS-Technik folgt in weiten Teilen dem gleichen Rational wie die Kreuzvalidierung. Auch bei der BMS geht es darum, zu überprüfen, inwiefern Maximum-Likelihood-Schätzer, die aus einem Kalibrierungsdatensatz gewonnen worden sind, sich dazu eignen, die Daten eines verwandten Validierungsdatensatzes zu erklären. Wie bereits erläutert, bedient sich die Kreuzvalidierung dazu der Teile, in die der Originaldatensatz unterteilt wurde: Wenn der Originaldatensatz zufällig in m Teile zerlegt wurde, wird für jeden dieser m Teile eine Parameterschätzung an den $m - 1$ übrig bleibenden Teilen vorgenommen und die Güte der Datenanpassung ermittelt, die die geschätzten Parameterwerte für die bei der Kalibrierung nicht herangezogenen Validierungsdaten erreichen. Aus den so erlangten Daten lässt sich anhand der in Tabelle 5.1 festgehaltenen Rechenvorschrift der Kreuzvalidierungsindex CV bestimmen, der zum Vergleich verschiedener Modelle und damit zur Modellselektion verwendet werden kann. Gegenüber Bootstrap-Modellselektions-Methoden hat dieses Verfahren jedoch einen Nachteil, der insbesondere bei Datensätzen mit relativ geringem Umfang zum Tragen kommt: Mit der Notwendigkeit, den Datensatz bei der Kreuzvalidierung in einzelne Teile zerlegen zu müssen, geht ein Informationsverlust einher, der nicht ohne weiteres ausgeglichen werden kann. Bei der Bootstrapping-Modellselektion kann in der Validierungsphase hingegen von der gesamten Stichprobe Gebrauch gemacht werden, da diese – wie sich gleich zeigen wird – nur mittelbar für die Generierung der Kalibrierungsdatensätze benötigt wird.

Um mit der BMS-Technik den Vorhersagefehler eines Modells zu bestimmen, beginnt man mit einem Resampling-Schritt, in dem aus dem Originaldatensatz des Umfangs n insgesamt B Bootstrap-Replikationen hergestellt werden. Als optimalen Umfang n^* der Bootstrap-Replikationen erachten Shao und Tu (1995) einen explizit gewählten Wert m , der dann optimal ist, wenn $m/n \mapsto 0$ und $m \mapsto \infty$. Efron und Tibshirani (1993) raten hingegen dazu, den Umfang des Bootstrapdatensatzes so zu wählen, dass die künstlichen Datensätze genauso groß sind wie der Originaldatensatz, also $n^* = n$ gilt. Für die neu entstandenen Datensätze werden anschließend Parameterschätzungen vorgenommen und ML-Werte bestimmt. Es resultiert damit für jeden der B Datensätze ein Vektor, der aus den für den jeweiligen Replikationsdatensatz i adäquaten Maximum-Likelihood-Schätzern ($\hat{\theta}_i^*$) und dem zugehörigen Log-

Likelihood-Wert ($\mathcal{LL}(\hat{\theta}_i^*|x_i^*)$), der die Datenanpassung des Modells an den jeweiligen Replikationsdatensatz wiedergibt, besteht.

In den folgenden Schritten lässt sich unschwer die Ähnlichkeit zum vorher skizzierten Kreuzvalidierungsverfahren erkennen. So werden die Bootstrap-Datensätze im Folgenden als Kalibrierungsstichproben begriffen, während die Originalstichprobe als Validierungsstichprobe genutzt wird. Das einfachste Bootstrap-Maß für den Vorhersagefehler (PE) entspricht Efron und Tibshirani (1993) zufolge dem arithmetischen Mittel der B Fehlerwerte, die auftreten, wenn die an den Kalibrierungsstichproben geschätzten Modelle genutzt werden, um die Originaldaten zu erklären. Eine Möglichkeit, PE auf der Basis von Likelihood-Werten zu bestimmen, wird in Gleichung 5.6 vorgeschlagen. Nach dieser Gleichung wird PE durch die mittlere negative \mathcal{LL} geschätzt, die bei der Erklärung der Originaldaten mit den an den Bootstrapstichproben geschätzten Parametern auftreten:

$$PE = - \sum_{i=1}^B \frac{\mathcal{LL}(\hat{\theta}_i^*|x)}{B} \quad (5.6)$$

Für Efron und Tibshirani (1993) stellt ein solches einfaches Verfahren zur Bestimmung eines BMS-Indexes zwar einen wichtigen Schritt in die richtige Richtung dar, die Methode selbst ist allerdings noch verbesserungswürdig. Die Autoren schlugen vor, auf dieser Basis einen Wert zu bestimmen, den sie Optimismus nennen. Dieser Optimismus-Wert entspricht der Differenz zwischen dem einfachen Vorhersagefehler und dem Fehler, der bei der Anpassung der Modelle an die Kalibrierungsstichprobe auftritt. Die wichtigste Idee liegt hierbei in einer Korrektur des einfachen Fehlermaßes um die Variabilität in der Datenanpassung, die auf die Kalibrierungsdatensätze als solche zurückgeht.

Im Rahmen des Einsatzes von Maximum-Likelihood-Methoden bietet es sich an, den PE -Schätzer um die $\mathcal{LL}(\hat{\theta}_i^*|x_i^*)$, die das geschätzte Modell bei der Anpassung an die Bootstrapdatensätze selbst erreicht, zu bereinigen. Der in Gleichung 5.7 vorgeschlagene Optimismus-Index ($Optim$) entspricht damit der mittleren Differenz zwischen den an den Bootstrapstichproben erreichten \mathcal{LL} und der \mathcal{LL} , die bei der Vorhersage der Originaldaten beobachtet werden kann:

$$Optim = \frac{1}{B} \cdot \sum_{i=1}^B \left(\mathcal{LL}(\hat{\theta}_i^*|x_i^*) - \mathcal{LL}(\hat{\theta}_i^*|x) \right) \quad (5.7)$$

Wie auch der CV-Index berücksichtigt der Optimismus-Index implizit beide Arten der Modellkomplexität, also sowohl die Zahl freier Parameter als auch die formale Komplexität der Modelle, während die Informationskriterien AIC und BIC nur die Anzahl freier Parameter des Modells respektive den Umfang der erhobenen Stichprobe berücksichtigen. Insofern stellt die BMS-Methode Informationen über Modelle bereit, die den Informationskriterien nicht entnommen werden können.

Mit der BMS-Methode wurde das letzte ML-basierte Verfahren geschildert, das in dem Modellvergleich, der im Folgenden dargestellt werden wird, eingesetzt wurde. Im

nächsten Kapitel wird als erstes die Datenbasis, die die empirische Grundlage dieser Vergleichsstudie darstellt, charakterisiert.

Teil II

Empirische Modellvergleiche

Kapitel 6

Datenbasis

In dem nun folgenden empirischen Teil der Arbeit wurden die in Kapitel 4 vorgestellten Modelle zur Remember-Know-Guess-Prozedur dahingehend verglichen, inwiefern sie zentrale Eigenschaften brauchbarer mathematischer Modelle besitzen. In diesem Teil wird dementsprechend behandelt, inwiefern die Modelle eine adäquate Beschreibung der Daten geben können, ob ihre Parameterschätzwerte plausibel erscheinen, und in welchem Ausmaß die Parameter der Modelle konstruktvaliden genannt werden können. Zusätzlich wurden Vorhersagen der Modelle überprüft, die für die Form von ROC-Kurven im z-transformierten Raum ableitbar sind. Für diese Analysen, die zum Vergleich der Modelle hinsichtlich ihrer Modellgüte dienten, wurden empirische Daten benötigt, zu deren Erklärung die infrage stehenden Modelle prinzipiell geeignet waren. In dem nun folgenden Kapitel soll deshalb die Datenbasis des durchgeführten Modellvergleichs näher vorgestellt werden.

Durchgängig bestand die Datenbasis aus Studien, in denen Rekognitionsuntersuchungen kombiniert wurden mit einer Remember-Know-Guess-Abfrage. Dabei war es nachrangig, ob die Alt-Neu-Entscheidung wie allgemein üblich vor der Remember-Know-Guess-Abfrage gefällt werden musste oder ob von den Probanden eine einzige, vierkategoriale Entscheidung verlangt wurde. Um eine ausreichend umfassende Datenbasis zu erreichen, wurden alle entsprechenden Datensätze verwendet, die in dem Zeitraum zwischen der ersten Verwendung der infrage stehenden Prozedur im Jahr 1996 und dem Ende des Jahres 2005 in einer Fachzeitschrift erschienen waren, die entweder in der Datenbank *PsycInfo* oder im *Social Sciences Citation Index (SSCI)* geführt wurden.

Zudem mussten die ausgewählten Experimente die zusätzliche Bedingung erfüllen, dass ihre Daten in ein modellierbares Format überführt werden konnten. Da zur Modellierung absolute Antworthäufigkeiten über Personengruppen hinweg zusammengefasst werden mussten, erfüllten fast alle der recherchierten Arbeiten diese Bedingungen. Mit hinreichender Genauigkeit ließ sich das gewünschte Datenformat immer dann herstellen, wenn eine Arbeit folgenden Bedingungen genügte:

- Relative Häufigkeiten für Remember-, Know- und Guess-Antworten für alle im Experiment realisierten Zielreiz- und Distraktorbedingungen mussten entnehmbar sein;

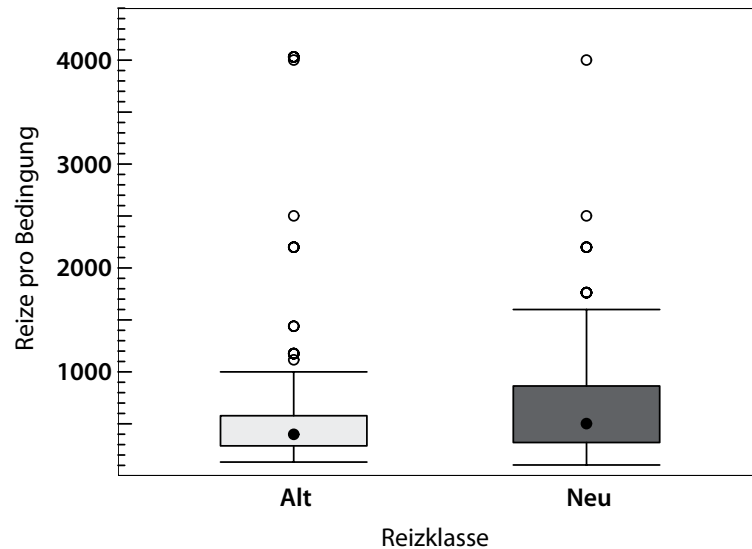


Abbildung 6.1: Dargestellt wird die absolute Anzahl der Reaktionen nach Itemtyp (Zielreize vs. Distraktoren). Entnehmbar sind der Median, das 1. und 3. Quartil, das 95%-Konfidenzintervall sowie einzelne Ausreißer.

- Die Anzahl der Reize, die in jeder Bedingung Verwendung fanden, sowie die Zahl der Probanden, die die Reize bearbeitet hat, mussten berichtet werden.

Zusammen genommen erlaubten diese Informationen einen relativ genauen Rückschluss auf die tatsächlich beobachteten Antworthäufigkeiten, wobei es allerdings aufgrund von Rundungsungenauigkeiten durchaus zu kleineren Abweichungen von den Daten, die dem Originalbericht zugrunde lagen, kommen konnte.

Insgesamt konnten in 32 Publikationen 46 Experimente gefunden werden, deren Ergebnisse sich im Rahmen des Modellvergleichs verwenden ließen. Eine Auflistung dieser Studien kann in Tabelle 6.1 gefunden werden. Die Experimente wurden, wie in dieser Tabelle ersichtlich, in zehn Gruppen (A-J) eingeteilt. Diese Gruppen ergaben sich durch die Berücksichtigung von vier unterschiedlichen Informationen, bei denen es sich um

1. die Anzahl der experimentellen Faktoren;
2. die Anzahl der Faktorstufen für jeden experimentellen Faktor;
3. die Vollständigkeit des faktoriellen Designs sowie
4. die Parameterklasse, auf die sich der experimentelle Faktor plausibler Weise auswirken sollte, handelte.

Während sich die Anzahl der experimentellen Faktoren sowie die Anzahl der Faktorstufen für jeden Faktor selbst erklären, bedarf es für die beiden weiteren Punkte

einer Erläuterung. Die Vollständigkeit des faktoriellen Designs ergibt sich aus dem Verhältnis der Anzahl der Distraktorbedingungen zu der Anzahl der Zielreizbedingungen. War dieses Verhältnis ausgeglichen, d. h. existierte zu jeder Zielreizbedingung eine Distraktorbedingung, so wird von einem vollständigen Design gesprochen. Dies war allerdings keineswegs bei allen Untersuchungen der Fall. Vielmehr ergaben sich häufig Konstellationen, in denen weniger Distraktorbedingungen erhoben wurden als Zielreizbedingungen. Dies konnte u. a. daran liegen, dass eine Manipulation während der Lernphase vorgenommen wurde, so dass in der Abrufphase mehrere Listen gelernter Reize zusammen mit nur einer Distraktorliste präsentiert wurden. In diesem Falle wird das Design als unvollständig bezeichnet.

Tabelle 6.1: *Überblick über die für den Modellvergleich verwendeten Studien.*

| Kürzel | Probanden | Literaturangabe |
|---|-----------|--|
| Effekt auf Sensitivität | | |
| Gruppe A: Unvollständiges 1-Faktor-Design (2 Bedingungen) | | |
| A1 | 16 | Gardiner, Java und Richardson-Klavehn (1996, Exp. 1) |
| A2 | 16 | Karayianni und Gardiner (2003) |
| A3 | 20 | Mantyla (1997, Exp. 1) |
| A4 | 24 | Nega (2005, Exp. 1) |
| A5 | 24 | Nega (2005, Exp. 2) |
| A6 | 24 | Nega (2005, Exp. 3) |
| A7 | 24 | Nega (2005, Exp. 4) |
| Gruppe B: Vollständiges 1-Faktor-Design (2 Bedingungen) | | |
| B1 | 128 | Bastin und Van der Linden (2003) |
| B2 | 24 | Dewhurst (2001, Exp. 2) |
| B3 | 20 | Dewhurst und Parry (2000, Exp. 2) |
| B4 | 60 | Huron und Danion (2002) |
| B5 | 24 | MacRae et al. (2002) |
| B6 | 36 | Mantyla (1997, Exp. 3) |
| B7 | 54 | Propper und Christman (2004) |
| B8 | 40 | Wehr und Wippich (2004, Fig. 2a) |
| B9 | 40 | Wehr und Wippich (2004, Fig. 2b) |
| Gruppe C: Vollständiges 1-Faktor-Design (3 Bedingungen) | | |
| C1 | 11 | Huron et al. (2002) |
| C2 | 36 | Huron et al. (2001) |
| C3 | 137 | Xu und Bellezza (2001, Exp. 1) |
| Gruppe D: Vollständiges 2-Faktor-Design (2*2 Bedingungen) | | |
| D1 | 96 | Bastin et al. (2004) |

| Kürzel | Probanden | Literaturangabe |
|---|-----------|--|
| D2 | 20 | Dewhurst und Parry (2000, Exp. 1) |
| D3 | 36 | Gardiner, Kaminska et al. (1996, Exp. 3) |
| D4 | 48 | Gardiner et al. (1997) |
| D5 | 56 | Maylor und Mo (1999) |
| Gruppe E: Unvollständiges 2-Faktor-Design (2*2 Bedingungen) | | |
| E1 | 16 | Curran und Hildebrandt (1999) |
| E2 | 64 | D'Argembeau, Van der Linden, Comblain und Etienne (2003) |
| E3 | 64 | D'Argembeau und Van der Linden (2004) |
| E4 | 48 | Gardiner et al. (2005, Exp. 1) |
| E5 | 32 | Gardiner et al. (1999, Exp. 1) |
| E6 | 40 | Gardiner et al. (1999, Exp. 2) |
| E7 | 48 | Huron et al. (2003) |
| E8 | 32 | Konstantinou und Gardiner (2005) |
| E9 | 32 | Mantyla (1997, Exp. 4) |
| E10 | 42 | Sonntag et al. (2003) |
| E11 | 40 | Wehr und Wippich (2004, Tab. 1) |
| E12 | 40 | Wehr und Wippich (2004, Tab. 2) |
| Gruppe F: Vollständiges 2-Faktor-Design (2*3 Bedingungen) | | |
| F1 | 40 | Comblain et al. (2004) |
| F2 | 48 | Dewhurst und Anderson (1999, Exp. 1) |
| Effekt auf Kriterien | | |
| Gruppe G: Vollständiges 1-Faktor-Design (2 Bedingungen) | | |
| G1 | 24 | Dewhurst (2001, Exp. 3) |
| G2 | 18 | Eldridge et al. (2002) |
| Gruppe H: Vollständiges 1-Faktor-Design (3 Bedingungen) | | |
| H1 | 24 | Dewhurst (2001, Exp. 1) |
| H2 | 150 | Xu und Bellezza (2001, Exp. 2) |
| H3 | 24 | Dewhurst und Farrand (2004) |
| Effekte auf Sensitivität und Kriterien | | |
| Gruppe I: Unvollst. 2-Faktor-Design (2*2 Bedingungen) | | |
| I1 | 46 | D'Argembeau, Van der Linden, Etienne und Comblain (2003) |
| I2 | 32 | Ramponi et al. (2004) |
| Gruppe J: Vollst. 2-Faktor-Design (2*3 Bedingungen) | | |
| J1 | 48 | Dewhurst und Anderson (1999, Exp. 2) |

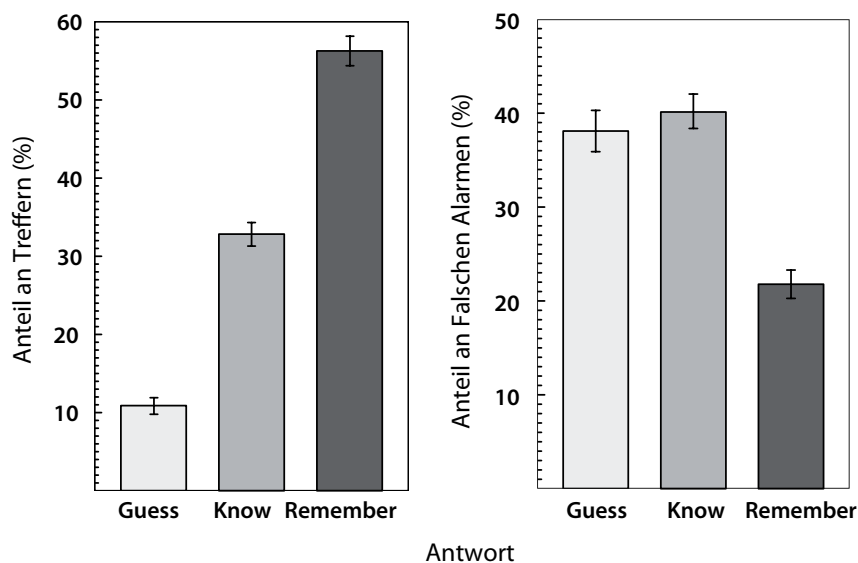


Abbildung 6.2: Dargestellt werden die mittleren prozentualen Anteile der Abruferebensarten an der Gesamtzahl positiver Rekognitionsentscheidungen mit 95%-Konfidenzintervallen.

Das unter Punkt 4 erwähnte Einteilungskriterium beruhte auf der Art der in den Untersuchungen realisierten Manipulationen. Diese wurden daraufhin untersucht, inwiefern von ihnen Effekte auf bestimmte Parameterklassen erwartet werden konnten. Eingeteilt wurden die Manipulationen dabei in zwei Gruppen, abhängig davon, ob ein Effekt auf die Gedächtnisleistung (Sensitivität) erwartet werden konnte oder ob für sie angenommen werden musste, dass sie die Antworttendenz (Kriterium) beeinflussten. Diese Einteilung wird in Tabelle 6.2 dargestellt.

Offensichtlich ließen sich deutlich mehr Manipulationen finden, von denen eine Wirkung auf die Gedächtnissensitivität erwartet werden konnte als solche, die einen Effekt auf Antwortkriterien postuliert werden musste. Insgesamt fanden sich unter den 27 realisierten Manipulationen nur fünf, für die Auswirkungen auf Antworttendenzprozesse behauptet werden konnten. Entsprechend musste auch nur für acht der 48 analysierten Datensätzen ein Effekt auf Antworttendenzen angenommen werden. Von diesen acht Untersuchungen waren wiederum nur fünf so konzipiert, dass ausschließlich Antworttendenzeffekte zu erwarten waren.

Die recherchierten Experimente umfassten insgesamt 148 Zielreiz- und 111 Distraktorbedingungen. Die absoluten Umfänge der Bedingungen, d. h. die Zahl von Beobachtungen, die in eine Bedingung einfließen, lässt sich Abbildung 6.1 entnehmen. Der Umfang einer Bedingung entspricht dabei dem Produkt der Anzahl der Items, die ihr zugeordnet werden konnten, und der Anzahl der Versuchspersonen, die diese Reize bearbeiteten. Abbildung 6.1 verdeutlicht, dass in den realisierten Bedingungen der Median für die Anzahl der Distraktoren mit 502 höher lag als der Median für die Ziel-

Tabelle 6.2: Die realisierten experimentellen Manipulationen wurden in zwei Gruppen eingeteilt: Solche, für die vermutet wurde, dass sie die Sensitivität beeinflussten, und solche, für die Effekte auf die Antworttendenz erwartet wurde.

| Beeinflusste Parameterklasse | |
|---|--|
| Sensitivität | Antworttendenz |
| Alter der Probanden (B1, D1, E3, F1) ^a | Anteil alter Reize an Testliste (D4, H2) |
| Antwortzeitbeschränkung (E4-6, E8) | Dysphorische Stimmung (I2) |
| Anzahl der Lerndurchgänge (D3) | Kategoriale Wiederholung (H1/3, J1) |
| Distinktheitsorientiertes Lernen (A3, B6, E9) | Kategoriengröße (G1) Soziale Ängstlichkeit (I1) |
| Emotionale Ladung (B3, D2, E2/3, F1, I1) | |
| Erwartungskonsistenz (B5) | |
| Exakte Stimuluswiederholung (F2) | |
| Händigkeit (B7) | |
| Lern-/Vergessensintention (E10) | |
| Lern-Test-Kongruenz (A4-7, D1/5, E4/9/11/12) | |
| Massierte Stimuluswiederholung (F2, J1) | |
| Retentionsintervall (C4) | |
| Salienz (B8/9) | |
| Schizophrenie (B4, E7/10) | |
| Sedativa (C2/3, E1) | |
| Typikalität (B2, D4) | |
| Verarbeitungstiefe (A1, E1/2/5/6/8, I2) | |

^aIn Klammern wird vermerkt bei welchen Untersuchungen die jeweiligen Manipulationen Verwendung fanden.

Tabelle 6.3: Entnehmbar sind relative Häufigkeiten, mit denen auf Zielreize respektive Distraktoren mit Remember(R), Know(K), Guess(G) bzw. Neu(N) reagiert wurde. Berichtet werden arithmetisches Mittel (M), Standardabweichung (SD) sowie Minimal- und Maximalwerte.

| | Zielreize | | | | Distraktoren | | | |
|------------|-----------|------|------|------|--------------|------|------|------|
| | R | K | G | N | R | K | G | N |
| <i>Min</i> | 0.05 | 0.07 | 0.00 | 0.03 | 0.00 | 0.00 | 0.01 | 0.20 |
| <i>M</i> | 0.40 | 0.22 | 0.07 | 0.31 | 0.05 | 0.08 | 0.08 | 0.79 |
| <i>SD</i> | 0.16 | 0.08 | 0.05 | 0.16 | 0.06 | 0.06 | 0.07 | 0.14 |
| <i>Max</i> | 0.80 | 0.60 | 0.31 | 0.74 | 0.42 | 0.33 | 0.37 | 0.99 |

reize (400). Gleiches galt, wenn auch nicht so eindeutig, für die mittlere Anzahl der Distraktoren (687.5, $SD = 543.23$) und die mittlere Anzahl der Zielreize (632.5, $SD = 748.62$). Minimal- und Maximalumfang von Distraktor- und Zielreizbedingungen hingegen waren nahezu identisch und betragen für Zielreize 132 und 4032 respektive 100 sowie 4000 Beobachtungen für Distraktorbedingungen.

Um aus den berichteten Daten die absoluten Antworthäufigkeiten für die einzelnen Antwortkategorien zu ermitteln, wurde in einem ersten Schritt aus den berichteten Daten für jede Antwortkategorie eine relative Häufigkeit bestimmt. Anschließend wurden die relativen Antworthäufigkeiten durch Multiplikation mit der absoluten Anzahl der Items in der jeweiligen Bedingung in absolute Häufigkeiten verwandelt. Ergaben sich dabei Ergebnisse mit Nachkommastelle, wurden diese probenhalber zu ganzen Zahlen auf- bzw. abgerundet. Resultierten daraus Häufigkeitstabellen, deren Gesamtsumme exakt die aufgrund von Probanden- und Itemzahl zu erwartenden Häufigkeiten widerspiegelte, so wurden diese Tabellen als Grundlage der späteren Modellierung verwendet. Im Falle von Abweichungen wurde das Runden wiederholt, um die aufgetretenen Ungenauigkeiten zu beseitigen. Dabei wurde per Zufall entschieden, bei welchen der infrage kommenden Kategorien zusätzlich auf- bzw. abgerundet werden sollte. Das Ergebnis wurde erst dann als Ausgangspunkt für die spätere Modellierung akzeptiert, wenn der Umfang der Häufigkeitstabelle dem auf der Basis von Probanden- und Itemzahl erwarteten Umfang genau entsprach. Die auf die geschilderte Weise abgeleiteten Antworthäufigkeiten können Anhang A entnommen werden.

Ein Überblick über die Häufigkeiten, mit denen auf Distraktoren respektive Zielreize mit Remember, Know, Guess bzw. Neu reagiert wurde, findet sich in Tabelle 6.3. Es zeigt sich hier insofern ein deutlicher Unterschied zwischen den beiden Reizklassen, als die Distraktoren mit einer mittleren Wahrscheinlichkeit von 0.79 ($SD = 0.14$) überaus häufig als solche zurückgewiesen wurden, während Zielreize weitaus häufiger zuerst in Alt-Antworten mündeten (0.69, $SD = 0.16$) und anschließend mit Remember-, Know- bzw. Guess-Antworten belegt wurden. Augenfällig ist, dass von diesen drei Möglichkeiten die Remember-Option mit einer mittleren Wahrscheinlichkeit von 0.40 ($SD = 0.16$) bei weitem am häufigsten gewählt wurde, während Know- (0.22, $SD = 0.08$) und Guess-Antworten (0.07, $SD = 0.05$) seltener erfolgten.

Des Weiteren offenbarten sich zwischen den Bedingungen insofern deutliche Unterschiede, als die Treffer-Raten sowie die Raten der Falschen Alarme über ein relativ breites Intervall variierten. Die entsprechenden Minimal- und Maximalwerte betragen 0.26 und 0.97 bei Treffern sowie 0.02 und 0.80 für die Falschen Alarme. Auch bei den minimalen und maximalen Wahrscheinlichkeiten von Remember-, Know- und Guess-Antworten war, wie Tabelle 6.3 entnommen werden kann, die Varianz zwischen den einzelnen Untersuchungen respektabel. Betrachtet man zusätzlich Abbildung 6.2, die zeigt, wie sich Treffer und Falsche Alarme auf die drei Abruferlebenskategorien aufteilten, so erhärtet sich der Eindruck, dass mit 56.27% aller Treffer das Gros selbiger in die Remember-Kategorie fiel, während sich die Falschen Alarme vorwiegend in der Know- (40.01%) und in der Guess-Kategorie (38.10%) wiederfanden.

Kapitel 7

Plausibilität der Parameterwerte

Im vorangegangenen Kapitel wurde die Datenbasis vorgestellt, die für den im Folgenden vorzustellenden Modellvergleich verwendet wurde. Wie gesehen werden konnte, bestanden die Datensätze aus insgesamt 148 Zielwortbedingungen, denen jeweils eine korrespondierende Distraktorbedingung zugeteilt werden konnte. Bei der Analyse der Rohdaten, d. h. der in den entsprechenden Arbeiten berichteten Häufigkeiten, mit denen von den Probanden Remember-, Know-, Guess- oder Neu-Antworten auf Zielreize oder Distraktoren produziert worden sind, wurde festgestellt, dass diese zwischen den einzelnen Untersuchungseinheiten deutlich variierten. Dieser Umstand stellt eine notwendige Bedingung für die im Folgenden vorzustellende Analyse dar, in der thematisiert werden sollte, inwiefern die konkurrierenden Modelle Parameterschätzungen lieferten, welche im Lichte des theoretischen Rahmens, auf den das Modell zurückgeht, plausibel erschienen.

Damit eine ausreichend große Anzahl verschiedener Parameterschätzungen gewonnen werden konnte, bediente sich diese Auswertung der kleinstmöglichen Dateneinheit, die mit Hilfe der zu evaluierenden Modelle analysiert werden konnte. Dabei handelte es sich um ein Datenfeld, das aus der Kombination der Antworthäufigkeiten einer Zielreizbedingung und den Antworthäufigkeiten der dazugehörigen Distraktorbedingung besteht. Diese Analyseeinheiten ergaben sich bei den Untersuchungen, die ein vollständiges Design aufwiesen, als diejenigen Zielreize und Distraktoren, die der gleichen Manipulationsbedingung angehörten. Für die Daten, die aus einem unvollständigen Versuchsplan hervorgegangen sind, gilt, dass die Distraktorbedingungen mehrmals verwendet wurden, damit jeder Zielreizbedingung eine Distraktorbedingung zur Seite gestellt werden konnte. In diesem Fall war den einander zugewiesenen Bedingungen gemeinsam, dass sie in derselben Abrufphase bearbeitet worden sind. Auf diese Weise ergaben sich insgesamt 148 Analyseeinheiten, bestehend aus insgesamt acht Reaktionshäufigkeiten: den Wahrscheinlichkeiten für Remember-, Know-, Guess- und Neu-Antworten für Zielwörter und für die jeweils zugehörigen Distraktoren.

An diese Analyseeinheiten wurden in einem nächsten Schritt alle sechs Modelle angepasst. Dabei wurde, wie in Kapitel 5.2 beschrieben, der größte aus zehn mit Hilfe des BFGS-Algorithmus bestimmten \mathcal{LL} -Werten als maximaler \mathcal{LL} -Wert und die in diesem Schätzdurchgang ermittelten Parameterwerte als ML -Schätzer festgehalten.

Außer für die beiden fünfparametrischen Modellen – SDM und HTSD – erübrigt sich eine Diskussion der Ergebnisse von Modellanpassungstests auf Basis der Likelihood-Quotienten-Statistik, da die übrigen Modelle die Häufigkeiten in den sechs freien Datenkategorien mit ebenso vielen Modellparametern beschrieben. Eine Betrachtung der Höhe der Parameterschätzwerte war hingegen für alle Modelle interessant und lieferte wichtige Anhaltspunkte für die Plausibilität der Modellannahmen. Insbesondere ließ sich für die Höhe der Parameterschätzwerte und deren Spannweite feststellen, inwiefern sie Werte einnahmen, die aufgrund ihres theoretischen Hintergrunds erwartet werden konnten, oder ob zwischen der Erwartung und den tatsächlichen Werten Diskrepanzen auftraten. Diese Fragestellung soll im Folgenden behandelt werden, indem ein Überblick über die Parameterschätzungen für alle 148 Analyseeinheiten gegeben wird. Die bei diesen Datenanpassungen erlangten Parameterschätzwerte können Anhang C entnommen werden.

7.1 Schätzungen für das SDM-Modell

Anders als bei den sechsparemetrischen Modellen, bei denen das Verhältnis von freien Datenkategorien zu der Anzahl der aus den Daten zu schätzenden Parameter ausgeglichen war, erlaubte die Datenlage für die fünfparametrischen Modelle SDM und HTSD eine Analyse der Datenanpassung anhand von Likelihood-Quotienten-Tests. Daher werden diese Tests im Rahmen der nächsten beiden Abschnitte diskutiert, bevor der Frage nach der Plausibilität der Parameterwerte nachgegangen wird. Dieses Vorgehen empfahl sich nicht zuletzt deshalb, weil eine ausreichend gute Datenanpassung durch das in Frage stehende Modell Voraussetzung dafür war, dass die Parameterschätzungen, die anhand des jeweiligen Modells gewonnen wurden, sinnvoll interpretiert werden konnten.

Bei der Anpassung des SDM-Modells an die Daten der 148 Analyseeinheiten, die aus den Antworthäufigkeiten je einer Zielreizbedingung und einer dazugehörigen Distraktorbedingung bestanden, fielen insgesamt 47 der Likelihood-Quotienten-Tests auf einem Signifikanzniveau von $\alpha = .05$ statistisch bedeutsam aus¹. Für 31.76% aller Analyseeinheiten konnten also statistisch bedeutsame Abweichungen zwischen den Modellvorhersagen und den tatsächlich beobachteten Daten festgestellt werden, die nicht durch Reparametrisierungen behoben werden konnten.

Für die weiteren Analysen, bei denen die Höhe der Parameterschätzwerte im Zentrum stand, wurden nur Schätzungen verwendet, die an Daten vorgenommen worden sind, zu deren Beschreibung sich das SDM-Modell als geeignet erwiesen hat. Die Datenbasis für die in Abbildung 7.1 dargestellten Parameterschätzwerte bestand also aus jenen 101 Analyseeinheiten, für die keine signifikanten Ergebnisse des Likelihood-Quotienten-Tests berichtet werden konnten. Die für jeden Parameter existierenden 101 Parameterschätzungen ergaben, wie in dieser Abbildung dargestellt, ein Gesamtbild, von dem man weitestgehend sagen kann, dass es den für dieses Modell zu erwartenden

¹Bei Annahme eines Signifikanzniveaus von $\alpha = .01$ vermindert sich die Anzahl der signifikanten Ergebnisse auf 33 (22.30%).

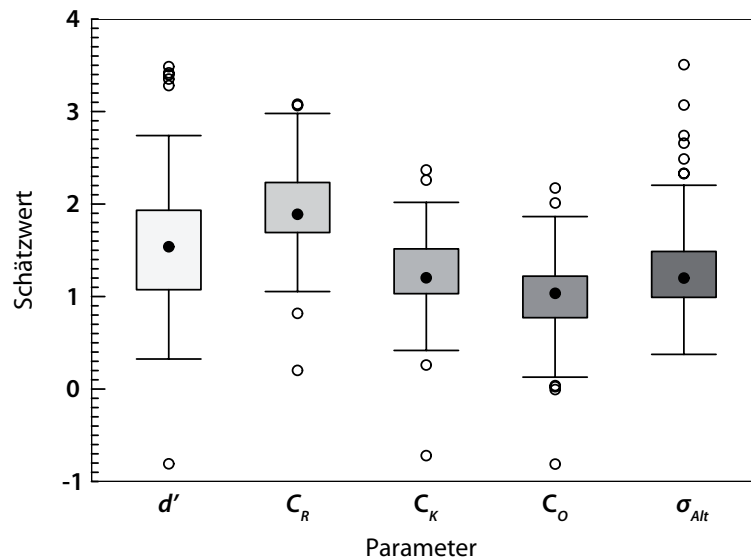


Abbildung 7.1: Dargestellt wird die Verteilung der anhand des SDM-Modells gewonnenen Parameterschätzungen. Berücksichtigt wurden alle 101 Bedingungen, bei denen der G^2 -Wert 3.84 betrug oder kleiner war.

Werten entsprach.

Die Kriteriumswerte fügten sich in die von den Modellannahmen vorgeschriebene Reihenfolge. C_R nahm mit 1.93 ($SD = 0.48$) im Mittel den höchsten Wert ein, darauf folgte das C_K -Kriterium mit 1.24 ($SD = 0.46$) und zuletzt C_G mit 0.97 ($SD = 0.49$). Vergleicht man diese Werte mit den mittleren Werten aus Dunns (2004) Metaanalyse über Remember-Know-Untersuchungen, die in Kapitel 3 behandelt wurde, so fällt auf, dass die Kriterien für die betrachteten Daten etwas konservativer gesetzt wurden als aufgrund der Daten von Dunn zu erwarten war. Das mittlere Remember-Kriterium wurde von Dunn bei 1.75 lokalisiert, während das Alt-Neu-Entscheidungskriterium bei 0.96 lag. Augenfällig ist, dass sich das Alt-Neu-Entscheidungskriterium verglichen mit dem von Dunn berichteten nur wenig verändert hat. Das Remember-Kriterium lag hingegen um einen Vertrauheitswert von 0.18 über dem von Dunn festgestellten Wert. Am deutlichsten ist wohl die Verschiebung des Know-Kriteriums, welches für die von Dunn behandelten Untersuchungen mit dem Alt-Neu-Kriterium zusammenfiel. Dieses war in den hier analysierten Daten um einen Betrag von 0.28 größer als der von Dunn ermittelte Wert.

Ein ähnlicher Vergleich ließ sich für die Schätzung der Gedächtnissensitivität anstellen. Der Wert des für das SDM-Modell relevanten Sensitivitätsparameters d' wich mit 1.58 ($SD = .72$) kaum von der konzeptuell verwandten Overall-Sensitivität d'_{RK} ab, die von Dunn auf 1.53 geschätzt wurde.

Betrachtet man den fünften Parameter, den das Modell bereithielt, um auf etwaige Streuungsunterschiede zwischen den Vertrauheitsverteilungen von Distraktoren und

Zielwörtern reagieren zu können, so wird ersichtlich, dass sich σ_{Alt} mit einem Mittelwert von 1.30 ($SD = .55$) in einem Bereich bewegte, der den Erwartungen über das Streuungsverhältnis von Distraktoren zu Zielreizen weitgehend entsprach. Bemerkenswert ist allerdings, dass immerhin ein Viertel der Schätzungen von σ_{Alt} einen Wert annahmen, der kleiner war als 1; insgesamt musste für 27 der 101 Datensätze davon ausgegangen werden, dass die Streuung der Distraktorverteilung entgegen der allgemeinen Erwartung kleiner war als die Streuung der zugehörigen Zielreizverteilung.

7.2 Schätzungen für das HTSD-Modell

Auch im Rahmen des HTSD-Modells werden, genau wie im zuvor diskutierten SDM-Modell, nur fünf Parameter genutzt, um das Abruferleben zu erklären. Dementsprechend wurden auch für das HTSD-Modell Datenanpassungsanalysen durchgeführt, um jene Analyseeinheiten zu identifizieren, die nicht mit dem Modell vereinbar waren. Von den dazu durchgeführten Likelihood-Quotienten-Tests waren insgesamt 66, also 44.59% auf einem Signifikanzniveau von $\alpha = .05$ signifikant². Genau wie im vorangegangenen Abschnitt stützen sich die folgenden Analysen nur auf die Parameterschätzungen der 82 Analyseeinheiten, für die bei $\alpha = .05$ kein signifikantes Abfallen der Datenanpassung beobachtet werden konnte. Einen Überblick über die für diese Untersuchungsbedingungen gewonnenen Parameterschätzungen bietet Abbildung 7.2.

Auch für diese Parameterschätzungen können die Ergebnisse der Metaanalyse von Dunn (2004) als Vergleichsmaßstab gelten, da sich das HTSD-Modell nur in der Annahme des durch den Parameter r erfassten Prozesses kontrollierter Erinnerung von dem einfachen Signalentdeckungsmodell unterscheidet, das Dunn zur Ermittlung seiner Schätzungen verwendete. Dieser Parameter r wurde im Mittel auf 0.16 geschätzt ($SD = 0.15$); demnach konnten Probanden in 16% aller Durchgänge mit Zielreiz die Reizpräsentation bewusst abrufen.

Im Vergleich zu dem von Dunn (2004) ermittelten mittleren d' fiel die auf dem HTSD-Modell beruhende Schätzung der Vertrautheit geringer aus: Sie lag für das HTSD-Modell im Mittel bei 1.20 ($SD = 0.54$), während Dunn einen Wert von 1.53 ermittelt hatte. Betrachtet man die Lage der Kriterien, so fällt auf, dass diese zu einem großen Teil konservativer waren als die von Dunn präsentierten Schätzungen. Das mittlere C_R lag mit 1.90 ($SD = 0.43$) über dem von Dunn berichteten Wert von 1.75 und das mittlere C_K war mit 1.19 ($SD = 0.45$) größer als das mittlere Alt-Neu-Kriterium in Dunns Arbeit (0.96). Einzig das Alt-Neu-Kriterium C_O lag mit einem Wert von 0.92 ($SD = 0.45$) knapp unter dem von Dunn ermittelten Kriteriumswert.

²Bei einem Signifikanzniveau von $\alpha = .01$ reduziert sich diese Zahl um 14 Analyseeinheiten auf 35.14%.

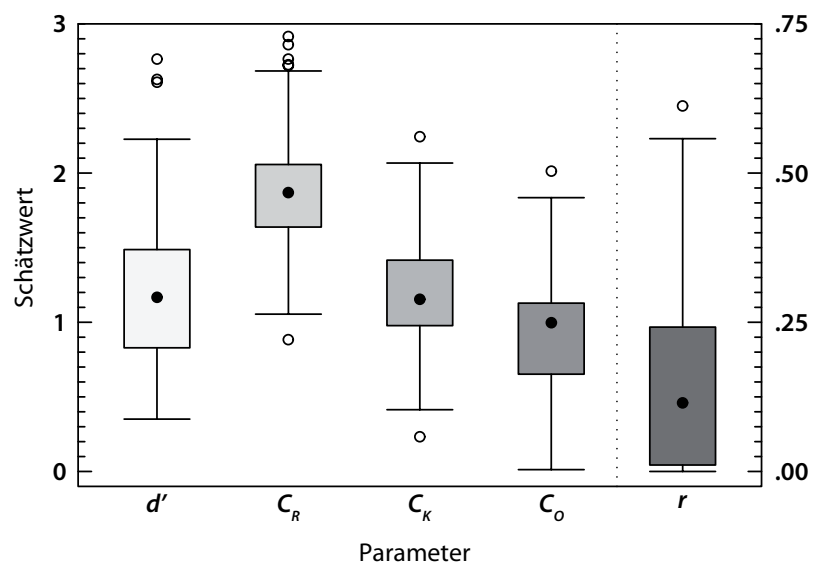


Abbildung 7.2: Dargestellt wird die Verteilung der anhand des HTSD-Modells gewonnen Parameterschätzungen. Berücksichtigt wurden jene 82 Bedingungen, bei denen der G^2 -Wert bei 3.84 oder darunter lag. Die Schätzungen für r werden auf der rechten Ordinatenachse abgetragen, während für die übrigen Parameter die linke Achse zur Orientierung dient.

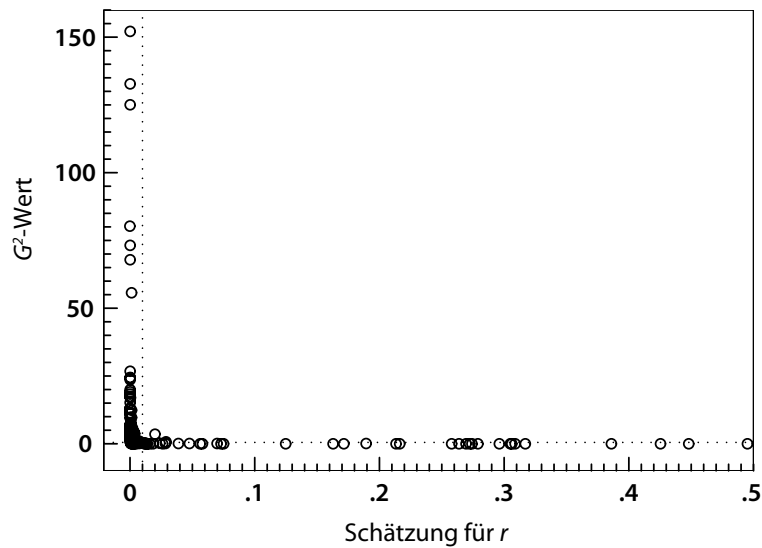


Abbildung 7.3: Für das HTUVSD-Modell wird die Verteilung von G^2 -Werten über den Schätzungen für r dargestellt. Gestrichelte Linien finden sich bei $G^2 = 0.5$ (waagrecht) und bei $r = .01$ (senkrecht).

7.3 Schätzungen für das HTUVSD-Modell

Im Unterschied zu den beiden vorangestellten Modellen sollte das Hochschwellen-Modell unter Annahme gleicher Varianzen nach der einfachen Zählregel zur Bestimmung der Freiheitsgrade optimal dazu geeignet sein, sich an die zu analysierenden Daten anzupassen. Bei der tatsächlichen Parameterschätzungen zeigten sich allerdings deutliche Abweichungen der Likelihood-Quotienten-Werte von ihrem Optimum. Die wahrscheinlichste Erklärung für dieses Anpassungsverhalten ist der Umstand, dass im Falle nicht optimaler Datenanpassung die Schätzung mindestens eines Parameters einen Wert liefert, der am Rand des für den jeweiligen Parameter geltenden Definitionsbereiches liegt. Von den sechs Parametern des HTUVSD-Modells haben vier einen Definitionsbereich, der die gesamte Menge aller reellen Zahlen \mathbf{R} umfasst: Das sind der Vertrautheitsparameter (d') und die drei Kriterien (C_R , C_K und C_G). Die anderen beiden Parameter weisen engere Definitionsbereiche auf; der Streuungsparameter σ_{Alt} muss positiv reellwertig sein, während der Parameter zur Erfassung kontrollierter Abrufprozesse (r) in dem Intervall zwischen 0 und 1 liegen muss.

Wenn man, wie in Abbildung 7.3 geschehen, den G^2 -Wert des Likelihood-Quotienten-Tests über den zugehörigen Schätzungen für den Parameter r abträgt, so fällt auf, dass hohe G^2 -Werte zu einem Großteil mit Schätzungen für r einhergingen, die nahe bei 0 lagen, und damit am linken Rand von dessen Definitionsbereich zu finden waren. Insgesamt fielen die Schätzungen von r bei 108 Datensätzen kleiner als .01 aus. Wurden diese Datensätze, wie durch die vertikale Linie in Abbildung 7.3 ange-

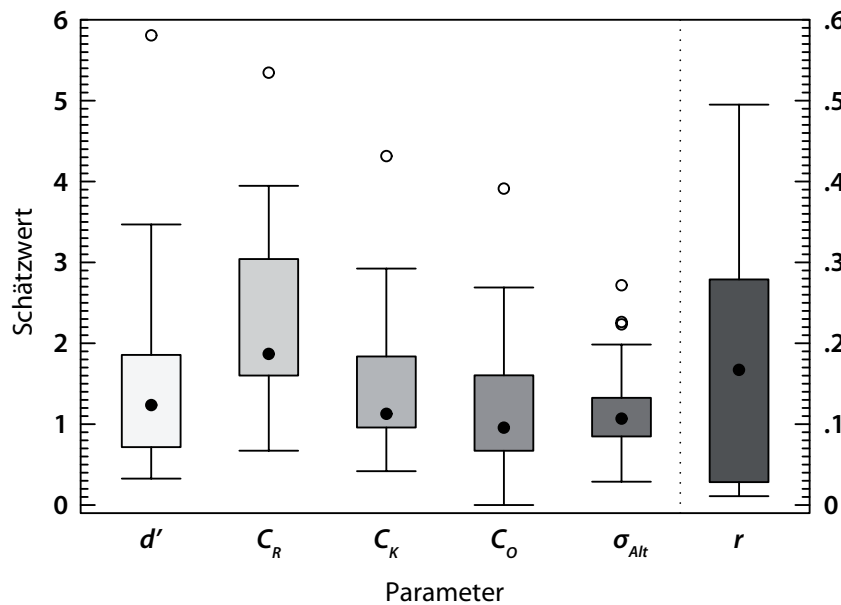


Abbildung 7.4: Dargestellt wird die Verteilung der anhand des HTUVSD-Modells gewonnen Parameterschätzungen. Berücksichtigt wurden die 38 Bedingungen, bei denen der G^2 -Wert kleiner als 0.5 war. Die Schätzungen für r werden auf der rechten Ordinatensachse abgetragen, während für die übrigen Parameter die linke Achse relevant ist.

deutet, von der Analyse ausgeschlossen, so reduzierte sich das mittlere G^2 von 8.17 ($SD = 22.23$) auf 0.13 ($SD = 0.57$). Unter den verbleibenden Datensätzen befanden sich dann nur noch zwei mit einem Wert im Likelihood-Quotienten-Test, der größer war als 0.5.

Bei zusätzlichem Ausschluss dieser beiden Datensätze verblieben 38 Datensätze, für die Parameterschätzungen anhand des HTUVSD-Modells gewonnen werden konnten. Das mittlere G^2 betrug für diese Datensätze 0.03 ($SD = 0.05$). In Abbildung 7.4 findet sich eine Darstellung der für diese 38 Datensätze durchgeführten Parameterschätzungen. Wie im Falle der beiden vorangegangenen Modelle bot sich auch hier ein Vergleich der Schätzwerte mit den von Dunn (2004) ermittelten Werten an. Zudem ließen sich die Parameterschätzungen sehr gut mit den beiden unmittelbar zuvor diskutierten Modellen vergleichen, da es sich bei beiden um Untermodelle dieses umfassenderen Modells handelte

Betrachtet man den Parameter r zur Erfassung kontrollierter Abrufprozesse, so fällt auf, dass dieser mit einem mittleren Wert von .17 ($SD = .14$) minimal höher ausfiel als die Schätzung kontrollierter Prozesse, die das HTSD-Modell geliefert hat. Würde man zusätzlich die 110 ausgeschlossenen Datensätze berücksichtigen, läge das mittlere r bei einem sehr kleinen Wert von .04 ($SD = 0.10$). Mit $d' = 1.50$ ($SD = 1.07$) fiel die HTUVSD-Schätzung der Sensitivität deutlich höher aus als die des HTSD-Modells. Mit diesem Wert blieb die Schätzung des HTUVSD-Modells nur minimal hinter der

des SDM-Modells und derjenigen von Dunn (2004) zurück.

Deutlich größer fielen hingegen die Abweichungen der Kriteriumsschätzer von denen Dunns sowie denen der beiden Tochtermodelle aus. Mit $C_R = 2.24$ ($SD = 1.02$), $C_K = 1.47$ ($SD = 0.86$) und $C_O = 1.20$ ($SD = 0.84$) erwiesen sich diese im Vergleich zu den Vergleichsschätzungen der übrigen Modelle und zu den Daten der Metaanalyse von Dunn (2004) als konservativer. Während sich das Verhalten der bisher diskutierten Parameter für die 38 analysierten Datensätze in dem Rahmen bewegten, den die anderen Modelle abgesteckt hatten, fiel der Streuungsparameter σ_{Alt} mit einem Wert von 1.17 ($SD = 0.52$) im Mittel sowohl im Vergleich zu der SDM-Schätzung als auch verglichen mit der allgemeinen Erwartung sehr hoch aus.

7.4 Schätzungen für das 4-States-Modell

Auch bei der Behandlung des 4-States-Modells empfiehlt sich vor der eigentlichen Erörterung der Parameterschätzungen eine Analyse des Datenanpassungsverhaltens. Dies ist insbesondere deshalb notwendig, da dieses Modell ausschließlich aus Parametern besteht, deren Definitionsbereiche auf das Intervall $[0; 1]$ beschränkt sind. Dieser Umstand machte es wahrscheinlich, dass Schätzungen für einzelne Datensätze auf den Rand des Parameterraumes fielen.

Der an allen 148 Analyseeinheiten gewonnene mittlere G^2 -Wert von 0.73 ($SD = 5.35$) deutete auf eine Abweichung der Datenanpassungsleistung vom Optimum hin. Diese Probleme ließen sich allerdings dadurch beheben, dass alle Datensätze ausgeschlossen wurden, bei denen mindestens ein Parameter auf .01 oder kleiner geschätzt wurde. Unter dieses Ausschlusskriterium fielen 21 Analyseeinheiten, so dass für die nachfolgenden Betrachtungen noch insgesamt 127 Datensätze zur Verfügung standen. Sowohl der an diesen Analyseeinheiten gewonnene mittlere G^2 -Wert als auch die zugehörige Standardabweichung verkleinerten sich merklich auf Werte, die ungefähr einem Zehntausendstel der Ausgangswerte entsprachen. Die auf dieser Basis gewonnenen Parameterschätzungen können in Abbildung 7.5 gefunden werden.

Zwischen dem 4-States-Modell und den übrigen Modellen konnten merklich weniger Parallelen gezogen werden als das für die bisher besprochenen Modelle möglich war. Hauptsächlich lag dies darin begründet, dass es sich bei dem 4-States-Modell, wie in Kapitel 4 erläutert worden ist, um ein reines Schwellenmodell handelt. Dadurch hebt es sich deutlich von den im Vorlauf diskutierten Modelle ab. Einzig in dem Parameter r des 4-States-Modells ließ sich eine Parallele zu den beiden Varianten des Yonelinas-Modells finden. Die Schätzung des ersten fiel allerdings mit einem mittleren Wert von 0.37 ($SD = 0.17$) ungefähr doppelt so hoch aus wie die Vergleichswerte, die HTSD- und HTUVSD-Modell lieferten. Für die übrigen beiden Gedächtnisparameter lagen hingegen keine Vergleichswerte vor. Ihre Schätzungen betragen für das Wissen ohne bewusste Erinnerung, die durch den Parameter k geschätzt wurde, 0.25 ($SD = 0.16$) und für die Distraktorentdeckung (d) 0.49 ($SD = 0.24$).

Augenfällig hoch fielen dabei die Schätzungen des Parameters d aus: Hiernach würde nahezu die Hälfte aller unbekanntten Reize bewusst als solche entdeckt. Dies

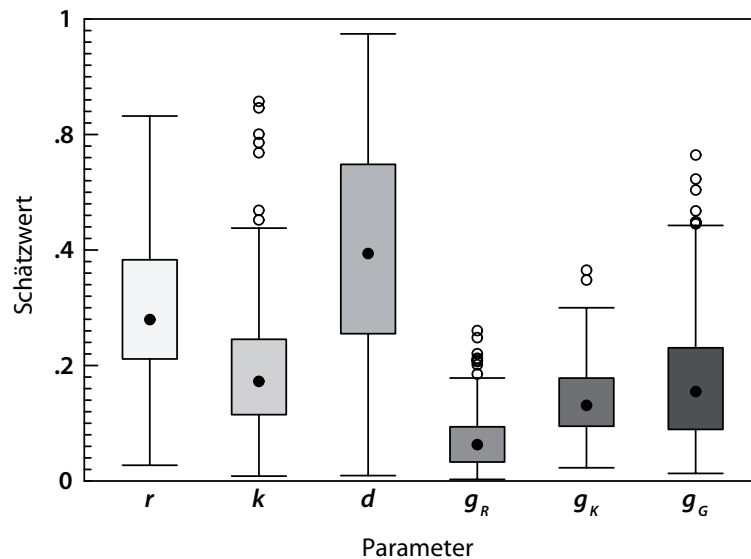


Abbildung 7.5: Dargestellt wird die Verteilung der anhand des 4-States-Modells gewonnen Parameterschätzungen. Berücksichtigt wurden die 127 Bedingungen, bei denen alle Parameterschätzungen größer als 0.1 waren.

steht in einem Missverhältnis zu der theoretischen Annahme, dass bewusste Distraktorentdeckung ein relativ seltenes Phänomen ist. Zudem wich es recht weit ab von dem oben bereits explizierten Schätzwert für das bewusste Erinnern, das bei nur 37.07% aller Zielreize auftrat. Eine Gleichsetzung der Parameter d und r , wie sie in verwandten Modellen häufig vorgenommen wird, erscheint dadurch für das 4-States-Modell unplausibel.

Betrachtet man die Schätzungen der Parameter, die Rateprozesse erfassen, so zeigt sich eine aufsteigende Funktion: Das mittlere g_R betrug 0.10 ($SD = 0.07$), etwas größer fiel g_K mit 0.18 ($SD = 0.08$) aus. Noch größer war nur der Wert für g_G mit 0.23 ($SD = 0.15$). Daraus lässt sich ableiten, dass die meisten der Items, die als Entscheidungen aus einem Unsicherheitszustand heraus gefällt wurden, mit Neu-Antworten versehen wurden (57.59%). Zudem wurde in nur 9.54% dieser Fälle Remember geraten. Die restlichen 32.78% der Entscheidungen, die aus einem Unsicherheitszustand resultierten, teilten sich relativ gleichmäßig auf die verbleibenden Antwortkategorien auf; etwas mehr als die Hälfte dieser, nämlich 16.86%, fielen dabei in die Guess-Kategorie.

7.5 Schätzungen für das Streak-Modell

Bei dem Streak-Modell handelt es sich um das erste der beiden zu behandelnden zweidimensionalen Signalentdeckungsmodelle. Anders als bei den bis zu diesem Zeitpunkt besprochenen Modellen, erschien es in Anbetracht der Datenanpassungsindizes nicht notwendig, vorab einzelne Analyseinheiten von dem weiteren Vorgehen auszuschlie-

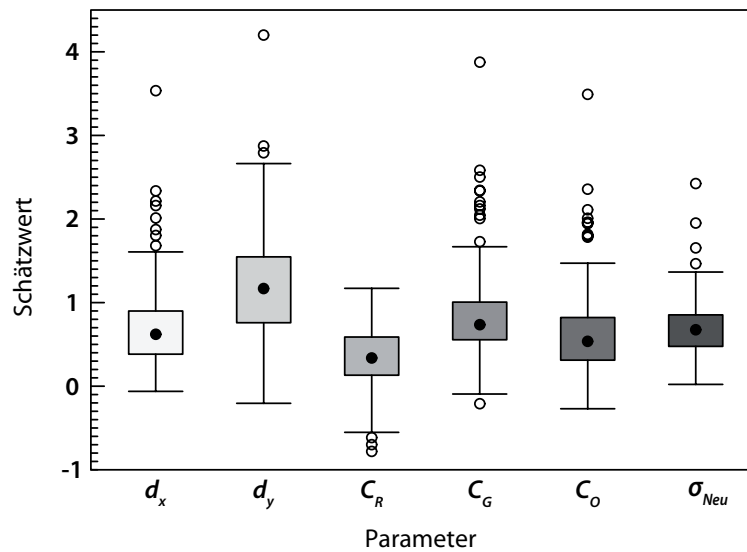


Abbildung 7.6: Dargestellt wird die Verteilung der anhand des Streak-Modells gewonnen Parameterschätzungen für alle 148 Untersuchungsbedingungen.

ßen; selbst der größte der an den 148 Analyseeinheiten bestimmten G^2 -Werte lag mit 0.04 sehr nah bei dem zu erwartenden Optimum von 0. In die durch Abbildung 7.6 veranschaulichten Parameterschätzungen flossen deshalb die Informationen aller 148 Datensätze ein.

Da das Streak-Modell die Gedächtnissensitivität in zwei Komponenten aufgliedert, ließen sich die mit Hilfe dieses Modells gewonnenen Parameterschätzungen nicht direkt mit den Schätzungen vergleichen, die anhand der eindimensionalen Signaldeckungsmodelle gewonnen wurden. Damit diese Werte dennoch miteinander verglichen werden konnten, bot sich eine Transformation an, die die beiden Sensitivitätsindizes d_x und d_y , die das Streak-Modell bereithält, um die Sensitivität des Gedächtnisses zu erfassen, in einen einzigen Wert verwandelte, der dem sonst üblichen Sensitivitätsmaß d' von der Bedeutung annähernd entspricht. Diese Bedingung erfüllt am ehesten die Länge der Strecke vom Mittelwert der Zielreizverteilung $M = (d_x | d_y)$ zum Koordinatenursprung (O). Die Länge dieser Strecke ließ sich anhand der in Gleichung 7.1 niedergelegten Rechenvorschrift bestimmen und konnte anschließend mit Schätzungen der Modelle verglichen werden, die d' als Sensitivitätsparameter verwendeten.

$$\overline{OM} = \sqrt{d_x^2 + d_y^2} \quad (7.1)$$

Der Wert dieses abgeleiteten eindimensionalen Sensitivitätsindex lag mit einem mittleren Wert von 1.42 ($SD = 0.77$) um einen Betrag von 0.11 bzw. 0.15 unter den Werten der vergleichbaren Schätzungen von Dunn (2004) und der auf der Basis des SDM-Modells ermittelten d' -Schätzung. Zudem fiel der abgeleitete Sensitivitätsindex zwischen die Schätzungen für d' , die auf der Basis der beiden diskutierten hybriden

Signalentdeckungsmodelle HTSD und HTUVSD gewonnen wurden.

Betrachtet man die Sensitivitätsschätzungen im einzelnen so fällt auf, dass die Schätzung der globalen Spurstärke d_x mit einem mittleren Wert von 0.74 ($SD = 0.50$) deutlich geringer ausfiel als die der spezifischen Spurstärke d_y , die bei 1.19 ($SD = 0.63$) lag.

Der dritte prinzipiell von der Gedächtnisleistung abhängige Modellparameter σ_{Neu} lag im Mittel bei 0.70 ($SD = 0.35$). Dieser Wert unterschritt den Wert von 0.80, der als Standard verwendet wird, wenn für Daten aus Remember-Know-Experimenten Parameter geschätzt werden (vgl. Rotello et al., 2004), um 12.5%.

Auch bei der Analyse der Kriteriumsparameter ist zu beachten, dass sie nicht mit den Parametern der bisher betrachteten Signalentdeckungsmodelle vergleichbar waren. Das Kriterium C_O , das zur Trennung zwischen alten Stimuli und Distraktoren verwendet wird, lag bei 0.64 ($SD = 0.54$), das Kriterium C_G , welches Guess-Antworten von Remember- und Know-Antworten abhebt, bei 0.86 ($SD = 0.56$) und das Kriterium C_R , anhand dessen zwischen Remember- und Know-Antworten unterschieden wird, bei 0.34 ($SD = 0.39$).

7.6 Schätzungen für das SAC-Modell

Das SAC-Modell stellt das zweite mehrdimensionale Signalentdeckungsmodell unter den Remember-Know-Guess-Modellen dar. Auch für dieses Modell musste keine der 148 Analyseeinheiten von den weiteren Betrachtungen ausgeschlossen werden: Der größte G^2 -Wert war deutlich kleiner als 0.01. Daher bilden die Parameterschätzungen für alle 148 Datensätze die Grundlage des in Abbildung 7.7 dargestellten Überblicks.

Ähnlich wie die Schätzungen für d_x und d_y , die auf der Basis des Streak-Modells gewonnen wurden, zeichneten sich die auf der Basis des SAC-Modells gewonnenen Schätzungen selbiger Parameter dadurch aus, dass die Schätzung für d_y mit einem Wert von 1.23 ($SD = 0.90$) höher ausfiel als die für d_x , die bei 0.90 ($SD = 0.76$) lag. Auch diese Sensitivitätsmaße ließen sich nicht direkt mit den Schätzungen für d' vergleichen, da letzteren – wie bereits erläutert – ein eindimensionales Sensitivitätskonstrukt zugrunde lag. Anhand der in Gleichung 7.1 festgehaltenen Transformation war es allerdings auch für die anhand des SAC-Modells bestimmten Sensitivitätsmaße möglich, ein eindimensionales Äquivalent zu bestimmen, das mit dem üblichen Sensitivitätsmaß d' verglichen werden konnte. Im Mittel lag dieser eindimensionale Sensitivitätswert mit einem Betrag von 1.62 ($SD = 1.06$) höher als alle vergleichbaren Werte. Allerdings war die Differenz zum größten der Vergleichswerte – dem d' des SDM-Modells, das einen Wert von 1.57 annahm – nur gering.

Der Streuungsparameter σ_{Neu} lag mit einem Mittelwert von 0.82 ($SD = 0.36$) sehr nahe bei dem Standardwert von 0.8, der häufig zur Modellierung von Remember-Know-Daten benutzt wird. Ganz im Gegenteil zu den üblicherweise getroffenen Annahmen nahm der Streuungsparameter σ_{Neu} jedoch bei immerhin einem Viertel der Analyseeinheiten Werte an, die über 1 lagen. In diesen Fällen fiel also die Streuung der Zielreizverteilung deutlich größer aus als die der Distraktorverteilung. Dieses Phäno-

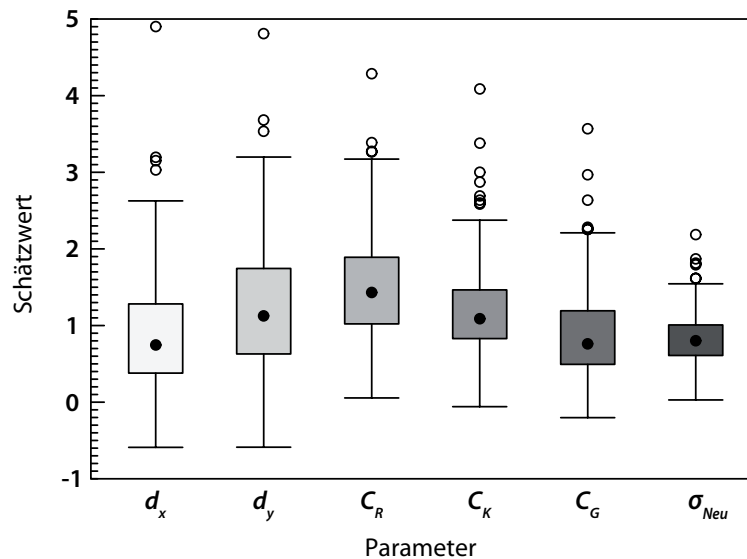


Abbildung 7.7: Dargestellt wird die Verteilung der anhand des SAC-Modells gewonnen Parameterschätzungen für alle 148 Bedingungen.

men wurde schon bei der Diskussion des SDM-Modells in ähnlichem Umfang beobachtet.

Die Schätzer der Kriterien lagen bei 1.52 ($SD = 0.64$) für das mittlere C_R -Kriterium, über dem die Episodenaktivierung liegen musste, damit Remember-Antworten gegeben wurden, 1.21 ($SD = 0.61$) für das mittlere C_K -Kriterium und 1.19 ($SD = 0.36$) für das mittlere C_G -Kriterium. Diese Schätzungen ließen sich wegen der deutlichen Unterschiede hinsichtlich ihrer Konzeptionalisierung nur schwer mit den Schätzungen der Antworttendenzmaße anderer signalentdeckungstheoretischer Modelle vergleichen.

7.7 Diskussion

In diesem Kapitel sind zwei wichtige Beobachtungen berichtet worden, die für den Modellvergleich von Relevanz sind. Zum einen wurden die Schätzungen der Modellparameter betrachtet, wie sie auf der Basis der verschiedenen Modelle gewonnen worden sind. Diese Schätzwerte wurden an den Daten der 148 Analyseeinheiten bestimmt, die aus jeweils vier Antworthäufigkeiten für eine Gruppe von Zielreizen und den Antworthäufigkeiten einer korrespondierenden Gruppe von Distraktoren bestanden. Bei der Analyse der Schätzwerte wurde allerdings nicht nur deutlich, wie sich diese über die Analyseeinheiten verteilten, zudem zeigte sich, dass sich die Modelle hinsichtlich ihrer Fähigkeit unterscheiden, die Daten der einzelnen Analyseeinheiten zu beschreiben. Während der erste dieser beiden Punkte zentral für den Inhalt dieses Kapitels war, weist der zweite Punkt voraus auf das nächste Kapitel, das sich mit der deskriptiven

Adäquatheit der Modelle beschäftigen wird. Dennoch sollen bei der Diskussion der in diesem Kapitel gewonnenen Erkenntnisse beide Punkte berücksichtigt werden.

Betrachtet man die in den durchgeführten Analysen an den Tag gelegte Datenanpassungsleistung der Modelle, so fällt auf, dass zwischen den Modellen deutliche Unterschiede bestanden. Augenfällig wurde dies an der Anzahl der Datensätze, die bei der Ermittlung der Schätzwertverteilungen berücksichtigt werden konnten. Während die beiden zweidimensionalen Signalentdeckungsmodelle durchweg so gute Datenanpassungswerte lieferten, dass für sie jeder Datensatz in die weiteren Analysen überführt werden konnte, zeigten sich bei den übrigen Modelle Defizite hinsichtlich der Datenanpassungsleistung. Wie viele Datensätze von den Analysen ausgeschlossen werden mussten, schwankte zwischen den einzelnen Modellen. Während die Ausschussrate für das 4-States-Modell mit 14.19% noch vergleichsweise gering ausfiel, stieg sie für das HTUVSD-Modell auf 74.32%. In beiden Fällen lag die Notwendigkeit für den Ausschluss von Datensätzen darin begründet, dass die Parameterschätzungen auf den Rand des Parameterraums fielen, was wiederum zur Folge hatte, dass die Datenanpassungsleistung – gemessen als Wert der Likelihood-Quotienten-Statistik G^2 – den für sie zu erwartenden Wert von 0 nicht erreichte. Für die beiden fünfparametrischen Modelle ließ sich ein klareres Ausschlusskriterium definieren, da die Umstände der Modellierung einen statistischen Test der Datenanpassungsleistung zuließen. In diesen Tests zeigte sich, dass das HTSD-Modell für die Beschreibung von 44.59% der Datensätze nicht geeignet war, während das SDM-Modell sich in 31.76% der Datensätze nicht bewährte.

Wenn man von diesen Ergebnissen vorläufig auf die deskriptive Adäquatheit der Modelle schließen möchte, so nahmen die beiden fünfparametrischen Modelle einen mittleren Platz ein: Sie lagen zwar hinter den beiden zweidimensionalen Signalentdeckungsmodellen und dem 4-States-Modell, schnitten aber besser ab als das um die Annahme ungleicher Varianzen ergänzte Yonelinas-Modell.

Eine zentralere Rolle als die Beobachtung, dass die meisten Modelle nicht mit allen Datensätzen vereinbar waren, kommt in diesem Kapitel den Parameterschätzungen zu. Besonders wichtig war es dabei zu betrachten, inwieweit diese Parameterwerte Annahmen, die für plausibel gehalten werden können. Entsprechend wird im Folgenden auf die Beobachtungen eingegangen werden, bei denen sich die Parameterschätzungen deutlich von den für sie formulierbaren Erwartungen unterscheiden. In einem solchen Sinne auffällige Parameterwerte konnten bei der Anpassung der Modelle verschiedentlich beobachtet werden, wobei sich die auftretenden Probleme relativ gleichmäßig über alle Modelle verteilten. So fielen die Schätzungen der Streuungsparameter, wie sie von manchen der Modelle mit signalentdeckungstheoretischer Fundierung angenommen wurden, für die meisten dieser Modelle nur im Mittel so aus, wie es erwartet wird. Bei einer nicht zu vernachlässigenden Zahl von Datensätzen zeigte sich bei den Modellen Streak, SAC und SDM, dass das Varianzverhältnis ins Gegenteil verkehrt war. Aufgrund der von diesen Modellen vorgenommenen Schätzungen müsste also darauf geschlossen werden, dass in jenen Fällen die Varianz der Distraktorverteilung größer ausfiel als die der Zielreizverteilung.

Auch bei dem HTUVSD-Modell wich der Streuungsparameter von den theo-

retischen Vorhersagen ab. Hier bestand das Problem darin, dass die mittlere Varianzschätzung sehr groß ausfiel. In der Literatur geht man davon aus, dass das Varianzverhältnis $\sigma_{Neu}/\sigma_{Alt}$ ungefähr 4/5 beträgt (vgl. Ratcliff et al., 1992; Rotello et al., 2004); die mit Hilfe des HTUVSD-Modells erlangte Schätzung des Verhältnisses lag mit 0.86 über diesem Wert.

Ebenfalls zu hoch – verglichen mit den allgemeinen Erwartungen – fiel die Schätzung des Distraktorentdeckungsprozesses d aus, der im Rahmen des 4-States-Modells angenommen wird. In der Regel wird dieser Prozess verstanden als eine bewusste Zurückweisung von Distraktoren, die recht selten vorkommen sollte. Gleichzeitig lag der Schätzwert für Prozesse der bewussten Erinnerung für das 4-States-Modell mit einem Mittelwert von .37 deutlich über den entsprechenden Schätzungen der Yonelinas-Modelle. Letztere schätzten kontrollierte Prozesse auf Werte von .16 im Falle des HTSD-Modells respektive .17 für das HTUVSD-Modell. Zusammengekommen legen diese Beobachtungen die Vermutung nahe, dass die anhand des 4-States-Modells gewonnenen Schätzungen kontrollierter Prozesse nach oben verzerrt sind. Eine mögliche Ursache für diesen Unterschied liegt in dem Umstand begründet, dass im Gegensatz zum 4-States-Modell beide Varianten des Yonelinas-Modells Vertrauenseinflüsse auf Remember-Antworten erlauben.

Weitere Vergleiche von Schätzwerten konzeptuell verwandter Parameter, die in unterschiedliche Modelle eingebettet sind, ließen sich bezüglich der Sensitivitätsparameter anstellen. Hier ermöglichte die Transformation der beiden Sensitivitätswerte d_x und d_y , wie sie von den beiden zweidimensionalen Signalentdeckungsmodellen angenommen werden, in ein eindimensionales Sensitivitätsmaß den Vergleich der Sensitivitätsschätzwerte untereinander und mit einem Außenkriterium. Als Außenkriterium dienten dabei die Schätzungen der Gedächtnissensitivität, die Dunn (2004) in seiner Metaanalyse über Remember-Know-Modelle vorgelegt hat.

Hierbei ließ sich beobachten, dass nahezu alle Schätzwerte in einem engen Bereich um den Wert von 1.53 lagen, den Dunn (2004) berichtet hat. Einzig das HTSD-Modell verfehlte diese Marke deutlich.

Die zuletzt geschilderten Vergleiche der Parameterschätzungen untereinander und mit geeigneten Außenkriterien sind insofern interessant, als sie zeigen, inwiefern die Bestimmung der Konstrukte, die von mehreren Modellen thematisiert werden, in vergleichbare Ergebnisse mündeten. Als Argument für oder gegen die Modellgültigkeit können solche Beobachtungen allerdings nicht gewertet werden, da in dieser Hinsicht ausschließlich Beobachtungen ins Gewicht fallen, die darauf hindeuten, dass Parameterschätzungen anders ausfallen, als es aufgrund ihrer theoretischen Einbettung zu erwarten ist. In diesem Sinne kritisch sind die Beobachtungen, dass sich das Varianzverhältnis bei allen Modellen mit signalentdeckungstheoretischer Fundierung für einen substantiellen Teil der Schätzungen umkehrt sowie die Beobachtung sehr hoher Schätzwerte für die Distraktorentdeckung, wie sie das 4-States-Modell zu Tage brachte.

Es kann an dieser Stelle festgehalten werden, dass sich mindestens ein Parameter aller für die Remember-Know-Guess-Prozedur vorgeschlagenen Modelle außerhalb des für ihn plausiblen Bereichs bewegte. Dieses Ergebnis muss jedoch nicht unbedingt

als Ausdruck fehlender Modellplausibilität gewertet werden – es ist auch denkbar, dass diese Abweichungen von den theoretisch zu erwartenden Parameterschätzungen zumindest zum Teil Besonderheiten der verwendeten Datensätze zugeschrieben werden können.

Kapitel 8

ROC-Analyse

Wie bereits in Kapitel 3 deutlich wurde, liefert der Vergleich empirischer ROC-Kurven mit ROC-Kurven, die auf Modellvorhersagen beruhen, Informationen über die Adäquatheit der Modelle. Bei der Betrachtung von ROC-Kurven wesentlich ist die Frage danach, inwiefern die Form empirischer ROC-Kurven mit der Struktur des Modells vereinbar ist (vgl. bspw. Wixted, 2007). ROC-Analysen werden daher im Rahmen dieser Arbeit als Methode genutzt, um die deskriptive Adäquatheit der Modelle zu betrachten, ohne auf Maximum-Likelihood-Verfahren zurückzugreifen. Entsprechend weisen die in diesem Kapitel erlangten Ergebnisse voraus auf jene, die im folgenden Kapitel berichtet werden.

Für die in diesem Kapitel durchgeführten Vergleiche ist der Umstand von zentraler Bedeutung, dass sich aus den Ergebnissen von Remember-Know-Guess-Untersuchungen nicht nur Zwei-Punkt-ROCs ableiten lassen, wie es für Daten, die in Remember-Know-Untersuchungen gewonnen wurden, möglich ist, sondern ROC-Kurven, die aus drei Datenpunkten bestehen. Damit erlauben Remember-Know-Guess-Daten – wie bereits in Kapitel 3 angeklungen – die Steigungen konfidenz- und abrufenerlebensbasierter ROC-Geraden in einer einzigen empirischen Untersuchung zu thematisieren. Zusätzlich liefern Drei-Punkt- z -ROC-Kurven genug Daten, um ihre Krümmung feststellen zu können.

Daher lässt sich anhand von Remember-Know-Guess-Daten zum einen überprüfen, inwiefern sich abrufenerlebensbasierte z -ROC-Geraden hinsichtlich ihrer Steigung von konfidenzbasierten unterscheiden. Zum anderen erlauben sie, zu untersuchen, inwiefern die für das SDM-Modell zentrale Vorhersage linearer z -ROC-Kurven von empirischen Daten erfüllt wird (vgl. Ratcliff et al., 1992). Diese Vorhersage über die Krümmung der z -ROC-Kurven wird im Verlaufe des Kapitels kontrastiert mit entsprechenden Vorhersagen, die aus den Modellgleichungen der übrigen Modelle mit Hilfe eines parametrischen Bootstrap-Verfahrens abgeleitet wurden. Wie gezeigt wird, lassen sich diese Modelle in zwei weitere Gruppen einteilen: Während die hybriden Modelle, die auf die Arbeit von Yonelinas et al. (1996) zurückgehen, umgekehrt u-förmige z -ROC-Kurven vorhersagen (vgl. auch Yonelinas, 1999), lassen die übrigen drei Modelle, wie die darzustellenden Simulationen ergaben, positiv beschleunigte z -ROC-Kurven erwarten.

Neben einer Diskussion der auf Remember-Know-Guess-Daten basierenden Drei-Punkt- z -ROC-Kurven unter dem Gesichtspunkt ihrer Krümmung werden die vorhandenen Daten zusätzlich bezüglich etwaiger Steigungsunterschiede zwischen z -ROC-Kurven, die auf Konfidenzurteilen in einfachen Rekognitionsexperimenten beruhen, und solchen, die auf in Remember-Know-Untersuchungen erhobenen Abruferlebensurteilen basieren, beleuchtet (vgl. Kapitel 3). Hierzu wurden aus den drei zur Verfügung stehenden ROC-Punkten zwei einzelne ROC-Geraden konstruiert. Während die eine dieser beiden Geraden auf zwei ROC-Punkten beruht, die sich hinsichtlich der Sicherheit unterscheiden, mit der Alt-Antworten gegeben wurden, beruht die andere Gerade auf Punkten, die sich hinsichtlich der Art des berichteten Abruferlebens unterscheiden. Die erste Gerade wird dabei aus korrekten und falschen Alt-Antworten einerseits und aus korrekten und falschen Remember- und Know-Antworten andererseits konstruiert und soll als Analog zu Konfidenzurteilen in Rekognitionsuntersuchungen dienen. Die zweite Gerade hingegen beruht auf Alt-Antworten einerseits und Remember-Antworten andererseits und soll in Analogie zu ROC-Kurven auf der Basis von Remember-Know-Daten stehen.

Die Untersuchung der Modellgüte mit Hilfe von ROC-Analysen wird in vier Abschnitten vorgestellt. Im ersten dieser Abschnitte wird auf die Analyse von ROC-Kurven für Remember-Know-Guess-Datensätze eingegangen, wobei insbesondere die Interpretation von Drei-Punkt- z -ROC-Kurven im Mittelpunkt stehen wird.

Im zweiten Abschnitt werden die Ergebnisse einer Simulationsstudie geschildert, die dazu diente, für alle Modelle Vorhersagen über die Form der z -ROC-Kurven zu generieren. Dabei werden zwei Punkte von besonderem Interesse sein: Zum einen ein Winkelmaß, das in direkter Beziehung zu der Krümmung der z -ROC-Kurven steht und zum anderen die Steigungswerte, welche sich für konfidenzbasierte bzw. abruferelebensbasierte Zwei-Punkt- z -ROC-Kurven ergeben.

In einem dritten Abschnitt werden dann die anhand von empirischen Daten ermittelten z -ROC-Kurven in einer Weise auf Krümmung und Steigung untersucht, die einen direkten Vergleich mit den Vorhersagen der einzelnen Modelle erlauben.

Im letzten Abschnitt werden die im zweiten und dritten Abschnitt erlangten Ergebnisse diskutiert und dahingehend verglichen, inwiefern sich Übereinstimmungen zwischen den empirischen Resultaten und den Vorhersagen der Modelle zeigen.

8.1 ROC-Kurven für RKG-Untersuchungen

Die Daten aus Remember-Know-Guess-Experimenten lassen sich, wie bereits erläutert, durch Drei-Punkt- z -ROC-Kurven darstellen. Dabei ergeben sich die einzelnen Punkte

1. aus der alleinigen Berücksichtigung von korrekten sowie falschen Remember-Antworten (R);
2. bei Berücksichtigung von richtigen respektive falschen Remember- und Know-Antworten (RK) und

3. unter Berücksichtigung aller Alt-Urteile (A).

Diese drei Punkte lassen sich ohne Informationsverlust durch zwei Geraden wiedergeben, die einen der drei Punkte teilen und durch je einen der beiden übrigen Punkte laufen. Anhand dieser Geraden lässt sich die Frage nach der Krümmung der z -ROC-Kurve klären, indem der Schnittwinkel betrachtet wird, der im Schnittpunkt der Geraden anfällt. Gleichzeitig ist es möglich, die Steigungen dieser beiden Geraden zu ermitteln.

Während für die Ermittlung der Krümmung unwesentlich ist, welcher der drei Punkte als Schnittpunkt der beiden Geraden verwendet wird, hängt die Interpretierbarkeit der Steigung von der Wahl des Schnittpunktes ab. Um eine inhaltlich sinnvolle Interpretation der Steigungswerte zu ermöglichen, wurde der Schnittpunkt so gewählt, dass die Punkte der einen Gerade sich nur hinsichtlich der Konfidenz – sichere (Z_{RK}) vs. weniger sichere (Z_A) Alt-Antworten – unterscheiden, während sich die Punkte der zweiten Gerade hinsichtlich des berichteten Abruferlebens – Remember-Antworten (Z_R) versus Alt-Antworten (Z_A) – unterscheiden.

Bei der Konstruktion beider Geraden wurde also der aus korrekten und falschen Alt-Antworten resultierende Punkt (Z_A) berücksichtigt. Von dort läuft die abruferlebensbasierte Gerade durch den ausschließlich auf den Remember-Antworten beruhenden Punkt Z_R , während die konfidenzbasierte Gerade durch den Punkt Z_{RK} läuft, in dem die Summe der Remember- und Know-Treffer zur Summe der entsprechenden Falsche-Alarm-Raten in Beziehung gesetzt wird.¹

Wie bereits erwähnt, können diese beiden Geraden nicht nur dazu genutzt werden, um Unterschiede zwischen den Steigungen konfidenz- und abruferlebensbasierter z -ROC-Geraden zu ermitteln. Zusätzlich kann aus der Lage dieser beiden Geraden zueinander die Krümmung der gesamten z -ROC-Kurve bestimmt werden. Als ein Maß für die Krümmung der aus drei Punkten bestehenden z -ROC-Kurve bietet sich der Winkel φ an, der zwischen der abruferlebensbasierten Geraden $\overline{Z_A Z_R}$ und der konfidenzbasierten Geraden $\overline{Z_A Z_{RK}}$ anfällt. Dieser Schnittwinkel (φ) lässt sich anhand der in Abbildung 8.1 dargestellten Größen bestimmen. Da der Sinus von φ dem Verhältnis aus der Länge des Lots $\overline{Z_R Z'_R}$ von Z_R auf $\overline{Z_A Z_{RK}}$ zu der Länge der Strecke $\overline{Z_A Z_R}$ multipliziert mit dem Vorzeichen der Differenz der Abszissenwerte von Z_R und Z'_R entspricht, lässt sich φ wie in Gleichung 8.1 festgehalten bestimmen:

$$\varphi = \text{sgn}(zFA_R - zFA'_R) \cdot \arcsin \left(\frac{\| \overline{Z'_R Z_R} \|}{\| \overline{Z_A Z_R} \|} \right) \quad (8.1)$$

Von den Werten, die φ annimmt, lässt sich leicht auf die Krümmung der Dreipunkt- z -ROC-Kurve schließen. Der Winkel φ nimmt nämlich den Wert 0 an, wenn sich die beiden Geraden zu einer einzigen vereinen, während sich eine Krümmung der

¹Eine zweite Konstellation wäre mit dem Ziel, zwei Geraden zu konstruieren, von denen die eine ausschließlich auf Konfidenzunterschieden beruht, während die zweite ausschließlich durch Unterschiede hinsichtlich des Abruferlebens gekennzeichnet ist, vorstellbar. Diese Fassung würde sich ergeben, wenn Z_{RK} als Schnittpunkt gewählt würde. Im Allgemeinen würde sich diese Variante nicht sehr von der gewählten unterscheiden.

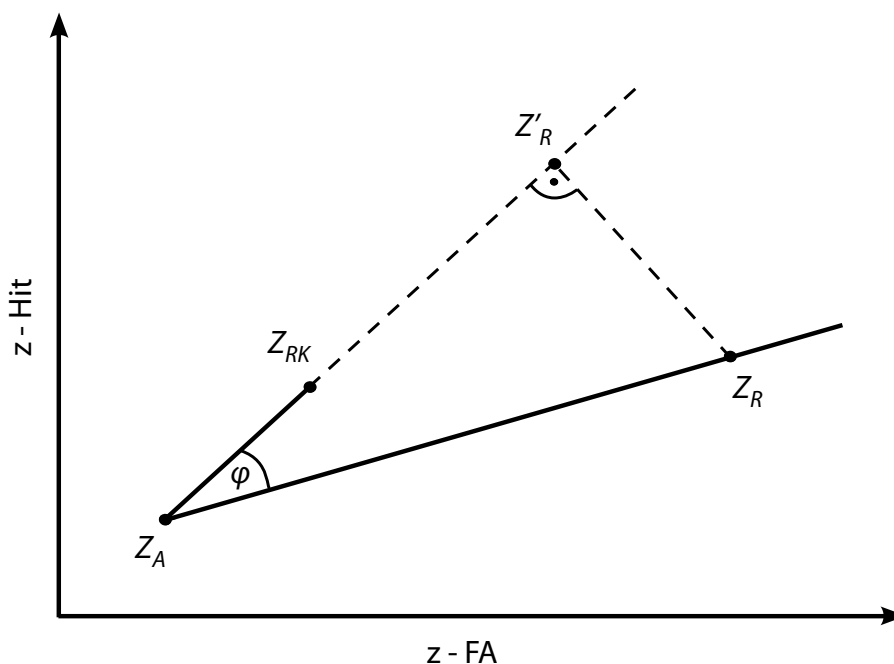


Abbildung 8.1: Dargestellt werden zwei z -ROC-Geraden – $\overline{Z_A Z_R}$ sowie $\overline{Z_A Z_{RK}}$, die sich im Punkt Z_A schneiden. Anhand des zusätzlich eingezeichneten Lots von Z_R auf $\overline{Z_A Z_{RK}}$ lässt sich der Schnittwinkel (φ) der beiden Geraden bestimmen.

Kurve in Abweichungen von Null äußert. Im Falle einer Abweichung der Werte des Winkels von 0 bedeuten negative Werte, dass die abruferlebensbasierte Gerade $\overline{Z_A Z_R}$ rechts des Schnittpunktes (Z_A) unterhalb der konfidenzbasierten Geraden ($\overline{Z_A Z_{RK}}$) verläuft. Negative Werte zeigen also einen umgekehrt-u-förmigen Verlauf der gesamten z -ROC-Kurve an. Aus positiven φ -Werten folgt hingegen, dass $\overline{Z_A Z_R}$ rechts des Schnittpunktes oberhalb von $\overline{Z_A Z_{RK}}$ verläuft. Positive φ -Werte zeigen also einen u-förmigen Verlauf der z -ROC-Kurve an.

Das Verhalten des Winkels φ gibt somit die Krümmung der Drei-Punkt- z -ROC-Kurve direkt wieder. Ein alternativer Ansatz, die Krümmung der ROC-Kurven zu bestimmen, liefe darauf hinaus, eine quadratische Funktion an die z -ROC-Punkte anzupassen. Dieses Vorgehen ist zwar gängiger, hätte darüber hinaus aber gegenüber der vorgeschlagenen Herangehensweise keinen bedeutenden Vorteil und ist außerdem rechnerisch aufwendiger. Da der eben vorgestellte Ansatz zur Analyse von Drei-Punkt- z -ROC-Kurven die Möglichkeit bietet, die Frage nach Steigungsunterschieden zwischen konfidenzbasierten und abruferlebensbasierten z -ROC-Kurven einerseits und die damit verwandte Frage nach der Krümmung der Drei-Punkt- z -ROC-Kurve andererseits in demselben Rahmen zu klären, wurde ersterer letzterem vorgezogen.

Deshalb werden im folgenden Abschnitt für jedes Modell Vorhersagen über die bei Gültigkeit des jeweiligen Modells zu erwartenden Steigungswerte für konfidenz- sowie abruferlebensbasierte ROC-Geraden und φ -Werte generiert, um diese anschließend mit den empirisch ermittelten Daten vergleichen zu können.

8.2 Ableitung von Vorhersagen für Schnittwinkel und Steigungen

Während die Vorhersagen des SDM-Modells für ROC-Kurven von Remember-Know-Guess-Untersuchungen identisch sind mit denjenigen, die Ratcliff et al. (1992) für Rekognitionsuntersuchungen formuliert haben, ist eine formale Lösung für die fünf anderen Modelle nicht leicht zu erreichen.

Wie bereits in Kapitel 3 diskutiert, zeigten Ratcliff et al. (1992), dass eindimensionale Signalentdeckungsmodelle im Allgemeinen und damit auch das SDM-Modell als Spezialfall selbiger lineare z -ROC-Kurven postulieren. Dabei entspricht die Steigung solcher Geraden – die Gültigkeit des Modells vorausgesetzt – dem Verhältnis der Varianzen der Vertrauheitsverteilungen für alte Stimuli und Distraktoren. Vor dem Hintergrund dieses Modells muss also erwartet werden, dass der Winkel φ einen Wert von Null annehmen sollte. Entsprechend müssten die Steigungen der beiden Geraden $\overline{Z_A Z_R}$ und $\overline{Z_A Z_{RK}}$ gleich groß sein.

Für die übrigen Modelle wurden Vorhersagen über Form und Verlauf der ROC-Kurven anhand einer parametrischen Bootstrap-Simulation gewonnen. Dabei wurden auf der Basis der Annahmen des jeweils im Mittelpunkt der Betrachtung stehenden Modells Vorhersagen bezüglich der erwarteten Treffer- und Falsche-Alarm-Wahrscheinlichkeiten für die drei Konfidenzbedingungen R , RK und A gewonnen; die Antworthäufigkeiten bestanden dabei aus korrekten und falschen Remember-Antworten (R), korrekten sowie falschen Remember- bzw. Know-Antworten (RK) und korrekten sowie falschen Alt-Antworten (A). Anhand der drei durch Remember-Antworten, Remember- und Know-Antworten sowie Alt-Antworten resultierenden z -ROC-Punkte wurden dann die Steigungen der beiden Geraden $\overline{Z_A Z_R}$ und $\overline{Z_A Z_{RK}}$ ermittelt, sowie der Winkel φ bestimmt, der zwischen den Geraden eingeschlossen wird. Ziel dieser Simulationsstudie war es, eine möglichst genaue Vorhersage für die im nächsten Abschnitt dargestellte Analyse der empirischen ROC-Kurven abzuleiten. Es war deshalb wichtig, dass die erzeugten Datensätze mit den empirischen Daten verwandt waren.

Damit die im Folgenden zu gewinnenden Modellvorhersagen auf die empirischen ROC-Kurven anwendbar waren, wurden einerseits die spezifischen Parameterwerte, anhand derer die Modellvorhersagen generiert wurden, aus den in Anhang C wiedergegebenen Schätzwerten gezogen, die anhand der Daten aus den einzelnen experimentellen Bedingungen bestimmt worden waren. Zusätzlich wurden in jedem einzelnen Simulationsdurchgang 148 ROC-Kurven erzeugt, die auf je drei Paaren von Wahrscheinlichkeiten beruhten, so dass jeder einzelne Simulationsschritt genauso viele Daten umfasste, wie die später zu analysierenden empirischen Daten.

Innerhalb der erzeugten Paare von Wahrscheinlichkeiten wurden Treffer-Wahrscheinlichkeiten zu den zugehörigen Falsche-Alarm-Raten in Bezug gesetzt. Dies geschah erstens auf der Ebene aller Alt-Antworten, zweitens auf der Ebene einer Sammelkategorie, in die Remember- und Know-Antworten einfließen, sowie drittens auf der Ebene einer Kategorie, die ausschließlich Remember-Antworten umfasste.

Die Vorhersage der Wahrscheinlichkeiten wurde für jedes Modell auf Basis der

in Kapitel 4 dargestellten Modellgleichungen vorgenommen. Um spezifische Vorhersagen zu erreichen, wurden die in den Modellgleichungen enthaltenen Parameter mit konkreten Werten versehen. Wie bereits erwähnt, wurden die einzusetzenden Parameterwerte aus den Schätzungen gewonnen, die sich, wie im vorangegangenen Kapitel beschrieben, für die 148 experimentellen Bedingungen ergeben hatten. Die einzelnen Werte wurden zufällig und weitgehend unabhängig voneinander gezogen und zu neuen Kombinationen zusammengesetzt, um die rekombinierten Parametervektoren möglichst unabhängig von den empirischen Parametersätzen zu gestalten. Eine Ausnahme von dieser Regel stellten Parametergruppen dar, die aufgrund der Modellannahmen durch Größer-Kleiner-Relationen zu Gruppen verbunden waren. Gruppen von auf diese Weise verbundenen Parametern existieren für alle Modelle, die ganz oder teilweise auf signalentdeckungstheoretischen Annahmen aufbauen. Bei den betroffenen Parametern handelt es sich ausnahmslos um Kriteriumsparameter. Für einige dieser Parameter muss, wie in Kapitel 4 verdeutlicht, gelten, dass das auf einer Dimension strengere Kriterium jeweils einen größeren Wert erhält als die auf der gleichen Dimension verankerten, aber laxeren Kriterien. Diese durch Größer-Kleiner-Relationen verbundenen Parametergruppen wurden gemeinsam aus einem der empirischen Parametersätze gezogen. So wurde sichergestellt, dass die relationalen Restriktionen stets gewahrt blieben.

Abhängig von der Form des jeweiligen Modells galten solche Abhängigkeitsregeln für einige oder alle der bereitgehaltenen Kriteriumsparameter, wobei jeweils nur solche Kriteriumsparameter betroffen waren, die auf derselben Dimension liegen. Für und die beiden hybriden Signalentdeckungsmodelle HTSD sowie HTUVSD mussten alle angenommenen Kriterien gemeinsam gezogen werden: Da die Kriterien auf einer gemeinsamen Vertrauheitsdimension variieren, muss das Kriterium C_R größer sein als das Kriterium C_K und dieses wiederum größer als das Kriterium C_O . Bei den Modellen Streak und SAC ist das Remember-Kriterium C_R auf einer von den beiden anderen Kriterien abgehobenen Ebene lokalisiert. Deshalb bildeten bei diesen zweidimensionalen Signalentdeckungsmodellen jeweils zwei Kriteriumsparameter eine relational verknüpfte Gruppe. Im Falle des Streak-Modells waren dies die Kriterien C_O und C_G , während aus den Annahmen des SAC-Modells folgte, dass die Kriterien C_O und C_K auf der Dimension der Detailaktivierung eine Gruppe relational verknüpfter Parameter bilden. Entsprechend wurden die jeweiligen Parameterpaare C_O und C_G für das Streak-Modell bzw. C_O und C_K für das SAC-Modell beim Sampling der Parameterwerte aus Parameterschätzungen gezogen, die an derselben Analyseeinheit gewonnen worden waren. Die verbundenen Parametergruppen und die Parameter, für die die Werte unabhängig von den übrigen ermittelt wurden, können Tabelle 8.1 entnommen werden.

Durch Einsetzen der unter den oben geschilderten Randbedingungen zufällig ausgewählten Parameterwerte in die in Kapitel 4 dargelegten Modellgleichungen ließen sich zunächst für alle Modelle Vorhersagen für Treffer- und Falsche-Alarm-Raten in den einzelnen Antwortkategorien generieren. Diese Antworthäufigkeiten wurden in zwei weiteren Schritten in z -ROC-Kurven verwandelt: Um die den ROC-Kurven zugrunde liegenden Wahrscheinlichkeiten zu erhalten, wurden im ersten Schritt alle auf

Tabelle 8.1: Überblick über das in den modellbasierten Simulationen angewandte Samplingschema. In Klammern zusammengefasste Parameter wurden gemeinsam gezogen.

| Modelle | | | | |
|----------|--------------|--------------|-------------------|-------------------|
| 4-States | SAC | Streak | HTUVSD | HTSD |
| r | d_x | d_x | r | r |
| k | d_y | d_y | d' | d' |
| d | σ | σ | σ | (C_R, C_K, C_O) |
| g_R | C_R | C_R | (C_R, C_K, C_O) | |
| g_K | (C_K, C_O) | (C_G, C_O) | | |
| g_G | | | | |

dem jeweiligen Konfidenzniveau zu berücksichtigenden Wahrscheinlichkeiten summiert. Anschließend wurden diese Wahrscheinlichkeiten z -transformiert. Das Ergebnis dieser Transformation bestand aus denjenigen z -Werten, die am oberen Ende der Standardnormalverteilung eine Fläche abschnitten, die den zu transformierenden Wahrscheinlichkeiten entsprachen.

Zur Ermittlung der Modellvorhersagen wurden für jedes Modell insgesamt 400 Replikationen des Originaldatensatzes generiert. Die Replikationsdatensätze bestanden wie die Originaldatensätze aus 148 Drei-Punkt- z -ROC-Kurven. Bei der Erzeugung der einzelnen ROC-Tripel wurde Sorge dafür getragen, dass sich die Treffer- und die Falsche-Alarm-Raten zwischen den Konfidenzstufen R sowie RK einerseits und A andererseits um mindestens .001 unterschieden. Durch diese Maßnahme sollte dem Problem entgegengewirkt werden, dass die Steigungen der ROC-Geraden gegen Null oder unendlich konvergierten.

Für jede einzelne ROC-Gerade wurden zudem Werte für den Winkel φ und die Steigungen der konfidenzbasierten Geraden (s_{RK}), sowie der abruferlebensbasierten Geraden (s_R) bestimmt. Auf der Ebene der Replikationsdatensätze ($n = 148$) wurden zusätzlich Aggregate berechnet, die die Basis für die in Tabelle 8.2 wiedergegebenen Ergebnisse bilden. Hier finden sich die mittleren vorhergesagten Steigungen s_{RK} und s_R sowie der mittlere Schnittwinkel φ . Zudem wird die mittlere Wahrscheinlichkeit $p_{\varphi > 0}$ berichtet, mit der φ in den Replikationsdatensätzen größer als 0 war.

Betrachtet man die mittleren Steigungen s_R und s_{RK} , so fällt auf, dass bei dreien der fünf Modelle die abruferlebensbasierten Geraden im Mittel steiler ausfielen als die konfidenzbasierten Geraden: Die anhand des SAC-Modell generierten Vorhersagen für s_R und s_{RK} betragen 1.44 respektive 0.92, für das Streak-Modell beliefen sich die Steigungen auf 1.16 sowie 0.75 und für das 4-States-Modell auf 1.06 respektive 0.75. Anders verhielt es sich bei den beiden Varianten des Yonelinas-Modells. Bei beiden Modellvarianten war s_R kleiner als s_{RK} : Für das HTSD-Modell belief sich die Steigung der abruferlebensbasierten z -ROC-Geraden auf 0.84, während die der konfidenzbasierten Geraden bei 0.88 lag. Für das HTUVSD-Modell betragen die Steigungen 0.98 respektive 0.99.

Tabelle 8.2: Überblick über die Ergebnisse von insgesamt 400 Replikationen des Originaldatensatzes. Dargestellt werden die vorhergesagten mittleren Steigungen der Geraden $\overline{Z_A Z_R}$ (s_R) und $\overline{Z_A Z_{RK}}$ (s_{RK}), ihr mittlerer Schnittwinkel (φ) sowie die relative Häufigkeit $p_{\varphi>0}$, mit der in den einzelnen Simulationsdurchgängen positive φ -Werte beobachtet wurden.

| | SAC | | | | Streak | | | |
|-------|-------|----------|-----------|-------------------|-----------|------------|-----------|-----------------|
| | s_R | s_{RK} | φ | $p_{\varphi>0}^a$ | $s_{A,R}$ | $s_{A,RK}$ | φ | $p_{\varphi>0}$ |
| MIN | 0.09 | 0.11 | -9.99 | .58 | 0.10 | 0.18 | -15.35 | .43 |
| M | 1.44 | 0.92 | 2.83 | .69 | 1.16 | 0.75 | 2.23 | .55 |
| SD(M) | 0.27 | 0.09 | 0.44 | .04 | 0.14 | 0.03 | 0.71 | .04 |
| MAX | 24.83 | 8.82 | 18.23 | .82 | 13.19 | 2.23 | 27.83 | .67 |

| | HTSD | | | | HTUVSD | | | |
|-------|-----------|------------|-----------|-----------------|-----------|------------|-----------|-----------------|
| | $s_{A,R}$ | $s_{A,RK}$ | φ | $p_{\varphi>0}$ | $s_{A,R}$ | $s_{A,RK}$ | φ | $p_{\varphi>0}$ |
| MIN | 0.11 | 0.19 | -6.04 | .00 | 0.15 | 0.22 | -6.17 | .00 |
| M | 0.84 | 0.88 | -0.88 | .00 | 0.98 | 0.99 | -0.31 | .00 |
| SD(M) | 0.02 | 0.01 | 0.09 | .00 | 0.04 | 0.03 | 0.07 | .00 |
| MAX | 1.00 | 1.00 | 0.00 | .00 | 2.51 | 2.52 | 0.00 | .00 |

| | 4-States | | | |
|-------|-----------|------------|-----------|-----------------|
| | $s_{A,R}$ | $s_{A,RK}$ | φ | $p_{\varphi>0}$ |
| MIN | 0.25 | 0.20 | -8.80 | .62 |
| M | 1.06 | 0.70 | 4.60 | .74 |
| SD(M) | 0.05 | 0.02 | 0.55 | .04 |
| MAX | 4.62 | 1.93 | 26.47 | .83 |

^aDie Basis für die Berechnung von Durchschnittswerten, Streuungen sowie Minimal- und Maximalwerten bildeten 400 relative Häufigkeiten, mit denen das angegebene Ereignis in den einzelnen Simulationsdurchgängen auftrat.

Obwohl die Differenzen zwischen den Gesamtmittelwerten insbesondere bei dem HTUVSD-Modell recht klein ausfielen, erwiesen sie sich allesamt als statistisch bedeutsam, wie der paarweise Vergleich der jeweiligen Steigungsmittelwerte zeigte, SAC: $t_{399} = 56.16$, Streak: $t_{399} = 59.03$, 4-States: $t_{399} = 138.94$, HTSD: $t_{399} = -217.37$, HTUVSD: $t_{399} = -89.23$.

Wie zu erwarten war, ließ sich das bei den mittleren Steigungswerten auftretende Muster von Ähnlichkeiten und Unterschieden zwischen den Modellen bei der Analyse der Schnittwinkel wiederfinden. Der mittlere Schnittwinkel φ fiel bei den drei Modellen, die im Mittel eine größere Steigung der abruferlebensbasierten ROC-Geraden vorhergesagt hatten, positiv aus, während er für die Yonelinas-Modelle negativ ausfiel. Am augenfälligsten von Null verschieden zeigte sich dabei der vor dem Hintergrund des 4-States-Modells ermittelte φ -Wert mit 4.60. Diesem folgten diejenigen des SAC-

Modells mit 2.83 und des Streak-Modells mit 2.23. Die auf der Basis der beiden hybriden Modelle HTSD und HTUVSD vorhergesagten φ -Werte beliefen sich auf -0.88 respektive -0.31 und lagen damit moderat unter Null.

Um zu überprüfen, inwiefern aus den Modellen auf das Vorzeichen der φ -Werte geschlossen werden kann, wenn – wie für die empirische Analyse geplant – 148 ROC-Tripel analysiert werden, wurden für jeden der 400 Replikationsdatensätze einzeln die statistischen Tests durchgeführt, die für die empirischen Datensätze verwendet werden sollten.

So wurde zunächst für jedes Modell in 400 t -Tests auf einem Signifikanzniveau von $\alpha = .05$ überprüft, ob die φ -Werte der Replikationsdatensätze von Null verschieden waren. Bei dieser Analyse fielen bei fünf Modellen 100% der Tests signifikant aus. Einzige Ausnahme blieb hier das Streak-Modell, für das 356 der 400 Tests signifikant ausfielen und durchgängig in positiven t -Werten resultierten, während die restlichen 44 Tests nicht signifikant waren. Für das SAC-Modell und das 4-States-Modell ergaben sich durchweg positive t -Werte, während die t -Werte für die beiden Yonelinas-Modelle durchweg kleiner als 0 waren.

Dieses Befundmuster ließ sich in 400 anschließend durchgeführten Vorzeichen-tests bestätigen, in denen überprüft wurde, ob die φ -Werte der einzelnen Simulationsdurchgänge überzufällig häufig von Null verschieden waren. Abgesehen von einer höheren Zahl insignifikanter Testergebnisse bekräftigten die Vorzeichentests die auf der Basis der t -Tests gewonnenen Eindrücke. Die Zahl signifikanter Ergebnisse verringerte sich für das SAC-Modell von 400 auf 398 und für das Streak-Modell von 356 auf 64 signifikante Ergebnisse. Für das 4-States-Modell wiesen alle 400 Tests darauf hin, dass die φ -Werte in mehr als der Hälfte der Fälle größer als Null waren. Auch alle Vorzeichentests, die für vor dem Hintergrund der Yonelinas-Modelle erzeugter Daten durchgeführt wurden, fielen signifikant aus. Dabei waren die φ -Werte immer kleiner als Null.

Gleiches wird deutlich, wenn man die in Tabelle 8.2 dargestellten relativen Häufigkeiten, mit denen φ in den 400 Simulationsdurchgängen im Mittel positiv ausfiel, betrachtet. Hier zeigt sich, dass dies für die beiden hybriden Modelle nie der Fall war. Beim Streak-Modell lag dieser Wert mit einer mittleren relativen Häufigkeit von .55 nur knapp über dem Zufallsniveau von .50, während die relativen Häufigkeiten für das SAC-Modell und das 4-States-Modell mit .69 respektive .74 vergleichsweise hoch ausfielen.

Die in den bis hierhin berichteten Analysen erzielten Resultate lassen sich als Vorhersagen der Modelle für empirische ROC-Geraden auffassen. Diese Vorhersagen können im Folgenden anhand der empirischen ROC-Kurven überprüft werden, die im Anschluss behandelt werden. Daher sollen die wesentlichen Erkenntnisse aus der durchgeführten Simulation an dieser Stelle noch einmal zusammengefasst werden.

Für die folgende Analyse besonders beachtenswert scheint, dass sich die Modelle darin unterscheiden, ob sie positiv oder negativ beschleunigte z -ROC-Kurven erwarten lassen. Drei der Modelle – das sind das Streak-, das SAC-, und das 4-States-Modell, – sagen für die empirischen Daten ein positives mittleres φ vorher, während das HTSD- und das HTUVSD-Modell ein negatives φ erwarten lassen. In gleicher

Weise unterscheiden sich die Modelle hinsichtlich der vorhergesagten Steigungen: Vor dem Hintergrund beider Versionen des Yonelinas-Modells sollte die Steigung der abruferslebenbasierten Geraden geringer ausfallen als die der konfidenzbasierten Geraden, während den drei übrigen Mehr-Prozess-Modellen zufolge die abruferslebenbasierte Gerade größer sein sollte als die konfidenzbasierte.

8.3 Analyse der empirischen ROC-Kurven

An dieser Stelle sollen die empirischen Antworthäufigkeiten in ROC-Kurven überführt und analysiert werden. Als Daten wurden dazu jene Daten verwendet, die in den 148 einzelnen experimentellen Bedingungen berichtet wurden. Die Auswertung der empirischen ROC-Kurven verlief dabei analog zu der Auswertung, die im Rahmen der im vorhergehenden Abschnitt geschilderten Simulationsstudie durchgeführt wurde. Im Folgenden werden also die Steigungen der abruferslebenbasierten und der konfidenzbasierten Geraden sowie der zwischen diesen Geraden eingeschlossene Schnittwinkel φ im Mittelpunkt des Interesses stehen.

Bei der Analyse der empirischen ROC-Kurven stellen sich zwei zentrale Fragen, die in den folgenden Betrachtungen einer Beantwortung zugeführt werden sollen: Zum einen muss überprüft werden, inwiefern die Drei-Punkt- z -ROC-Kurven mehrheitlich positiv respektive negativ beschleunigt sind, und zum anderen ist zu untersuchen, welche Werte die Steigungen s_{RK} und s_R in den empirischen Daten annehmen.

Diese Fragen stellen sich insbesondere deshalb, weil die Antworten dazu beitragen können zwischen – im Sinne von deskriptiver Adäquatheit – besseren und schlechteren Modellen zu diskriminieren. Diese Möglichkeit ist dem Umstand zu verdanken, dass sich aus den Modellen mehr oder minder klare Vorhersagen über den Ausgang dieser Analysen ableiten ließen: Aus den Annahmen des eindimensionalen Signalentdeckungsmodells lässt sich, wie geschildert, die Hypothese ableiten, dass φ gleich Null oder zumindest nicht überzufällig häufig von Null verschieden sein sollte. Aufgrund der im vorangegangenen Abschnitt beschriebenen Simulationen ließen sich für die übrigen Modelle ähnliche Hypothesen konstruieren. Bei Gültigkeit eines der Yonelinas-Modelle wäre beispielsweise zu erwarten, dass φ mehrheitlich negative Werte annimmt, während die übrigen Modelle mehrheitlich positive φ -Werte vorhersagten. Ebenso wäre vor dem Hintergrund der Yonelinas-Modelle ein negatives mittleres φ zu erwarten, während die übrigen Modelle ein mittleres φ vorhersagten, das größer als Null sein sollte. Da die Untersuchung von φ und die Ermittlung globaler Steigungsunterschiede insofern eng miteinander verzahnt sind, als s_R immer dann größer als s_{RK} ausfällt, wenn φ größer als 0 ist, und im gegenteiligen Falle kleiner als s_{RK} ist, genügt es festzustellen, welche Werte die Steigungen annehmen. Die für die einzelnen experimentellen ermittelten ROC-Indizes können Anhang B entnommen werden.

Ein Überblick über die Ergebnisse liefert Tabelle 8.3. Man sieht hier, dass die mittlere Steigung der abruferslebenbasierten Geraden s_R mit einem Wert von 0.98 deutlich über der mittleren Steigung der konfidenzbasierten Geraden s_{RK} lag, die 0.70 betrug.

Tabelle 8.3: Überblick über die Ergebnisse der Analyse von 148 empirisch gewonnenen ROC-Datensätzen. Dargestellt werden die mittleren Steigungen der Geraden $\overline{Z_A Z_R}$ (s_R) und $\overline{Z_A Z_{RK}}$ (s_{RK}), ihr mittlerer Schnittwinkel (φ) sowie die relative Häufigkeit $p_{\varphi>0}$, mit der positive φ -Werte beobachtet wurden.

| | Steigung | | Winkel | |
|-----|----------|----------|-----------|-----------------|
| | s_R | s_{RK} | φ | $p_{\varphi<0}$ |
| Min | 0.29 | 0.02 | -6.49 | – |
| M | 0.98 | 0.70 | 4.56 | .83 |
| SD | 0.43 | 0.35 | 5.68 | – |
| Max | 2.68 | 2.43 | 21.57 | – |

Dieser Größenunterschied setzte sich im mittleren Schnittwinkel der beiden ROC-Geraden fort; mit einem mittleren Wert von 4.56 fiel φ positiv aus und erwies sich als statistisch von Null verschieden, $t_{147} = 9.77$. Zudem war φ für 123 der 148 Bedingungen größer als 0. Damit fielen die φ -Werte in 83% der Bedingungen positiv. Dies war mehr als aufgrund des bloßen Zufalls erwartet werden konnte, $p_{binom}(123, 148) = .01$.

8.4 Diskussion

Die in diesem Kapitel vorgestellten Resultate können aus zwei Perspektiven heraus betrachtet werden. Zunächst lassen sich die hier festgestellten Eigenschaften der empirischen ROC-Kurven in einen Zusammenhang bringen mit den in Kapitel 3 vorgestellten Untersuchungen, in denen ROC-Geraden aus Remember-Know-Untersuchungen verglichen wurden mit ROC-Kurven aus Rekognitionsuntersuchungen, in denen Konfidenzurteile erfragt wurden. Darüber hinaus liefern die empirischen ROC-Kurven wichtige Hinweise darüber, welche der in Kapitel 4 vorgestellten Modelle von den empirischen Daten gestützt werden. Um diesen zweiten Punkt näher zu beleuchten, braucht es eine Zusammenschau der in den beiden vorigen Abschnitten gelieferten Daten bezüglich der abgeleiteten bzw. simulierten Modellvorhersagen einerseits und die empirischen ROC-Kurven betreffend andererseits.

Dadurch, dass sich die aus Remember-Know-Guess-Untersuchungen konstruierbaren Drei-Punkt-ROC-Kurven in zwei Geraden unterteilen lassen, von denen eine den ROC-Kurven ähnlich ist, die in Rekognitionsuntersuchungen erhoben werden, während die andere enger mit herkömmlichen Remember-Know-ROC-Geraden in Beziehung steht, lassen sich Fragen aus der Literatur aufgreifen, die bis dahin schwer zu beantworten waren. Dabei handelt es sich insbesondere um die von Rotello et al. (2004) aufgeworfene und von Wixted und Stretch (2004) aufgegriffene Frage nach Unterschieden in den Steigungen zwischen konfidenzbasierten ROC-Kurven einerseits und Remember-Know-ROC-Geraden andererseits.

Die im vorhergehenden Abschnitt präsentierten Resultate weisen auf deutliche Un-

terschiede zwischen den Steigungen dieser beiden Arten von ROC-Kurven hin. Dabei sind diese Unterschiede anders als in der Arbeit von Wixted und Stretch (2004) insofern durchgängig, als sie nicht nur in den mittleren Steigungen wiedergefunden werden können, sondern zusätzlich auch in den Häufigkeiten aufscheinen, mit denen abruferlebensbasierte ROC-Geraden steiler waren als ihre konfidenzbasierten Gegenstücke.

Der Unterschied zwischen den Steigungen der beiden Geraden trat in den Ergebnissen deutlich zu Tage: Während abruferlebensbasierte ROC-Geraden im Mittel eine Steigung von 0.98 aufwiesen, lag die mittlere Steigung der konfidenzbasierten ROC-Geraden nur bei 0.70. Dieser Unterschied ist insofern interessant, als sich in diesen Ergebnissen die Werte wiederfinden lassen, die Rotello et al. (2004) für ROC-Geraden auf der Basis von Remember-Know-Daten berichtet haben. Wie in Kapitel 3 erläutert, stellten die genannten Autoren fest, dass ROC-Kurven, die auf Abruferlebensberichten fußten, eine Steigung von ungefähr 1 aufwiesen. Dies steht im Widerspruch zu den Resultaten von ROC-Analysen für Rekognitionsexperimente, in denen ROC-Kurven anhand von Konfidenzurteilen gewonnen wurden. Glanzer et al. (1999) ermittelten für solche Untersuchungen eine deutlich geringere Steigung. Während, wie Wixted und Stretch (2004) ausführten, gegen den direkten Vergleich von Steigungswerten aus unterschiedlichen Arten von Untersuchungen spricht, dass Unterschiede in der mittleren Steigung auch auf Unterschiede im Untersuchungsdesign zurückgeführt werden können, gilt dieses Argument für die vorliegenden Daten nicht. Da allen Schätzungen die gleichen Daten zugrunde lagen, muss die Erklärung für die gefundenen Steigungsunterschiede in der Natur des Abruferlebens selbst bzw. in den dem Abruferleben zugrunde liegenden Prozessen gesucht werden.

Für sich genommen stellt dieser Befund bereits eine interessante Beobachtung dar. Für den Vergleich der in Kapitel 4 vorgestellten Modelle bot es sich allerdings an, noch einen Schritt weiter zu gehen und die empirischen ROC-Daten mit den auf der Basis der Modelle abgeleiteten bzw. simulierten ROC-Kurven zu vergleichen. Bei einem solchen Vergleich fällt auf, dass die empirischen ROC-Kurven weder den Vorhersagen entsprachen, die vor dem Hintergrund des eindimensionalen Signalentdeckungsmodell zu erwarten waren, noch mit den Vorhersagen der Yonelinas-Modelle kompatibel waren. Die Ergebnisse sprechen dabei sowohl bezüglich der Steigungsdifferenzen als auch bezüglich des Schnittwinkels eine klare Sprache: Die alleinige Analyse von Remember-Antworten führte verglichen mit der Analyse von Remember- und Know-Antworten zu einem Effizienzverlust, der sich darin ausdrückt, dass mehr Falsche Alarmer in die Remember-Kategorie fielen als es zu erwarten gewesen wäre, wenn die Annahmen des eindimensionalen Signalentdeckungsmodells gültig wären.

Dieses Datenmuster ist prinzipiell mit allen drei übrigen Modellen kompatibel. Ein Ausbleiben dieses Resultates hätte für das SAC- sowie das 4-States-Modell allerdings schwerer gewogen als für das Streak-Modell, da für die ersteren alle bzw. nahezu alle 400 Simulationen zu signifikanten Unterschieden in die gegebene Richtung geführt hatten, das Streak-Modell hingegen zusätzlich mit Daten vereinbar gewesen wäre, die den Vorhersagen des eindimensionalen Signalentdeckungsmodells entsprochen hätten.

Betrachtet man die ermittelten Steigungswerte sowie den mittleren Schnittwinkel und vergleicht diese mit den simulierten Werten, so lagen die empirischen Daten am

nächsten an den Vorhersagen des 4-States-Modells. Dessen Vorhersagen wichen mit 1.06 für die Steigung der abrufersbasierten ROC-Geraden und 0.70 für die konfidenzbasierte Gerade nur wenig bzw. gar nicht von den empirisch ermittelten Daten ab. Genauso lag der empirische Winkelwert mit 4.56 nur knapp unter dem vor dem Hintergrund des multinomialen Bearbeitungsbaummodells zu erwartenden mittleren φ von 4.60. In allen Maßen wichen die Vorhersagen der beiden anderen infrage kommenden Modelle weiter von den empirisch erhaltenen Werten ab.

Diese Ergebnisse weisen auf die im nächsten Kapitel zu behandelnde deskriptive Adäquatheit der zu vergleichenden Modelle hin. Darin wird jedoch insofern über die hier betrachtete Übereinstimmung zwischen theoretisch vorhergesagten und empirischen Datenmustern hinausgegangen, als Maximum-Likelihood-Verfahren eingesetzt wurden, um den Grad der Übereinstimmung von Modellvorhersagen und empirischen Befunden zu erfassen.

Kapitel 9

Deskriptive Adäquatheit

In den beiden vorausgegangenen Kapiteln sind die Daten der in den Modellvergleich aufgenommenen Studien in einzelne Bedingungen dekomponiert und auf dieser Basis analysiert worden. In den nun folgenden Teilen wurden die in der Literatur berichteten Experimente jeweils als Ganzes betrachtet und analysiert. In diesem Kapitel wird es dabei um die deskriptive Adäquatheit der Modelle gehen, also um die Frage, inwiefern die Modelle in der Lage waren, die empirisch erlangten Daten zu beschreiben. Verwendet wurden hierzu – wie in Kapitel 5 vorgeschlagen – ML-basierte Maße, die zur Bestimmung der deskriptiven Adäquatheit geeignet sind. Dabei handelte es sich vor allem um die Likelihood-Quotienten-Statistik G^2 sowie das Informationskriterium AIC_c .

Das folgende Kapitel gliedert sich in zwei Teile. In dem ersten der beiden Teile wird der Vergleich der Modelle hinsichtlich ihrer Datenanpassungsleistung behandelt, wobei auf die oben genannten Indizes zurückgegriffen wurde. Im zweiten Teil wurden diese Indizes erweitert um eine bootstrappingbasierte Abschätzung des Modelloptimismus. Der Modelloptimismus, der als Ausdruck des Vorhersagefehlers betrachtet werden kann, wurde, wie von Efron und Tibshirani (1993) vorgeschlagen, mit der Methode der Bootstrap-Modellselektion bestimmt. Anhand des Vorhersagefehlers sollte zusätzlich zu der Untersuchung der deskriptiven Adäquatheit festgestellt werden, ob die in der ersten Analyse beobachtete Datenanpassungsleistung unabhängig von der Komplexität der einzelnen Modelle ist. Dies war, wie bereits erläutert, unerlässlich, um feststellen zu können, inwiefern die Komplexität eines Modells der Generalisierbarkeit seiner Vorhersagen im Wege steht. Sollten die Ergebnisse der beiden Analysen keine Übereinstimmung zeigen, so steht zu befürchten, dass die Parameterschätzungen der betroffenen Modelle stärker durch Messfehler verzerrt waren als die Parameterschätzungen der anderen Modelle.

9.1 Likelihood-Quotienten-Tests und AIC-basierte Analysen

Der nun folgende Abschnitt befasst sich mit dem Vergleich der Modelle hinsichtlich der Datenanpassungsleistung, die sie bei der Anpassung an komplexere Datensätze erreichen. Die Dateneinheiten waren hierbei insofern komplexer, als die Ausgänge vollständiger Experimente modelliert wurden. Anhand dieser Daten wurden Parameterschätzungen vorgenommen und Datenanpassungsindizes bestimmt, die den anschließend durchgeführten Modellvergleichen zugrunde liegen. Die Parameterschätzmethode entsprach dabei dem in Kapitel 7 dargestellten Vorgehen: Wie auch bei den dort geschilderten Schätzungen wurden für jedes Modell insgesamt zehn Schätzungen vorgenommen; berichtet werden jeweils die Daten des Durchgangs, der den größten \mathcal{LL} -Wert aufwies. Dieser maximale \mathcal{LL} -Wert wurde anschließend anhand der in Gleichung 5.5 festgehaltenen Rechenvorschrift in einen G^2 -Wert transformiert. Zusätzlich wurde entsprechend der in Tabelle 5.1 niedergelegten Formel der auf die Verwendung kleiner Stichproben angepasste AIC -Wert¹ berechnet.

Anders als bei der Analyse der einzelnen Bedingungen wurde bei der Bestimmung der deskriptiven Adäquatheit auf der Basis ganzer Experimente darauf verzichtet, das Fehlen von Distraktorbedingungen durch mehrmaliges Verwenden einzelner Antworthäufigkeiten aufzufangen. Diese Maßnahme erübrigte sich bei der angestrebten Analyse, da sich die Modelle durch das Gleichsetzen von Parametern in Varianten überführen ließen, die sich dazu eigneten, die Daten der Experimente so wiederzugeben, wie sie erhoben wurden. Die Modellvarianten, mit denen aus Untersuchungen, die ein unvollständiges Design aufwiesen, – bei denen also zu einer einzigen Distraktorklasse zwei oder mehr Gruppen alter Reize gehörten – Parameterwerte geschätzt wurden, wurden dadurch erzeugt, dass alle Parameter gleich gesetzt wurden, die sowohl in den Modellgleichungen der miteinander verbundenen Gruppen alter Stimuli als auch in den Gleichungen für die jeweils verbindenden Distraktoren auftauchen.

Die Anpassung an unvollständige Designs bedeutete für SDM-, HTSD- und HTUVSD-Modell sowie SAC-Modell, dass zwischen den verbundenen Zielreizklassen alle Kriteriumsparameter C_R , C_K und C_O bzw. C_R , C_K und C_G paarweise gleich gesetzt wurden. Entsprechend wurden für das 4-States-Modell die Rateparameter g_R , g_K und g_G zwischen den verbundenen Reizklassen gleichgesetzt.

Für das Streak-Modell musste allerdings von diesem generellen Verfahren abgewichen werden. Da in den Modellgleichungen dieses Modells alle Parameter – von dem Streuungsparameter einmal abgesehen – sowohl in den für die Zielreize relevanten Gleichungen als auch in den Gleichungen für Distraktoren vorhanden sind, wäre eine Gleichheitsrestriktion nach oben geschildertem Muster gleichbedeutend mit der Behauptung, dass es keine Unterschiede zwischen den verschiedenen Zielreizbedingungen geben dürfe. Um die deskriptive Adäquatheit dieses Modells vergleichbar streng zu erfassen wie die der übrigen Modelle, wurden daher auch für das Streak-Modell

¹Im Folgenden werden die Begriffe AIC und AIC_c synonym verwendet. Bei den Werten, auf die damit Bezug genommen wird, handelt es sich durchgängig um AIC_c -Werte.

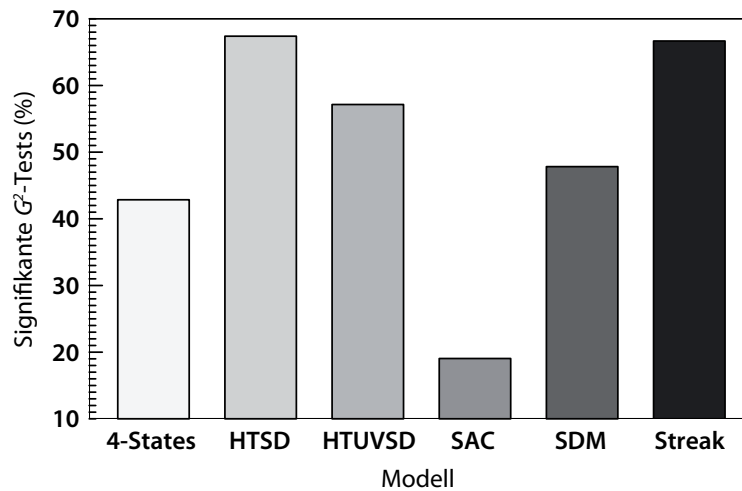


Abbildung 9.1: Dargestellt werden die relativen Häufigkeiten, mit denen G^2 -Anpassungstests für die zu vergleichenden Modelle bei $\alpha = .05$ in signifikante Ergebnisse mündeten. Relativiert wurde dabei an der Anzahl von Datensätzen, die mehr freie Datenkategorien bereithielten als für die Schätzung der Parameter des jeweiligen Modells erforderlich waren.

nur die Kriteriumsparameter C_R , C_G und C_O zwischen den verbundenen Kategorien gleichgesetzt. Die Sensitivitätsparameter d_x und d_y durften hingegen zwischen den Zielwortbedingungen variieren. Für die Modellierung der Anwohnhäufigkeiten in den Distraktorklassen wurden die benötigten Sensitivitätsparameter von der jeweils ersten der zugehörigen Zielreizklassen übernommen. Die Sensitivitätsschätzer der übrigen Zielreizklassen konnten frei variieren.

Im ersten Schritt zur Bestimmung der deskriptiven Adäquatheit wurden für jedes Modell G^2 -Datenanpassungstests auf einem Signifikanzniveau von $\alpha = .05$ durchgeführt. Dies war für die fünfparametrischen Modelle, also für das *HTSD*- und das *SDM*-Modell, für alle 46 Untersuchungen möglich, da die Anzahl der frei variierenden Datenkategorien durchgängig die Anzahl der aus den Daten zu schätzenden Modellparameter übertraf. Für die vier sechsparemetrischen Modelle entsprach die Anzahl der freien Datenkategorien in der Regel der Anzahl der aus den Datensätzen zu schätzenden Parameter. Nur für die Experimente, die zu den unvollständigen Designs zählten, standen auch für diese Modelle freie Datenkategorien zum Test der Modellgültigkeit zur Verfügung.

In Abbildung 9.1 finden sich die relativen Häufigkeiten, mit denen G^2 -Tests statistisch bedeutsame Abweichungen der Datenanpassungsgüte von der des Null-Modells anzeigten. Die Anzahl signifikanter Tests wurde dabei relativiert an der Anzahl möglicher Tests; wie oben dargelegt, waren das für die fünfparametrischen Modelle alle 46 Experimente und für die sechsparemetrischen Modelle jene 21 Untersuchungen mit unvollständigen Designs. Wie man in der Abbildung sieht, schnitt das SAC-Modell mit

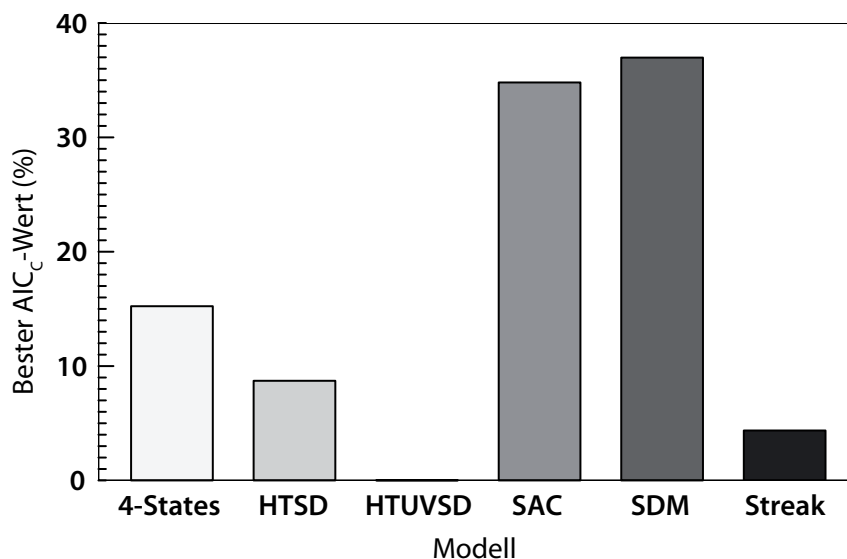


Abbildung 9.2: Dargestellt wird die Häufigkeit, mit der die verglichenen Modelle als im Sinne des AIC-Wertes bestpassendes Modell klassifiziert wurden.

19% Zurückweisungen am besten ab. Die restlichen Modelle wurden teilweise deutlich häufiger zurückgewiesen. So bewährte sich das 4-States-Modell am zweitbesten, wurde aber in 42.86% der Tests zurückgewiesen. Darauf folgten das SDM-Modell mit 47.83% Zurückweisungen und das HTUVSD-Modell mit 57.14%. Die höchsten Zurückweisungsrate wiesen das Streak-Modell mit 66.67% und das HTSD-Modell mit 67.40% auf.²

In einem zweiten Schritt wurden die Modelle hinsichtlich ihrer AIC-Werte miteinander verglichen. Die AIC-Werte lieferten aus drei Perspektiven informative Auskünfte über die deskriptive Adäquatheit der Modelle. Zum einen ließ sich mit Hilfe dieses Indexes feststellen, welches der betrachteten Modelle zu einem gegebenen Datensatz am besten passte. Dieses AIC-bestpassende Modell zeichnete sich durch den geringsten AIC-Wert aus. Zweitens ließ sich feststellen, inwiefern ein Modell in Termini von AIC nicht bedeutsam schlechter war als das AIC-bestpassende Modell. Drittens schließlich ließen sich die AIC-Werte der verschiedenen Modelle für jeden einzelnen Datensatz in eine Rangreihe bringen, so dass die Modelle dahingehend verglichen werden konnten, welche Rangposition sie relativ zu den übrigen Modellen im Mittel einnahmen.

Die Ergebnisse der ersten Analyse finden sich in Abbildung 9.2 wieder. Wie man dieser Abbildung entnehmen kann, waren zwei Modelle nahezu gleich häufig das AIC-bestpassende Modell. Bei diesen beiden Modellen handelte es sich zum einen um das eindimensionale Signalentdeckungsmodell SDM und zum anderen um das

²Eine Übersicht über die Datenanpassungsindizes, die für die hier berichteten Analysen ermittelt wurden, findet sich in Anhang F.

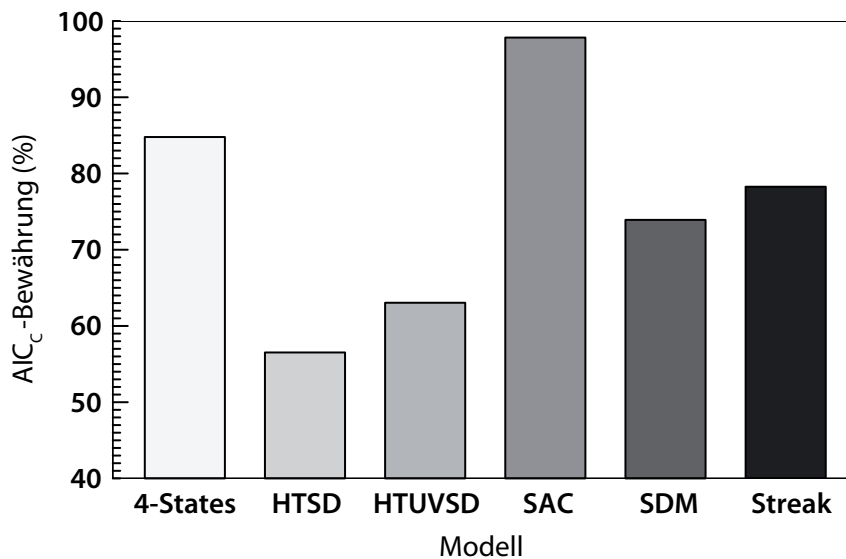


Abbildung 9.3: Dargestellt wird die Häufigkeit, mit der die einzelnen Modelle im Sinne von AIC -Differenzwerten (ΔAIC) als nicht bedeutsam von dem AIC -bestpassenden Modell verschieden klassifiziert wurden. Als Bedeutsamkeitskriterium wurde ein ΔAIC -Wert von 10 oder größer verwendet.

zweidimensionale Signalentdeckungsmodell SAC. Dabei war das SDM-Modell bei 17 der 46 Experimente – also bei 36.96% aller Experimente – das AIC -bestpassende Modell. Das SAC-Modell hingegen erreichte bei 34.78% aller Experimente den geringsten AIC -Wert und war damit für 16 der 46 Datensätze das bestpassende Modell. Bei nur noch 7 der 46 Experimente erwies sich das 4-States-Modell als AIC -bestpassendes Modell. Mit seiner Rate von insgesamt 15.22% übertraf es jedoch das darauf folgende HTSD-Modell um nahezu das Doppelte. Dieses nahm mit insgesamt 4 der 46 Experimente, bei denen es als AIC -bestpassendes Modell abschnitt, den vierten Platz ein. Bei den beiden verbleibenden Datensätzen erwies sich das Streak-Modell als das AIC -bestpassende Modell. Anders als das HTSD-Modell, das sich immerhin für 8.70% der Experimente als AIC -bestpassend gezeigt hat, war der AIC -Wert, den das HTUVSD-Modell erreichte, durchgängig größer als der AIC -Wert von mindestens einem der konkurrierenden Modelle.

Neben der Bestimmung des bestpassenden Modells ließ sich differenzieren zwischen solchen Modellen, deren AIC -Wert sich bedeutsam von dem des bestpassenden Modells unterschied, und den Modellen, deren AIC -Wert nicht bedeutsam größer war als der AIC -Wert des bestpassenden Modells. Um eine entsprechenden Analyse vorzunehmen, wurden in einem ersten Schritt für jedes Experiment die Differenzen zwischen dem für jedes Modell bestimmten AIC -Wert und dem AIC -Wert des bestpassenden Modells ermittelt. Auf diese Weise resultierten AIC -Differenzwerte, die als ΔAIC -Werte bezeichnet werden. Für das bestpassende Modell beträgt ΔAIC immer

Null, die Werte der übrigen Modelle bewegen sich oberhalb von Null. In einem zweiten Schritt ließen sich die ΔAIC -Werte der einzelnen Modelle in Analogie zu herkömmlichen Signifikanztests mit einem kritischen Wert vergleichen, der bei 10 angesiedelt wurde. Eine Differenz von 10 oder größer zeigt erfahrungsgemäß einen bedeutsamen Unterschied zwischen zwei AIC -Werten an (vgl. Burnham & Anderson, 2002). Betrug also der ΔAIC -Wert eines Modells 10 oder mehr, so wurde davon ausgegangen, dass dieses Modell bedeutsam schlechter zu den Daten passte als das AIC -bestpassende Modell. Eine AIC -Differenz, die kleiner als 10 war, wurde hingegen als Hinweis darauf gedeutet, dass das gegebene Modell nicht bedeutsam schlechter abschnitt als das AIC -bestpassende Modell.

In Abbildung 9.3 wird der Anteil der Experimente dargestellt, bei denen der ΔAIC -Wert eines gegebenen Modells weniger als 10 betrug. In dieser Wertung schnitt das SAC-Modell, das aus der vorherigen Analyse als zweitbestes Modell hervorgegangen war, am besten ab. Sein AIC -Wert war für 45 der 46 Experimente weniger als zehn Punkte kleiner als der AIC -Wert des AIC -bestpassenden Modells. Dieses Ergebnis entsprach einer Bewährungsrate von 97.83%. Als zweitbestes Modell dieser Analyse erschien das 4-States-Modell. Es war bei insgesamt 39 Experimenten das bestpassende Modell oder nicht bedeutsam schlechter als das bestpassende Modell. Die Bewährungsrate des 4-States-Modells lag mit 84.78% über derjenigen des Streak-Modells, die 78.26% betrug. Damit bewährte sich das Streak-Modell bei 36 Experimenten. An vierter Stelle folgte das SDM-Modell, welches bei insgesamt 34 der 46 Experimente das AIC -beste Modell war oder nicht bedeutsam schlechter als das bestpassende Modell abschnitt. Die Bewährungsquote des SDM-Modells fiel mit 73.91% um mehr als 10% größer aus als die des HTUVSD-Modells. Die Bewährungsrate dieses Modells betrug 63.04%. Von einer Bewährung des HTUVSD-Modells konnte bei 29 der 46 Datensätze ausgegangen werden. Zuletzt folgte die zweite Variante des Yonelinas-Modells: Das HTSD-Modell war bei 26 Experimenten nicht bedeutsam schlechter als das jeweils bestpassende Modell und bildete mit einer Bewährungsquote von 56.52% das Schlusslicht unter den konkurrierenden Modellen.

In dem folgenden Analyseschritt wurde das Ziel verfolgt, die Modelle hinsichtlich ihrer Datenanpassungsleistung direkt miteinander zu vergleichen. Einen statistisch abgesicherten Vergleich zwischen den Modellen erlauben multiple Vergleiche, die auf der Basis der von den Modellen geäußerten Datenanpassungsleistung durchgeführt werden. Da es unüblich ist, statistische Tests über Indizes der Datenanpassungsleistung durchzuführen, wurden diese – wie im Folgenden beschrieben – in ordinale Daten überführt.

Dazu wurden auf der Basis der von den Modellen erreichten AIC -Werte für jedes Experiment Rangplätze vergeben. Das AIC -beste Modell erhielt dabei den Rangplatz 1, das Modell mit dem zweitkleinsten AIC -Wert den Rangplatz 2 und so weiter. Kleinere Rangplätze wiesen also auf ein besseres Abschneiden hin.

Der Vergleich der Modelle wurde letztlich in zwei Schritten vorgenommen: Im ersten Schritt wurde auf globaler Ebene überprüft, ob sich die Modelle systematisch darin unterschieden, wie häufig sie einen geringeren AIC -Wert erreichten als die konkurrierenden Modelle. Anschließend wurde, auf diesen globalen Unterschieden zwi-

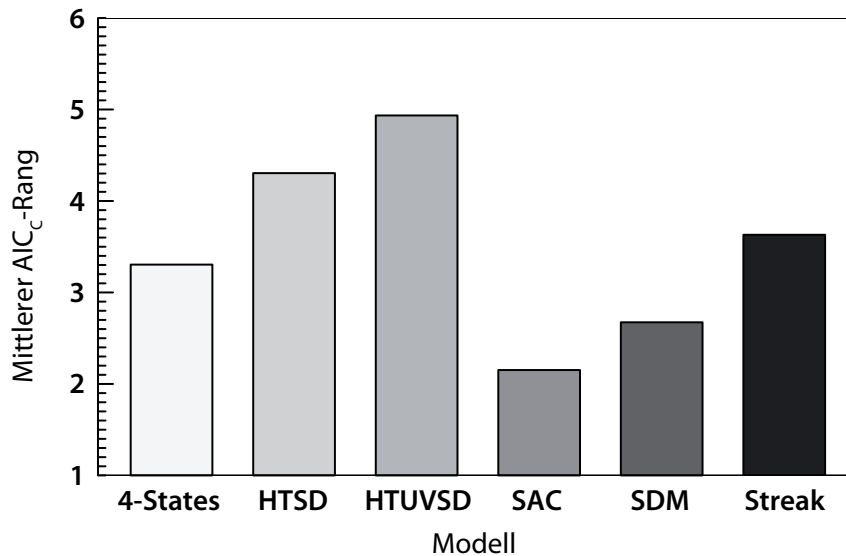


Abbildung 9.4: Dargestellt werden die mittleren AIC-Ränge, die die einzelnen Modelle über alle Datensätze hinweg eingenommen haben. Dabei repräsentieren kleinere Ränge bessere AIC-Werte.

schen den Modellen aufbauend, untersucht, zwischen welchen Modellen im Einzelnen statistisch bedeutsame Differenzen in den Häufigkeiten auftraten, mit denen ein bestimmtes Modell über die Experimente hinweg tendenziell einen geringeren AIC-Wert aufwies als ein zweites Modell, mit dem es verglichen wurde.

Die mittleren Rangplätze, die die einzelnen Modelle über alle Experimente hinweg eingenommen haben, können Abbildung 9.4 entnommen werden. Auch bei der Betrachtung der Rangplätze schnitt das SAC-Modell mit einem mittleren Rang von 2.15 besser ab als die übrigen Modelle. Den zweitbesten Rang nahm das SDM-Modell ein, das einen mittleren Rangplatz von 2.67 erreichte. Im Mittel auf einen Rang von 3.30 kam das 4-States-Modell. Darauf folgte das Streak-Modell mit einem mittleren Rang von 3.63. Die Endpositionen belegten die Yonelinas-Modelle: Von diesen beiden Modellen erreichte das HTSD-Modell einen mittleren Rangplatz von 4.30 und das HTUVSD-Modell ein Rangmittel von 4.93. Letzteres erhielt damit von den sechs konkurrierenden Modellen im Mittel den schlechtesten Rangplatz.

Ob sich die eingenommenen Rangplätze im statistischen Sinne bedeutsam voneinander unterschieden, ließ sich mit Hilfe eines Quade-Tests klären (vgl. Conover, 1999). Wesentlicher Bestandteil der Teststatistik des Quade-Tests sind die Differenzen zwischen den jeweils erreichten Rängen und dem Mittel der möglichen Ränge. Bei sechs möglichen Rängen entsprach der letztgenannte Wert genau $3\frac{1}{2}$. Die Differenzen zwischen dem tatsächlich erreichten Rang und 3.5 wurden zusätzlich mit der Spannweite auf der den Rängen zugrunde liegenden abhängigen Variable gewichtet. Im Falle der AIC-Werte bestand die für jedes einzelne Experiment angewandte Gewichtung aus

Tabelle 9.1: Dargestellt werden multiple Vergleiche hinsichtlich der Rangpositionen, die den Modellen anhand der in den einzelnen Experimenten erlangten *AIC*-Werte zugeteilt wurden. Berichtet werden, wie von Conover (1999) vorgeschlagen, Quadratsummendifferenzen. Ein signifikantes Ergebnis lag dann vor, wenn diese Werte den kritischen Wert von 822.34 überschritten. Statistisch bedeutsame Abweichungen zwischen den Modellen werden durch * angezeigt.

| Modelle | Modelle | | | | |
|----------|---------|-------|----------|--------|--------|
| | SAC | SDM | 4-States | Streak | HTUVSD |
| SDM | 921* | | | | |
| 4-States | 1492* | 571 | | | |
| Streak | 1660* | 739 | 168 | | |
| HTUVSD | 3060* | 2139* | 1568* | 1400* | |
| HTSD | 3112* | 2191* | 1620* | 1452* | 52 |

der Differenz zwischen dem maximal erreichten *AIC*-Wert und dem *AIC*-Wert des bestpassenden Modells. Der Schritt der Gewichtung stellt den wichtigsten Unterschied zwischen dem Quade-Test und der Friedman-Varianzanalyse über Rangdaten dar, die zum gleichen Zweck eingesetzt werden kann, aber gemeinhin als weniger teststark gilt. Eine genaue Beschreibung des Vorgehens bei der Durchführung eines Quade-Tests kann bei Conover (1999, S. 373 ff.) gefunden werden. Das Vorgehen, das dem im Folgenden zu berichtenden Quade-Test sowie den weiterführenden multiplen Vergleichen zu Grunde lag, folgte den von Conover explizierten Anweisungen.

Die Unterschiede zwischen den Modellen hinsichtlich der von ihnen eingenommenen Rangpositionen erwiesen sich in einem entsprechenden Test als statistisch bedeutsam ($F_{5,225} = 16.96$). Entsprechend ließen sich in einem zweiten Schritt multiple Vergleiche über die Rangpositionen durchführen, die die Modelle bei der Anpassung an die jeweiligen Experimente erreicht haben. In Tabelle 9.1 werden die Ergebnisse dieser multiplen Vergleiche berichtet. Festgehalten werden in dieser Tabelle die für jeweils zwei Modelle paarweise voneinander subtrahierten Summen der gewichteten Rangdifferenzen. Der Betrag dieser Differenzen ist in der oben genannten Tabelle festgehalten worden und konnte mit einem kritischen Testwert verglichen werden, der sich auf der Basis der *t*-Statistik berechnen ließ (vgl. Conover, 1999, S. 375). Bei einem α -Fehler-Niveau von 5% betrug der kritische Testwert für die anstehenden Tests 822.34. In der Ergebnistabelle werden alle Differenzbeträge, die größer ausfielen als der kritische Testwert durch ein Sternchen markiert.

Signifikante Unterschiede ließen sich zwischen dem SAC-Modell und allen übrigen Modellen nachweisen. Das SAC-Modell erwies sich in dieser Analyse also als statistisch bedeutsam besser als alle übrigen Modelle. Das SDM-Modell, das Streak-Modell und das 4-States-Modell unterschieden sich hinsichtlich des von ihnen erreichten Ranges nicht voneinander. Alle drei Modelle schnitten aber signifikant besser ab als die beiden Hochschwellen-Signalentdeckungsmodelle. Die letzten beiden Model-

le, das HTSD- und das HTUVSD-Modell belegten zwar signifikant schlechtere Rangplätze als die übrigen Modelle, untereinander waren ihre Rangpositionen hingegen vergleichbar.

9.2 Bootstrap-Modellselektion

Im vorherigen Abschnitt wurde die Datenanpassungsleistung anhand zweier Datenanpassungsindizes – der Likelihood-Quotienten-Statistik G^2 und dem Informationskriterium AIC_c – analysiert. Sowohl bei der Auswertung der Datenanpassung anhand der Likelihood-Quotienten-Statistik als auch bei der Analyse des Akaike-Informationskriteriums wurde dabei für die Modellkomplexität nur durch die Berücksichtigung der Anzahl freier Parameter korrigiert. Bei den vorgestellten G^2 -Tests geschah dies durch die Wahl der geeigneten Vergleichsstatistik. Als solche wird im Allgemeinen die χ^2 -Verteilung angesehen, wobei die Freiheitsgrade von der Zahl der aus den Daten zu schätzenden Parameter abhängen. Bei der AIC -Statistik floss die Zahl der zu schätzenden Parameter in den Strafterm ein, um den die ML korrigiert wurde. Weitere denkbare Quellen, die dazu beitragen, dass ein Modell komplexer ist als ein anderes, wie z. B. Unterschiede in der mathematischen Form, wurden in keinem dieser Indizes berücksichtigt.

Inwiefern es solchen anderen Formen der Modellkomplexität zugeschrieben werden muss, dass einzelne der Modelle besser abschnitten als andere, sollte in den nun folgenden Analysen näher betrachtet werden. Dazu war zunächst ein Maß erforderlich, das jene Modellkomplexität erfasst, die nicht in der Anzahl freier Parameter begründet liegt. Wie in Kapitel 5 aufgezeigt wurde, lässt sich die Modellkomplexität indirekt über Verfahren der Bootstrap-Modell-Selektion erfassen. Die Bootstrap-Modell-Selektion zielt darauf ab, durch iterative Verfahren ein Maß für den Optimismus eines Modells zu bestimmen, das umso größer ausfällt, je flexibler das Modell bei der Datenanpassung auf oberflächliche Veränderungen der Daten reagiert. Da die Flexibilität eines Modells direkt durch seine Komplexität bedingt ist, lässt sich der geschätzte Optimismus eines Modells als Ausdruck seiner Komplexität interpretieren.

Damit für jedes der 46 Experimente eine Schätzung des Optimismus vorgenommen werden konnte, wurde durch zufälliges Ziehen der Beobachtungen mit Zurücklegen entsprechend ihrer Auftretenswahrscheinlichkeit in den Originaldaten insgesamt 300 Replikationen eines jeden Datensatzes gewonnen. Anhand der Replikationsdatensätze wurden ML-Schätzer bestimmt. Die resultierenden Parameterschätzwerte wurden anschließend daraufhin untersucht, in welchem Umfang sie mit den Originaldaten vereinbar waren. Als Maß für den Grad dieser Vereinbarkeit wurde die rechnerisch ermittelte \mathcal{LL} verwendet, die die so bestimmten Parameterschätzwerte bei den Originaldaten erreichten. Wie in Kapitel 5 vorgestellt, wurde als Optimismusindex das Mittel der Differenzen zwischen dem logarithmierten Maximum-Likelihood-Wert, der bei der Schätzung der Bootstrap-Parameterschätzungen ermittelt worden ist, und der \mathcal{LL} , mit der die Originaldaten anhand der Bootstrap-Parameterschätzungen erklärt werden konnten, verwendet.

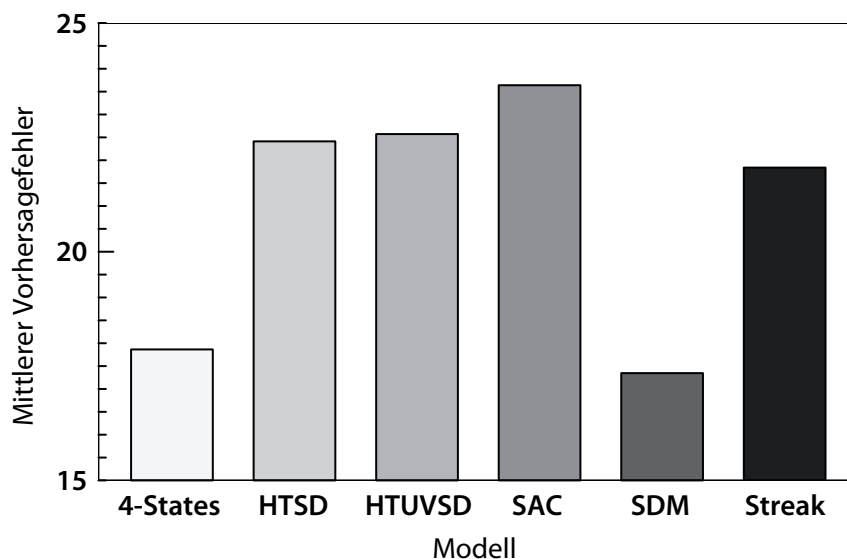


Abbildung 9.5: Dargestellt wird der mittlere Modelloptimismus.

Abbildung 9.5 liefert einen Überblick über die mittleren Optimismus-Werte der zu vergleichenden Modelle. Insgesamt bewegte sich der mittlere Optimismus zwischen 17.35 für das fünfparametrische SDM-Modell und 23.64 für das sechsparemetrische SAC-Modell. Dazwischen lagen das 4-States-Modell mit 17.82, das Streak-Modell mit 21.84 sowie die beiden Yonelinas-Modelle mit 22.42 (HTSD) und 22.57 (HTUVSD). An den Korrekturwerten ist erkennbar, dass die Anzahl der freien Parameter zwar einen Einfluss auf die Vorhersagegenauigkeit der Modelle hatte, dass die parametrische Komplexität jedoch nicht die einzige Quelle für Vorhersageungenauigkeiten gewesen zu sein scheint. So war zwar der Optimismus-Index des fünfparametrischen SDM-Modells der beste aller Modelle, das zweite fünfparametrische Modell HTSD hingegen folgte erst an vierter Stelle nach den beiden sechsparemetrischen Modellen 4-States und Streak. Ein weiterer Einfluss kann in der mathematischen Komplexität der Modelle vermutet werden, die bei den Modellen, die auf die signalentdeckungstheoretische Basisannahmen eines normalverteilten Evidenzmaßes zurückgreifen, augenscheinlich höher ist als die Komplexität des reinen Schwellenmodells 4-States.

In der im Folgenden dargestellten Auswertung wurde der Optimismus-Wert von dem Maximum-Likelihood-Wert abgezogen, der bei der Datenanpassung der jeweiligen Modelle an die empirischen Daten ermittelt wurde. Damit wird der offensichtliche Fehler, der verhindert, dass das Modell die Daten perfekt wiedergeben kann, um das Ausmaß der auf die Modellkomplexität zurückzuführenden Verzerrung korrigiert (Efron & Tibshirani, 1993, S. 252). Auf der Basis dieser um die Modellkomplexität bereinigte Schätzung der Datenanpassungsleistung wurden die Modelle erneut hinsichtlich ihrer deskriptiven Adäquatheit verglichen.

Da für den so gewonnenen Datenanpassungsindex weder bekannt ist, ob er einer

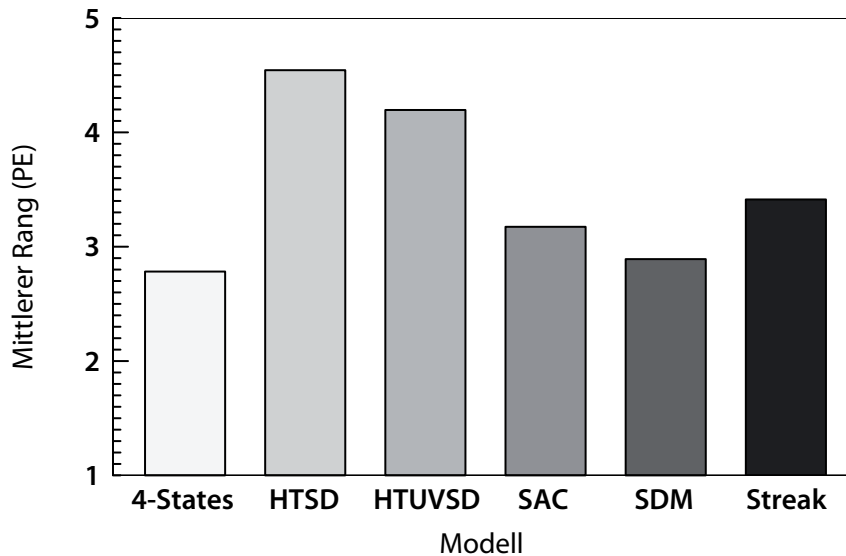


Abbildung 9.6: Dargestellt werden die mittleren Ränge, die die einzelnen Modelle über alle Datensätze hinweg eingenommen haben. Die Ränge wurden dabei anhand der um den Modelloptimismus korrigierten Likelihood ermittelt; kleinere Ränge deuten ein besseres Abschneiden an.

bestimmten Verteilungsfunktion folgt, wie es für die G^2 -Statistik galt, noch bekannt ist, ob er geeignet ist, die Adäquatheit verschiedener Modelle direkt miteinander zu vergleichen, wie es für die AIC -Statistik angenommen werden konnte, waren unter Berücksichtigung des Optimismus weniger Analysen durchführbar als es bei Vernachlässigung der mathematischen Komplexität der Modelle möglich war. Bei einem solchen Verzicht auf die quantitative Interpretation der durch die BMS-Technik ergänzten Datenanpassungsindizes ist jedoch nach wie vor die Analyse der ordinalen Struktur der Datenanpassungsleistung möglich. Wie bei der Analyse der AIC -Rangpositionen wurde dabei auch für die um den Optimismus korrigierte Datenanpassungsleistung für jedes Experiment ermittelt, wie gut die einzelnen Modelle hinsichtlich dieses Maßes abgeschnitten haben.

Abbildung 9.6 gibt die mittlere Rangposition der Modelle wieder. In der Abbildung wird deutlich, dass das Zwei-Hochschwellen-Modell 4-States mit einem mittleren Wert von 2.78 den besten Rang einnahm. An zweiter Stelle folgte das eindimensionale Signalentdeckungsmodell SDM mit 2.89, an das sich mit 3.17 (SAC) und 3.41 (Streak) die beiden zweidimensionalen Signalentdeckungsmodelle anschlossen. Die letzten Positionen nahmen die beiden Varianten des Yonelinas-Modells ein. Von diesen beiden erreichte das sechsparametrische HTUVSD-Modell mit einem mittleren Rang von 4.17 einen besseren Platz als das nur fünf Parameter umfassende HTSD-Modell, das einen mittleren Rangwert von 4.54 aufwies. Diese Rangplätze erwiesen sich in einem Quade-Test auf globaler Ebene signifikant von einer Lösung verschie-

Tabelle 9.2: Dargestellt werden multiple Vergleiche hinsichtlich der Rangpositionen, die den Modellen anhand der den Modelloptimismus umfassenden Datenanpassungswerte, die sie in den einzelnen Experimenten erlangt haben, zugewiesen wurden. Berichtet werden, wie von Conover (1999) vorgeschlagen, Quadratsummendifferenzen. Ein signifikantes Ergebnis liegt dann vor, wenn diese Werte den kritischen Wert von 887.39 überschreiten. Statistisch bedeutsame Abweichungen zwischen den Modellen werden durch * angezeigt.

| Modelle | Modelle | | | | |
|---------|----------|-------|-------|--------|--------|
| | 4-States | SDM | SAC | Streak | HTUVSD |
| SDM | 22 | | | | |
| SAC | 445 | 423 | | | |
| Streak | 539 | 517 | 94 | | |
| HTUVSD | 1747* | 1725* | 1302* | 1208* | |
| HTSD | 2152* | 2130* | 1707* | 1613* | 405 |

den, bei der alle Modelle im Mittel denselben Rang einnahmen ($F(5, 225) = 8.21$). Entsprechend wurde in multiplen Vergleichen genauer analysiert, welche der Modelle sich signifikant voneinander unterschieden

Die Ergebnisse der durchgeführten multiplen Vergleiche auf der Basis des Quade-Tests werden in Tabelle 9.2 wiedergegeben. Anhand der dort dargestellten Testergebnisse ließ sich feststellen, dass sich die besten vier Modelle nicht bedeutsam voneinander unterschieden. Weiter zeigte sich, dass die Rangpositionen dieser Gruppe von Modellen statistisch bedeutsam kleiner war als die Ränge der beiden Varianten des Yonelinas-Modells. Für diese beiden Modelle hingegen gilt, dass sie eine zweite Gruppe bildeten, deren Rangpositionen sich nicht bedeutsam voneinander unterschieden.

9.3 Diskussion

In diesem Kapitel wurden verschiedene Wege genutzt, um die deskriptive Adäquatheit der konkurrierenden Modelle zu quantifizieren. Verwendet wurden dazu zum einen Likelihood-Quotienten-Tests und zum anderen *AIC*-basierte Tests. Auf der Basis des letzteren Maßes wurden neben den Häufigkeiten, mit denen ein Modell das *AIC*-bestpassende Modell war und der Anzahl der Untersuchungen, für die das Modell nicht bedeutsam schlechter abschnitt als das bestpassende Modell, *AIC*-Rangreihen bestimmt.

Anhand der *AIC*-Rangreihen ließ sich ein genaueres Bild von dem Gefüge der Modelle zeichnen, indem bestimmt wurde, wie hoch der Rangplatz eines jeden Modells im Mittel ausfiel. Als Basis für solche Rangreihenanalysen wurden im ersten Abschnitt dieses Kapitels *AIC*-Werte verwendet, in denen die Komplexität der Modelle nur im Sinne der parametrischen Komplexität berücksichtigt wurde. Im zweiten Abschnitt wurden vergleichbare Rangreihenanalysen durchgeführt, bei denen die

Datenanpassung um den für jeden Datensatz und für jedes Modell spezifischen Optimismus korrigiert wurde. Auf der Annahme aufbauend, dass der Modelloptimismus umso größer ist, je flexibler ein Modell ist, ließ sich vermuten, dass auf diese Weise die generelle Komplexität der Modelle berücksichtigt werden konnte.

Alles in allem lässt sich festhalten, dass die durchgeführten Tests zu einem vergleichsweise klaren Ergebnismuster geführt haben. Von der Analyse der *AIC*-Differenzen einmal abgesehen, in der das Streak-Modell den dritten Platz und das SDM-Modell den vierten Platz eingenommen hat, schnitt die eine Hälfte der Modelle durchgängig besser ab als die andere: Hinsichtlich der deskriptiven Adäquatheit schienen das SAC-Modell, das 4-States-Modell und das SDM-Modell den übrigen Modellen überlegen zu sein.

Betrachtet man ausschließlich die Daten der Analysen, bei denen herkömmliche Datenanpassungsindizes verwendet wurden, so schien sich das SAC-Modell noch einmal von den beiden anderen Modellen abzusetzen. Es war zwar nur am zweithäufigsten das *AIC*-bestpassende Modell, hinsichtlich der anderen drei Kriterien schnitt es jedoch durchgängig am besten ab; es wurde also von allen Modellen am seltensten in G^2 -Tests als nicht passend zurückgewiesen, war unter ihnen im Sinne von *AIC*-Differenzen am häufigsten nicht bedeutsam schlechter als das jeweilige bestpassende Modell und wies den kleinsten und damit besten mittleren *AIC*-Rangplatz auf.

Das SDM-Modell und das 4-States-Modell hoben sich in diesem Sinne nicht deutlich voneinander ab. Während das SDM-Modell von allen Modellen am häufigsten als *AIC*-bestpassendes Modell abschnitt und den zweitniedrigsten *AIC*-Rangplatz einnahm, wurde es in G^2 -Tests häufiger als das 4-States-Modell zurückgewiesen und bewährte sich in *AIC*-Differenztests seltener als das Zwei-Hochschwellen-Modell. Letzteres schnitt in G^2 -Tests und hinsichtlich der Bewährungsrate im Sinne von *AIC*-Differenzen als zweitbestes Modell ab und war von allen Modellen am dritthäufigsten das *AIC*-bestpassende Modell. Zudem erreichte es den drittkleinsten *AIC*-Rangplatz, als die Modellkomplexität ausschließlich durch die Zahl freier Modellparameter erfasst wurde.

Diese vermutete Rangreihe fand sich auch in den statistischen Tests auf Unterschiede zwischen den Modellen hinsichtlich der von ihnen eingenommenen *AIC*-Rangplätze wieder. Es zeigte sich hier, dass das SAC-Modell nicht nur von allen Modellen den kleinsten mittleren Rangplatz belegte, sondern sich zudem bedeutsam von allen anderen Modellen unterschied. Eine zweite Gruppe bildeten die Modelle SDM, 4-States und Streak, die sich ihrerseits nicht nur von dem überlegenen SAC-Modell unterschieden, sondern zugleich bedeutsam besser abschnitten als das HTUVSD-Modell und das HTSD-Modell.

Die Korrektur der Modellkomplexität durch den Modelloptimismus brachte eine deutliche Veränderung der Rangpositionen mit sich. Während das SDM- und das 4-States-Modell von allen Modellen die geringsten Optimismus-Werte aufwiesen, war das SAC-Modell durch den größten Modelloptimismus gekennzeichnet; es produzierte also den größten Vorhersagefehler aller Modelle. Entsprechend verschlechterte sich die Rangposition des SAC-Modells bei Berücksichtigung des Vorhersagefehlers; sie sank unter diesen Bedingungen auf den drittbesten Wert ab. Vor ihm lagen das 4-States-

Modell und das SDM-Modell. Statistisch bedeutsam unterschieden sich die Rangpositionen dieser drei Modelle jedoch weder voneinander noch von der Rangposition des viertplazierten Streak-Modells. Andererseits unterschieden sich diese vier Modelle hinsichtlich der von ihnen eingenommenen Rangpositionen bedeutsam von den beiden Varianten des Yonelinas-Modells. Wie auch bei der Analyse der *AIC*-Rangplätze nahmen HTSD- und HTUVSD-Modell den letzten respektive vorletzten mittleren Rang ein.

Alles in allem schnitten die Yonelinas-Modelle hinsichtlich der deskriptiven Adäquatheit besonders schlecht ab, während sich die anderen Modelle recht gut bewährten. In vielen Tests zeichnet sich das SAC-Modell durch eine besonders gute Leistung aus. Allerdings wies dieses Modell auch einen besonders hohen Modelloptimismus auf. Anders verhielt es sich bei dem 4-States- und dem SDM-Modell; diese agierten ebenso wie das Streak-Modell bei Berücksichtigung des Modelloptimismus nicht mehr bedeutsam schlechter als das SAC-Modell.

Diese Ergebnisse machen deutlich, dass weitere Vergleichsschritte unentbehrlich sind, um die Adäquatheit der Modelle umfassend bewerten zu können. Im nächsten Kapitel wird deshalb die Konstruktvalidität der Modelle, die eine wichtige Eigenschaft brauchbarer Modelle darstellt, näher betrachtet.

Kapitel 10

Konstruktvalidität

Während sich die vorherigen Kapitel damit beschäftigten, inwiefern die zu vergleichenden Modelle dazu geeignet waren, die vorhandenen empirischen Daten zu beschreiben, wird sich dieses Kapitel mit der den Modellparametern zugeschriebenen Bedeutung auseinandersetzen. Im Rahmen der Konstruktvalidierung interessiert dabei im Wesentlichen, inwiefern die Parameter der einzelnen Modelle tatsächlich jene theoretischen Entitäten erfassen, zu deren Erfassung sie gedacht sind.

In der Regel geschieht eine solche Konstruktvalidierung experimentell, indem gezeigt wird, dass die Modellparameter dem Einfluss solcher Manipulationen unterliegen, für die begründet erwartet werden kann, dass sie sich auf genau jene theoretischen Größen auswirken, zu deren Erfassung die entsprechenden Modellparameter bestimmt sind. Zugleich sollten Modellparameter, die zur Erfassung einer bestimmten theoretischen Größe vorgesehen sind, nicht durch Manipulationen beeinflusst werden, für die begründet angenommen werden kann, dass sie theoretische Größen betreffen, zu deren Erfassung im Modell andere Parameter dienen sollen. Bei der Konstruktvalidierung einzelner Modellparameter steht also neben der konvergenten Validität, das ist die Frage danach, inwiefern Parameter durch theoretisch verbundene Manipulationen beeinflusst werden, die diskriminante Validität, also die Frage, ob die Parameter *nur* von theoretisch verbundenen Manipulationen affiziert werden, im Zentrum des Interesses.

Im Rahmen der vorliegenden Arbeit ist die Validierung der einzelnen Modellparameter nicht geeignet, um die Modelle hinsichtlich ihrer Konstruktvalidität miteinander zu vergleichen. Gegen ein solches Vorgehen sprechen zum einen praktische Erwägungen: Eine derartige Konstruktvalidierung erfordert die Planung von Experimenten, die spezifisch auf die Parameter eines Modells zugeschnitten sind. Entsprechend findet man Experimente, die dazu geeignet sind, die Konstruktvalidität eines einzelnen Modells zu prüfen, nicht in einer Auswahl von Experimenten, die, wie die vorliegenden, im Hinblick auf andere, meist inhaltliche Fragestellungen entworfen worden sind. Zum anderen spricht ein inhaltlicher Punkt gegen eine Konstruktvalidierung einzelner Modellparameter im Rahmen des Modellvergleichs: Ein Vergleich zwischen verschiedenen Modellen ist immer dann gut vorzunehmen, wenn die Testbedingungen für die konkurrierenden Modelle konstant gehalten werden. Dies ist bisher dadurch gewährleistet worden, dass die Modelle vor dem Hintergrund einer einheitlichen Datenbasis

Tabelle 10.1: *Festgehalten wird, welche Modellparameter benötigt werden, um Antworthäufigkeiten für Distraktorklassen, Zielreizklassen bzw. Distraktor- und Zielreizklassen zu erklären.*

| Theoret. Konstrukt | Modelle | | |
|--------------------|---------------|-------------------|-------------------|
| | 4-States | SAC | Streak |
| Sensitivität | r, k, d | $d_x/d_y, \sigma$ | $d_x/d_y, \sigma$ |
| Antworttendenz | $g_r/g_k/g_g$ | $C_R/C_K/C_O$ | $C_R/C_G/C_O$ |

| Theoret. Konstrukt | Modelle | | |
|--------------------|---------------|---------------|---------------|
| | HTSD | HTUVSD | SDM |
| Sensitivität | r/d' | $r/d'/\sigma$ | d'/σ |
| Antworttendenz | $C_R/C_K/C_O$ | $C_R/C_K/C_O$ | $C_R/C_K/C_O$ |

miteinander verglichen wurden. In einer Arbeit, die sich mit der Konstruktvalidierung von Modellen über einzelne Modellparameter beschäftigen würde, müsste man sich hingegen zwangsläufig davon abwenden, alle Modelle an denselben Daten zu erproben. Vielmehr wäre es notwendig, für jedes Modell einen eigenen Satz kritischer Experimente zu entwerfen, falls sich – wie es bei den hier zu vergleichenden Modellen der Fall ist – die theoretische Einbettung der Modellkomponenten zwischen den Modellen unterscheidet.

Es lässt sich jedoch ein Ausweg aus dem Dilemma zwischen der Notwendigkeit zur Modellspezifität bei der Bestimmung der Konstruktvalidität einzelner Modellkomponenten und dem Bedürfnis nach möglichst einheitlichen Testbedingungen, die einen Vergleich verschiedener Modelle hinsichtlich ihrer Konstruktvaliditäten ermöglichen, finden: Solange sich einzelne Teile der Modelle hinreichend ähnlich sind, um den Einflüssen desselben Satzes an Manipulationen unterworfen zu sein, lässt sich auch an einem einzigen Satz von Experimenten überprüfen, bis zu welchem Grad die einzelnen Modelle das Kriterium der Konstruktvalidität erfüllen.

In dem vorliegenden Modellvergleich wird es deshalb nicht so sehr um die Konstruktvalidierung einzelner Modellparameter gehen, sondern vielmehr darum, wie die Konstruktvalidität der Modelle anhand der bekannten Datensätze festgestellt werden konnte. Dazu musste jedoch eine Betrachtungsebene gefunden werden, auf der die Modellkomponenten aller Modelle so zusammengefasst waren, dass sie einheitliche theoretische Entitäten erfassten. Für die sechs in dieser Arbeit betrachteten Modelle war die niedrigste Ebene, für die das zutrifft, eine Klassifikation der Modellparameter in zwei Gruppen. Für alle Modelle ließ sich eine Gruppe von Parametern identifizieren, die dazu dienen, die Gedächtnissensitivität zu erfassen, und eine zweite Klasse, der alle Parameter zugeschrieben werden konnten, die Antworttendenzen erfassen. Jeder Modellparameter eines jeden Modells konnte, wie in Tabelle 10.1 dargestellt, in genau eine dieser beiden Klassen eingeordnet werden. Gemeinsam liefern die beiden Klassen damit eine erschöpfende Beschreibung der Bestandteile aller Modelle.

Analog zu diesem Vorgehen ließen sich, wie in Kapitel 6 gezeigt wurde, auch die in

den Untersuchungen verwendeten Manipulationen in zwei Klassen einteilen: Für eine Klasse von Manipulation konnte vermutet werden, dass sie das Erinnern erleichterten oder schwerer machten und damit die Gedächtnissensitivität affizierten, während für eine zweite Klasse von Manipulationen Einflüsse auf die Antworttendenz, die die Probanden ihren Urteilen zugrunde legen, erwartet wurden.

Zusammengenommen ermöglichten es diese komplementären Einteilungen der Modellparameter einerseits und der Manipulationen andererseits, dass die Modelle hinsichtlich ihrer Konstruktvalidität miteinander verglichen werden konnten. Wie sich auf der Basis dieser Einteilungen ein Konstruktvaliditätsindex bestimmen ließ und die einzelnen Modelle hinsichtlich dieses Kriteriums bewertet werden konnten, wird in den nächsten beiden Abschnitten behandelt werden.

10.1 Konstruktvaliditätsindex

Durch die klare Zuordnung aller Manipulationen zu einer von zwei Klassen ließen sich – wie im vorherigen Abschnitt angedeutet – alle Modelle durch Parameterrestriktionen in Submodelle überführen, anhand derer sich die konvergente sowie die diskriminante Validität der Modelle untersuchen ließen. Zum Test der diskriminanten Validität wurden dabei Veränderungen zwischen den einzelnen Bedingungen nur für die Parameter zugelassen, die zu der jeweiligen Manipulation in Beziehung standen, während die zur anderen Klasse gehörenden Parameter zwischen den Bedingungen gleichgesetzt wurden. Dadurch entstand für jedes Modell genau ein Submodell, das zu den Daten passen sollte, wenn die Annahme zutraf, dass die beiden Parameterklassen nur den Einflüssen der ihr zugeordneten Manipulationen, nicht aber den Einflüssen von Manipulationen der anderen Klasse unterliegen.

Die so spezifizierten Submodelle ließen ausschließlich Unterschiede zwischen jenen Maßen zu, die zu den mutmaßlich manipulierten theoretischen Konstrukten korrespondierten. Anhand dieser Untermodelle ließ sich also feststellen, inwiefern Manipulationen ausschließlich diejenigen Parameter beeinflussten, zu denen sie vermittels des verbindenden theoretischen Konstruktes der Sensitivität bzw. der Antworttendenz in Beziehung standen. Diese Fragestellung wiederum ist gleichbedeutend mit der Frage danach, inwiefern die Modelle sich im Sinne der diskriminanten Validität als konstruktvalide erwiesen.

Neben einer Restriktion der Modelle anhand derer sich ihre diskriminante Validität erfassen ließ, konnten die Modelle auch so restringiert werden, dass sich Submodelle ergaben, mit deren Hilfe die konvergente Validität der Modelle erfasst werden konnte. Die Frage nach der konvergenten Validität eines Modells ließ sich, wie zuvor bereits erörtert wurde, so umformulieren, dass zu klären war, ob eine experimentelle Manipulation in Veränderungen genau jener Parameter resultierte, auf die sie den theoretischen Vorhersagen gemäß Einfluss nehmen sollte.

Submodelle, die diese Hypothese prüfen konnten, mussten derart gestaltet sein, dass sie für ein gegebenes Experiment keine Unterschiede zwischen den Parametern erlaubten, die dem Einfluss der im jeweiligen Experiment vorgenommenen Manipula-

tionen unterliegen sollten. Das heißt, dass in den entsprechenden Modellen die sensitivitätserfassenden Parameter mehrerer Bedingungen gleichgesetzt wurden, wenn sich die Bedingungen in einer Variable unterschieden, für die eine Wirkung auf die Gedächtnissensitivität erwartet wurde. Wurde hingegen von der zwischen den Bedingungen veränderten Variable eine Auswirkung auf die Antworttendenz erwartet, so wurden gerade die antworttendenz erfassenden Parameter auf Gleichheit restringiert. In beiden Fällen war eine Veränderung der Parameter, die der jeweils anderen Klasse zugerechnet werden, erlaubt.

Als Maß für die Konstruktvalidität einzelner Modelle im Sinne der konvergenten Validität sprach in diesem Aufbau, wie häufig die derart restringierten Submodelle in geeigneten Tests zurückgewiesen wurden. Eine schlechte Datenanpassung und damit eine Modellzurückweisung zeigte an, dass das restringierte Modell nicht zu den verwendeten Daten. Dies konnte jedoch zwei Ursachen haben: Einerseits war es möglich, dass das Modell an sich nicht in der Lage war, die Daten zu beschreiben; andererseits konnte es sein, dass nur die eingeführte Restriktion unzulässig war, während das Modell als solches grundsätzlich in der Lage war, die Daten zu beschreiben. Von diesen Ursachen ist nur die letztgenannte Ausdruck konvergenter Validität. Das Modell ist in diesem Fall in der Lage, die Daten zu beschreiben, und der Effekt ist genau bei jenen Modellparametern zu finden, für die seine Wirkung vorhergesagt wurde. Die erstgenannte Ursache mangelnder Datenanpassungsleistung hingegen ist kein Ausdruck konvergenter Validität, sondern vielmehr die Folge unzureichender deskriptiver Adäquatheit des jeweils im Fokus der Analysen stehenden Modells.

Fehlt einem Modell die deskriptive Adäquatheit, so wirkt sich dies zusätzlich auf die Ergebnisse der Tests zur Feststellung der diskriminanten Validität aus, da auch in diesen Tests ein Mangel an deskriptiver Adäquatheit zu signifikanten Testergebnissen führt: Die Submodelle werden nicht wegen der Restriktion zurückgewiesen, die sie vom Ursprungsmodell unterscheidet, sondern wegen fehlender Adäquatheit des Ursprungsmodells an sich. Hierbei stellt sich die Lage allerdings anders dar als im Falle der konvergenten Validität, da Zurückweisungen des Submodells gegen die diskriminante Validität des jeweiligen Modells sprechen, während die Modellbewährung als Anzeichen für die Modellgüte gewertet wird. Ein Mangel an deskriptiver Adäquatheit führt also zu einer ins Positive verzerrten Schätzung der konvergenten Validität und gleichzeitig zu einer ins Negative verzerrten Schätzung der diskriminanten Validität.

Dieses Dilemma ist gleichzeitig grundlegendes Problem und Teil der im Folgenden vorzuschlagenden Lösung desselben: Betrachtet man konvergente und diskriminante Validität nämlich nicht als zwei unabhängige Maße, sondern als Ausdruck des Gütekriteriums der Konstruktvalidität, so lässt sich in der Verrechnung dieser beiden Maße ein Ausweg finden.

Wie eingangs erwähnt, sollte sich ein konstruktvalides Modell durch konvergente und diskriminante Validität auszeichnen. Bisher wurde herausgearbeitet, dass signifikant ausfallende Datenanpassungstests für eine Klasse von Submodellen mit fehlender diskriminanter Validität oder fehlender deskriptiver Adäquatheit erklärt werden müssen, während sie für eine zweite Klasse von Submodellen konvergente Validität anzeigen oder aus der fehlenden deskriptiven Adäquatheit des Modells folgen. So-

lange man annehmen kann, dass sich das Fehlen deskriptiver Adäquatheit in beiden Analysen im gleichen Maße manifestiert, kann die Kontamination der einzelnen Maße durch die Bildung ihrer Differenz unschädlich gemacht werden. Die Differenz zwischen den Häufigkeiten, mit denen hypothesendiskonform bzw. hypothesenkonform spezifizierte Modelle zurückgewiesen werden, beinhaltet unter diesen Umständen ausschließlich Informationen über die Validität der Modelle: Es handelt sich bei ihr um die Differenz zwischen konvergenter Validität und fehlender diskriminanter Validität.

Eine alternative Auswertungsform, bei der darauf verzichtet werden konnte, die Daten betreffend der konvergenten und der diskriminanten Validität zu einem einzigen Validitätsmaß zusammenzufassen, stellt der Einsatz von G^2 -Differenztests dar. In diesen Tests wurde die Verschlechterung von \mathcal{LL} -Werten genutzt, um die Zulässigkeit von Restriktionen zu überprüfen. Dazu wurde, wie in Kapitel 5 beschrieben, die doppelte Differenz zwischen den \mathcal{LL} -Werten eines stärker restringierten Modells und denen eines Obermodells, das das stärker restringierte Modell beinhaltet, gebildet und auf Signifikanz überprüft. Als Teststatistik wurde dabei die χ^2 -Statistik herangezogen, wobei deren Freiheitsgrade der Anzahl der Parameter entspricht, die das Obermodell mehr enthielt als das restringierte Modell. Zulässig sind solche Tests allerdings nur dann, wenn das Ausgangsmodell, welches in diesem Falle dem verwendeten Obermodell entspricht, selbst gültig ist. Dies hatte zur Folge, dass zum Test der konvergenten und diskriminanten Validität nur jene Datensätze aufgenommen werden konnten, für die das jeweils in Frage stehende Modell eine adäquate Datenanpassung aufwies.

10.2 Ergebnisse

Um die Modelle daraufhin zu untersuchen, inwiefern sie unter der hypothesenkonformen und der hypothesendiskonformen Restriktionsbedingung in der Lage waren, sich an die Daten anzupassen, wurden Tests auf Basis der Likelihood-Quotienten-Statistik vorgenommen. Diese Tests wurden jeweils für α -Niveaus von .05 und .01 durchgeführt, um zu überprüfen, inwiefern die erhaltenen Ergebnisse von der Wahl des Fehlerniveaus abhingen, auf dem getestet wurde.¹

Damit die Daten für jedes Experiment so viele freie Datenkategorien aufwiesen, dass für jedes der Modelle eine Anpassung der entsprechenden Submodelle möglich wurde, erfuhren die Datensätze, die aus unvollständigen Designs stammten, eine besondere Behandlung. Für diese Datensätze war es erneut notwendig, die Zahl der Distraktorbedingungen der Zahl der Zielreizbedingungen anzupassen. Anders als bei den in Kapitel 7 geschilderten Analysen, in denen die Betrachtung der Parameterschätzungen im Vordergrund stand, war in diesem Analyseabschnitt die Datenanpassungsleistung von primärer Bedeutung. Daher wurde bei der nun zu diskutierenden Analyse darauf verzichtet, einzelne Beobachtungen mehrmals zu verwenden, um eine künstliche Erhöhung der Teststärke zu vermeiden. Stattdessen wurden die in den betreffenden Experimenten auf Distraktoren anfallenden Reaktionen so zu gleichen Teilen

¹Die Datenanpassungsindizes, die bei der Anpassung hypothesenkonformer und -diskonformer Modellvarianten resultierten, liefern Anhang D und Anhang E.

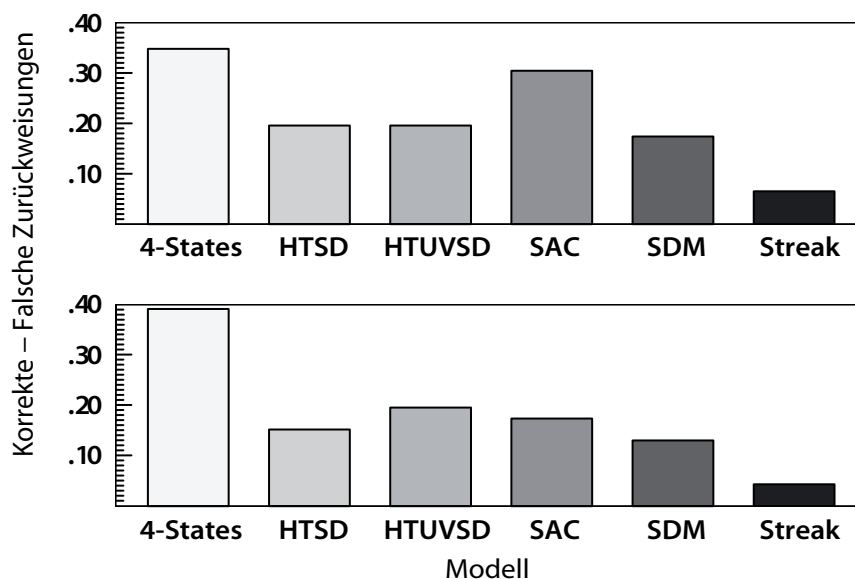


Abbildung 10.1: Dargestellt wird die Differenz der Häufigkeiten von Modellzurückweisungen zwischen hypothesendiskonform und hypothesenkonform spezifizierten Submodellen. Getestet wurde die Datenanpassungsgüte anhand von G^2 -Tests unter Verwendung der α -Niveaus von .01 (oben) und .05 (unten).

aufgeteilt, dass in der Analyse für jede Zielreizbedingung eine Gruppe von Distraktoren existierte. Durch dieses Verfahren blieb die absolute Zahl der modellierten Beobachtungen konstant, so dass eine künstliche Erhöhung der Teststärke ausgeschlossen werden konnte.

In Abbildung 10.1 werden die Ergebnisse dieser Tests wiedergegeben. Berichtet wird die Differenz der Häufigkeiten für signifikante G^2 -Tests zwischen den beiden oben beschriebenen Subtestklassen. Dabei wurde die relative Zahl unerwünscht positiver Tests von der relativen Zahl an Tests mit erwünscht positivem Ausgang abgezogen. Die Konstruktvalidität eines Modells kann also als umso höher eingeschätzt werden, je größer die in dieser Abbildung dargestellte Differenz ausfiel.

Wie Abbildung 10.1 entnommen werden kann, hatte das Signifikanzniveau nur einen geringen Einfluss auf die Ergebnisse des Modelltests: Einzig das Ergebnis des SAC-Modells fiel unter den strengeren Testbedingungen augenfällig besser aus als unter den laxeren Testbedingungen. Für die übrigen Modelle veränderte sich der Differenzwert nur minimal.

Insgesamt schnitt das 4-States-Modell unter beiden Testbedingungen mit Differenzwerten von $.39^2$ respektive $.35$ am besten ab. Bei Verwendung eines α -Niveaus von .01 folgte darauf mit einem Differenzwert von $.30$ das SAC-Modell. Bei Verwen-

²Die erstgenannten Werte beziehen sich, falls nichts anderes mitgeteilt wird, auf die laxere Testbedingung mit einem α -Niveau von $\alpha = .05$, während sich die an zweiter Stelle genannten Werte auf die strengere Testbedingung mit $\alpha = .01$ beziehen.

dung des 5%- α -Fehlerniveaus lag der Differenzwert des SAC-Modells hingegen nur bei .17. Während das SAC-Modell bei strengeren Tests also den zweithöchsten Validitätsindex aufwies, nahm es unter den laxeren Testbedingungen lediglich den dritten Rang ein; hier erreichte das HTUVSD-Modell mit .20 den zweithöchsten Wert. Der Validitätswert des HTUVSD-Modells veränderte sich zwischen den Testbedingungen nicht, er betrug auch unter der strengeren Testbedingung .20. Auf einem α -Niveau von .01 erreichte das HTSD-Modell mit .20 den gleichen Wert wie die um die Annahme ungleicher Varianzen ergänzte Variante des Yonelinas-Modells, wohingegen ersteres auf dem 5%-Niveau nur einen Wert von .15 annahm. Damit schnitt es auf beiden α -Niveaus etwas besser ab als das SDM-Modell (.13, .17). Das schlechteste Verhältnis von erwünschten zu unerwünschten Zurückweisungen wies das Streak-Modell mit Differenzwerten von .04 respektive .07 auf.

Als Alternative zu dem Differenzwert, der die Konstruktvalidität der Modelle bestimmbar macht, wurden G^2 -Differenz-Tests vorgeschlagen. Diese haben allerdings den Nachteil, dass sie nur dann anwendbar sind, wenn die Modelle in unrestringierter Form zu den Daten passen.

Wie im vorausgegangenen Kapitel deutlich geworden ist, unterschieden sich die Modelle vor dem Hintergrund der vorhandenen Daten hinsichtlich ihrer deskriptiven Adäquatheit. Folglich war zu erwarten, dass die Anzahl der Datensätze, die bei der anstehenden Analyse berücksichtigt werden konnten, zwischen den einzelnen Modellen variieren würde. Um überprüfen zu können, ob die Modelle in ihrer Ausgangsform eine adäquate Beschreibung der Daten darstellten, wurden für die fünfparametrischen Modelle Likelihood-Quotienten-Tests durchgeführt. Die deskriptive Adäquatheit der sechsparametrischen Modelle konnte hingegen lediglich dadurch überprüft werden, dass ermittelt wurde, ob ihr G^2 -Wert nahe bei Null lag. Da diesen Modellen ebenso viele Parameter zur Verfügung standen wie die Daten freie Kategorien bereithielten, sollten diese Modelle die Daten perfekt wiedergeben können und entsprechend einen Datenanpassungswert von annähernd Null aufweisen. Durch die Verwendung eines Vergleichswertes von 0.01 wurde dafür Sorge getragen, dass Schätzungenauigkeiten in der Analyse unberücksichtigt blieben.

Wie Tabelle 10.2 entnommen werden kann, schwankte die Anzahl der nach dieser Prüfung der deskriptiven Adäquatheit verwendbaren Datensätze abhängig von dem jeweils betrachteten Modell zwischen 2 für das HTUVSD-Modell und 45 für das Streak-Modell. Dazwischen lagen das HTSD-Modell mit 17, das SDM-Modell mit 26, das 4-States-Modell mit 36, sowie das SAC-Modell mit 44 Datensätzen, bei denen G^2 -Differenztests möglich waren.

Wegen der geringen Anzahl von Datensätzen, für die sich das HTUVSD-Modell als passend erwiesen hat, wurde dieses von den weiteren Analysen ausgenommen. Für die restlichen Modelle wurde ermittelt, inwiefern jene Restriktionen, die zu im oben dargelegten Sinne hypothesenkonformen bzw. -diskonformen Submodellen führten, in statistisch bedeutsame Verschlechterungen der Datenanpassung mündeten. Die Zahl der Datensätze, für die die durchgeführten Tests zu signifikanten Verschlechterungen führten, kann ebenfalls Tabelle 10.2 entnommen werden. Zudem findet sich dort das Verhältnis von signifikanten Ergebnissen bei hypothesendiskonformen Re-

Tabelle 10.2: Überblick über die Ergebnisse von G^2 -Differenztests zur Analyse der Konstruktvalidität. Berichtet wird die Zahl zulässiger G^2 -Differenztests (N_{Tests}), die Zahl signifikanter G^2 -Differenztests bei hypothesenkonformer Parameterrestriktion ($N_{konform}$) sowie bei hypothesendiskonformer Parameterrestriktion ($N_{diskonform}$). Zusätzlich werden Wettquotienten aufgeführt, die angeben, wie viele erwünschte signifikante Ergebnisse auf ein statistisch bedeutsames, aber unerwünschtes Ergebnis kommen ($\frac{N_{diskonform}}{N_{konform}}$).

| Modell | N_{Tests} | Signifikante Befunde | | $\frac{N_{diskonform}}{N_{konform}}$ |
|----------|-------------|----------------------|------------------|--------------------------------------|
| | | $N_{konform}$ | $N_{diskonform}$ | |
| SDM | 26 | 14 | 18 | 1.29 |
| HTSD | 17 | 8 | 12 | 1.5 |
| HTUVSD | 2 | -/- ^a | -/- | -/- |
| SAC | 44 | 19 | 26 | 1.37 |
| Streak | 45 | 30 | 32 | 1.07 |
| 4-States | 34 | 12 | 26 | 2.17 |

^aFür dieses Modell sind G^2 -Differenztests für zu wenige Datensätze möglich, als die Ergebnisse interpretierbar wären.

striktionen zu der Zahl signifikanter Ergebnissen bei hypothesenkongformen Restriktionen. Wie bereits erläutert wurde, sprechen erstere für die konvergente Validität des Modells, während das Auftreten letzterer gegen die diskriminante Validität des jeweiligen Modells spricht. Entsprechend lässt sich das berichtete Maß deuten als die Anzahl erwünschter Effekte relativiert an der Anzahl statistisch bedeutsamer aber unerwünschter Effekte. Entsprechend gilt, dass größere Werte in stärkerem Maße für die Konstruktvalidität des Modells sprechen als kleinere Werte. Problematisch an der Betrachtung dieser Wettquotienten ist jedoch, dass die Anzahl der möglichen Tests in ihnen keine Berücksichtigung findet.

Die erhaltenen Ergebnisse wiesen dennoch in dieselbe Richtung wie die Befunde der vorhergehenden Analyse zum Vergleich der Konstruktvaliditäten: Der höchste Quotient ließ sich mit 2.17 für das 4-States-Modell beobachten. Darauf folgten das HTSD-Modell mit einem Quotienten von 1.5 und das SAC-Modell mit einem von 1.37. Die kleinsten Quotienten wiesen auch in diesem Test das SDM-Modell mit 1.29 sowie das Streak-Modell mit 1.07 auf. Diese Reihenfolge entsprach damit weitgehend der Reihenfolge, die bei der zuvor berichteten Analyse der Differenzwerte beobachtet wurden.

10.3 Diskussion

In diesem Kapitel wurde die Frage danach gestellt, inwiefern sich die sechs Modelle hinsichtlich ihrer Konstruktvalidität unterscheiden. Dazu wurden in Anlehnung an das

Vorgehen bei der experimentellen Konstruktvalidierung einzelner Modelle Indizes für konvergente und diskriminante Validität bestimmt. Damit ein Vergleich der Modelle möglich werden konnte, war es notwendig, eine Betrachtungsebene zu finden, auf der sich die Modellbestandteile der einzelnen Modelle so zusammenfassen ließen, dass die resultierenden Parametergruppen erstens für alle Modelle gleichbedeutend waren und zweitens ein Bezug zu den in der Literatur realisierten Manipulationen hergestellt werden konnte. Beides gelang auf einer Ebene, auf der eine Klasse sensitivitätserfassender Parameter von einer zweiten Klasse antwortendenzerfassender Parameter unterschieden wurde: Da sich eine analoge Einteilung für alle Manipulationen vornehmen ließ, konnte diese Unterscheidung dazu genutzt werden, um konvergente und diskriminante Validität der Modelle zu bestimmen.

Zur Erfassung der konvergenten sowie der diskriminanten Validität wurden verschiedene Parametrisierungsvarianten der Modelle an die einzelnen Datensätze angepasst. Dabei wurde beobachtet, wie häufig korrekt spezifizierte Modelle fälschlicherweise zurückgewiesen wurden bzw. wie häufig fehlspezifizierte Modelle zu Recht zurückgewiesen wurden. Das erste Datum wurde dabei als mangelnde konvergente Validität gedeutet, während letzteres als Hinweis diskriminanter Validität gewertet wurde.

Diese beiden Datenanpassungsdurchgänge wurden auf zwei Arten dazu eingesetzt, um Konstruktvaliditätsindizes zu bestimmen, die einen Vergleich der Modelle erlaubten. Da beiden Ansätzen spezifische Vor- und Nachteile zu Eigen sind, ist es umso interessanter, dass sich die Ergebnisse der Analysen weitgehend entsprachen.

Unabhängig von der zu ihrer Bestimmung eingesetzten Methode schnitt das 4-States-Modell besser ab als die übrigen Modelle. Ebenfalls eine positive Bilanz wies das SAC-Modell auf, das abhängig von der eingesetzten Methode sowie von dem angelegten Signifikanzniveau den zweithöchsten bzw. dritthöchsten Validitätsindex aufwies. In den beiden Fällen, in denen das SAC-Modell als drittbestes Modell abschloss, lag es in hinter dem HTSD-Modell bzw. dem HTUVSD-Modell.

In den übrigen Fällen erreichten die beiden Yonelinas-Modelle unauffällige Konstruktvaliditätswerte, die es ihnen erlaubten, einen mittleren Platz zwischen den Modellen einzunehmen. Allerdings konnte für das HTUVSD-Modell nur eine der Analysen durchgeführt werden, da aufgrund seiner mangelnden deskriptiven Adäquatheit zu viele Datensätze von den G^2 -Differenztests ausgeschlossen werden mussten.

Das SDM-Modell sowie das Streak-Modell wiesen unter den Modellen die geringsten Konstruktvaliditätswerte auf, wobei das Streak-Modell durchgängig die kleinsten Werte erzielte, während das eindimensionale Signalentdeckungsmodell durchgängig die zweitkleinsten Werte produzierte.

Zusätzlich wäre an dieser Stelle eine statistische Absicherung der zwischen den Modellen festgestellten Unterschieden wünschenswert. Eine solche lässt sich mit den heutzutage zur Verfügung stehenden Methoden jedoch nicht erreichen. An dieser Stelle muss deshalb die dargestellte deskriptive Betrachtung der Ergebnisse ausreichen.

Kapitel 11

Abschließende Diskussion

In den vorherigen Kapiteln sind die in Kapitel 4 vorgestellten Modelle hinsichtlich zentraler Modellgütekriterien verglichen worden. Neben der Plausibilität der Parameterschätzwerte wurde dabei die deskriptive Adäquatheit der Modelle untersucht. Zudem wurde die Konstruktvalidität der Modelle näher beleuchtet und in ROC-Analysen beobachtet, inwiefern die empirischen ROC-Kurven mit den Vorhersagen der Modelle übereinstimmten. In diesem Kapitel sollen zunächst die Ergebnisse dieses Modellvergleichs zusammengefasst und kritisch bewertet werden. Anschließend wird ein Ausblick auf die Fragen geliefert, die sich vor dem Hintergrund der erlangten Erkenntnisse für zukünftige Forschungsbemühungen stellen.

Wie in Kapitel 7 festgestellt, schienen die Parameterschätzwerte, die die Modelle lieferten, im Allgemeinen mit den theoretisch begründeten Erwartungen übereinzustimmen. In einigen Bedingungen allerdings wichen die Schätzungen von den erwarteten Werten ab. Besonders betroffen waren davon die Streuungsparameter, die die meisten der zu vergleichenden Modelle mit signalentdeckungstheoretischer Fundierung annehmen, sowie der Distraktorentdeckungsparameter des Zwei-Hochschwellen-Modells.

Unter den Streuungsparametern fiel besonders die Schätzung des HTUVSD-Modells aus dem Rahmen. Während man gemeinhin von einem moderaten Unterschied zwischen den Varianzkennwerten der Zielreizverteilung auf der einen Seite und der Distraktorverteilung auf der anderen Seite ausgeht, fiel der Unterschied zwischen diesen beiden Maßen für das HTUVSD-Modell mit einem mittleren Verhältnis von etwa 5/4 allzu groß aus.

Bei den übrigen signalentdeckungstheoretischen Modellen, die Varianzunterschiede zulassen, fiel ein jeweils substantieller Anteil der Schätzungen so aus, dass die Varianz der Distraktorverteilung entgegen der allgemeinen Erwartung größer war als die Varianz der Zielreizverteilung. Anders als bei dem HTUVSD-Modell bewegten sich jedoch die mittleren Streuungsverhältnisse im erwarteten Bereich.

Wie bereits erwähnt, wich die mittlere Schätzung der Distraktorentdeckung, die im Rahmen des 4-States-Modells postuliert wird, deutlich von den üblichen Erwartungen ab; vor dem Hintergrund, dass gemeinhin angenommen wird, dass es sich bei der bewussten Zurückweisung von Distraktoren um ein seltenes Phänomen handelt, fielen

die Schätzwerte überaus hoch aus. Ob, wie vermutet, das 4-States-Modell dazu neigt, bewusste Gedächtnisprozesse generell zu überschätzen, ist eine Frage, die an dieser Stelle nicht beantwortet werden kann. Für diese Vermutung spricht jedoch, dass auch die auf Basis des 4-States-Modells gewonnene Schätzung des r -Parameters im Mittel höher ausfiel als diejenigen Schätzwerte für rekollektive Gedächtnisprozesse, die die beiden Varianten des Yonelinas-Modells lieferten.

Generell ist bei dieser Form der Betrachtung jedoch unklar, ob die Besonderheiten einzelner Schätzwerte auf mangelnde Plausibilität der Modellannahmen zurückgeführt werden kann oder vielmehr mit besonderen Eigenschaften der analysierten Daten erklärt werden muss. Festgehalten werden kann hier allerdings, dass die Abweichungen der Schätzwerte von den allgemeinen Erwartungen für das HTUVSD-Modell und das 4-States-Modell am auffälligsten waren. Andererseits ging auch keines der übrigen Modelle vollkommen unbeschadet aus dieser Analyse hervor.

Im Sinne des Modellvergleichs erhellender sind die mit der Analyse von ROC-Kurven verbundenen Resultate. In Kapitel 8 konnte festgestellt werden, dass die Form der empirisch beobachteten ROC-Kurven am besten den Vorhersagen der beiden zweidimensionalen Signalentdeckungsmodelle SAC und Streak sowie denen des 4-States-Modells entsprach. Im Gegensatz zu den übrigen Modellen sind die Annahmen dieser Modelle vereinbar mit der Beobachtung, dass der auf Remember-Antworten beruhende z -ROC-Punkt überzufällig häufig oberhalb der z -ROC-Geraden zu finden war, die durch die Punkte für Alt-Antworten einerseits und Remember-/Know-Antworten andererseits verlief.

Berücksichtigt man zusätzlich, inwiefern die Modelle anhand empirischer Daten hätten falsifiziert werden können, so schneiden das SAC-Modell und das 4-States-Modell besser ab als das Streak-Modell. Letzteres erwies sich in Simulationen als ebenfalls mit Daten vereinbar, in denen kein systematischer Unterschied hinsichtlich der Steigungen von abrufersbasierten ROC-Geraden und konfidenzbasierten ROC-Geraden bestand. Für die erstgenannten Modelle gehört diese Beobachtung hingegen genauso wie diejenige eines negativen Schnittwinkels der beiden z -ROC-Geraden zu der Gruppe von Beobachtungen, die nicht mit den Modellen vereinbar sind. Aus diesen Gründen ist die Beobachtung eines positiven mittleren Schnittwinkels für das SAC- und das 4-States-Modell bedeutsamer als für das Streak-Modell.

Als auf der Basis der ROC-Geraden falsifiziert gelten können die beiden hybriden Modelle und das SDM-Modell. Während aus ersteren ein negativer Schnittwinkel abgeleitet wurde, entspräche es den Vorhersagen des letzteren, wenn die beiden ROC-Geraden kollinear wären. Der Schnittwinkel hätte demnach also im Mittel Null entsprechen müssen bzw. nicht überzufällig häufig positiv oder negativ ausfallen dürfen.

Eine genauere Betrachtung der ermittelten ROC-Kennwerte wies weiterhin eine überraschend hohe Übereinstimmung mit den auf Basis des 4-States-Modells ermittelten Vorhersagen auf. Das 4-States-Modell hob sich also in dieser Hinsicht positiv von den beiden zweidimensionalen Signalentdeckungsmodellen ab. Dieser Umstand wiederholt sich in manchen der folgenden Analysen. So fand sich das 4-States-Modell bezüglich der deskriptiven Adäquatheit unter den bestabschneidenden Modellen wieder, als die Modellkomplexität in der Analyse durch den Modelloptimismus erfasst

wurde. In den Analysen der *AIC*-Rangplätze hingegen schnitt das SAC-Modell signifikant besser ab als das 4-States-Modell, das SDM-Modell und das Streak-Modell. In beiden Fällen schnitten die Yonelinas-Modelle bedeutsam schlechter ab als die anderen Modelle. Die übrigen Analysen der deskriptiven Adäquatheit anhand von herkömmlichen Datenanpassungstests bestätigten diese Reihenfolge: Ihnen zufolge weisen das SAC-Modell, das SDM-Modell und das 4-States-Modell eine höhere deskriptive Adäquatheit auf als die restlichen Modelle.

Allerdings war gerade das SAC-Modell durch einen besonders hohen Modelloptimismus gekennzeichnet, der zu der Annahme veranlasste, dass diesem Modell eine besonders hohe Komplexität zu Eigen ist, die nicht gänzlich durch die Anzahl der freien Parameter des Modells erfasst wird. Dieser Verdacht ließ sich an den Daten bezüglich der Konstruktvalidität der Modellbestandteile insofern bestätigen, als das SAC-Modell nicht nur dann an Daten passte, wenn es in hypothesenkonformer Weise restringiert wurde, sondern auch häufig mit den Daten vereinbar war, wenn ein hypotheseninkonformes Submodell an die Daten angepasst wurde. Dadurch erreichte dieses zweidimensionale Signalentdeckungsmodell hinsichtlich des Konstruktvaliditätsindex, in dem konvergente und diskriminante Validität zusammengefasst wurden, nur ein mittleres Ergebnis. Im Vergleich dazu schnitt das 4-States-Modell deutlich besser ab. Auffällig ist, dass sich in den Analysen der Konstruktvalidität das eindimensionale Signalentdeckungsmodell und das Streak-Modell deutlich schlechtere Ergebnisse zeigten als die übrigen Modelle.

Betrachtet man diese Ergebnisse im Überblick, so bewährten sich nur das SAC-Modell und das 4-States-Modell in allen Analysen. Die übrigen Modelle wiesen spezifische Defizite auf, die sie in einzelnen Kriterien schlechter agieren ließen als die konkurrierenden Modelle. So demonstrierten die Yonelinas-Modelle insbesondere im Bereich der deskriptiven Adäquatheit Mängel. Das Streak-Modell und das SDM-Modell schnitten hingegen in der Analyse der Konstruktvalidität durchgängig schlechter ab als die übrigen Modelle. Besondere Beachtung verdient der Punkt, dass sich diese Befunde nicht nur anhand parametrischer Methoden finden ließen, sondern – wie zuvor dargestellt – ebenfalls in der Analyse der ROC-Kurven durchschienen.

Offen bleibt allerdings die Frage, welches der beiden Modelle das bessere ist: Während das SAC-Modell bei der Analyse der deskriptiven Adäquatheit anhand herkömmlicher Datenanpassungsindizes besser abschnitt als das 4-States-Modell, zeichnete sich das letztere durch eine höhere Konstruktvalidität aus.

Eine klare Entscheidung zwischen diesen beiden Modelle erfordert gezielte Konstruktvalidierungsstudien, bei denen die theoretische Basis der beiden Modelle näher beleuchtet und auf ihre Gültigkeit hin untersucht wird. Hierzu bieten sich abhängig von dem jeweils im Zentrum der Untersuchung stehenden Modells verschiedene Herangehensweisen an. Für das SAC-Modell empfehlen sich vor allem solche Untersuchungen, in denen die Natur der Gedächtnisparameter d_x und d_y näher beleuchtet wird. Ein Ansatz, der aus diesem Blickwinkel in den Vordergrund rückt, besteht in der Verwendung von Manipulationen, die die Salienz der Enkodierepisode einerseits bzw. die Verarbeitung der Iteminformation andererseits erhöhen. Solche Manipulationen sollten sich den Annahmen des Modells gemäß auf d_y respektive d_x auswirken. Eine Orientie-

rung für den Entwurf solcher Untersuchungen bieten die Arbeiten zum kontextuellen Fächer-(Fan-)Effekt (vgl. Anderson, 1974) von Reder und Mitarbeitern. Diana et al. (2004) haben beispielsweise festgestellt, dass Wörter, die während der Lern- und Abrufphase in einer Schriftart standen, die für viele Reize verwendet wurde, korrekte Alt-Antworten seltener waren als bei Reizen, die in einer weniger häufig vorkommenden Schriftart präsentiert wurden. Für Distraktoren ergab sich ein gegenteiliges Bild: Hier traten dann mehr Fehler auf, wenn Distraktoren in der Schriftart präsentiert wurden, die mit mehr Zielreizen assoziiert war. Dieser kontextuelle Fächer-Effekt lässt sich, wie Diana, Reder, Arndt und Park (2006) erläutern, sehr gut mit dem SAC-Modell erklären, da für die Schriftart als Kontextinformation angenommen werden kann, dass sie mit dem Episodenknoten verbunden wird, und darüber hinaus davon ausgegangen werden kann, dass sich die Aktivierung der Kontextinformation entsprechend der Größe ihres Fächers auf die mit ihr verbundenen Episodenknoten aufteilt. Aus diesen beiden Annahmen folgt, dass die Episodenaktivierung für Episoden, die mit einer Vielzahl anderer Reize in Beziehung stehen, geringer sein muss als für sehr auffällige Reize, die ihre Merkmale nur mit wenigen anderen Episoden teilen. In den Parameterschätzungen des SAC-Modells sollte sich diese Manipulation also in einem Abfall der Schätzungen für die Episodenaktivierung d_y auswirken. Entsprechend lässt sich mit dieser Aufgabe die Vertrauenswürdigkeit des SAC-Modells empirisch überprüfen

Für das 4-States-Modell lassen sich drei unterschiedliche Wege erkennen, die weiter verfolgt werden können, um seine theoretische Basis näher zu beleuchten. Der erste Ansatz besteht in der Durchführung von Untersuchungen, die sich mit der Natur des Distraktorentdeckungsprozesses auseinandersetzen. In diesem Rahmen sollte insbesondere nach einer Antwort auf die Frage gesucht werden, weshalb die Parameterschätzungen für diesen Prozess in den hier verwendeten Untersuchungen höher ausfielen als erwartet. Damit verbunden stellt sich die Frage, ob innerhalb dieses Modells unter Distraktorentdeckung tatsächlich eine bewusste Entdeckung von Distraktoren zu verstehen ist, oder ob in diesem Parameter zusätzlich die Funktion eines Neuheitsdetektorprozesses erfasst wird, wie er beispielsweise von Tulving, Markowitsch, Kapur, Habib und Houle (1994) vorgeschlagen wurde.

Evidenz für einen solchen Neuheitsdetektorprozess fanden Habib, Nyberg und Tulving (2003) in Untersuchungen zum Novelty-Effekt. In solchen Studien wird der Zusammenhang zwischen der relativen Neuheit von Reizen und ihrer Enkodierbarkeit untersucht, indem der eigentlichen Lernphase eines Gedächtnisexperimentes eine Familiarisierungsphase vorgeschaltet wird. In dieser Phase werden die Probanden mit einem Teil der im Experiment zu lernenden Reize vertraut gemacht. Ein späterer Gedächtnistest offenbart für die Reize, die in der Familiarisierungsphase präsentiert worden sind, in der Regel eine schlechtere Gedächtnisleistung als für die im Kontext des nachfolgenden Experiments vollkommen neuen Items. Dieses Befundmuster deuteten die Autoren als Beleg für die Existenz eines Gedächtnissystems, das der Entdeckung von Neuheit dient und sich zusätzlich auf die Enkodierung von Reizen auswirkt.

Untermauert wurde diese Annahme von Habib et al. (2003) zusätzlich durch neuropsychologische Untersuchungen, in denen anhand von bildgebenden Verfahren gezeigt werden konnte, dass die Neuheitsdiskrimination auf der Aktivität umfassender neuro-

naler Netzwerke basiert, in denen der Hippocampus nur einen Teil ausmacht. Da dem Hippocampus in einer Vielzahl von Untersuchungen eine wichtige Funktion beim rekollektiven Abruf von Informationen zugeschrieben wurde (für einen Überblick siehe Yonelinas, 2002), erscheint es möglich, dass Neuheitsentdeckung nicht nur auf kontrollierten Prozessen beruht, sondern auch durch vertrautheitsähnliche, automatische Prozesse vermittelt werden kann. Habib (2001) ging sogar soweit, Neuheits- und Vertrautheitserfassung als zwei Seiten einer Medaille zu betrachten.

Entsprechend kann eine wichtige Frage für zukünftige Forschungsanstrengungen darin gesehen werden, zu ermitteln, inwiefern die Leistung dieses Neuheitserfassungssystems im Rahmen des 4-States-Modells reflektiert wird. Sicher ließe sich durch die Annahme solch eines Systems erklären, weshalb die Schätzungen für den Distraktorentdeckungsparameter d höher ausfielen als erwartet.

Um dieser Frage auf den Grund zu gehen, wären Untersuchungen dienlich, in denen den Probanden wie in den Untersuchungen von Habib et al. (2003) in einer vorgelegten Familiarisierungsphase Distraktoren gezeigt werden. Wenn der Distraktorentdeckungsparameter sensitiv für die Funktionsweise des vermuteten Neuheitsdetektionssystems ist, sollte die Schätzung der Distraktorentdeckung unter dem Eindruck der Familiarisierung kleiner werden. Auf diese Weise könnte ein Beleg dafür gewonnen werden, dass der Parameter d des 4-States-Modells neben den Prozessen bewusster Distraktorentdeckung eine als vertrautheitsbasiert zu betrachtende Form der Neuheitsentdeckung erfasst.

Weitere Fragestellungen, die dazu geeignet sind, die Verallgemeinerbarkeit des 4-States-Modells zu überprüfen, ergeben sich aus dessen enger Verwandtschaft mit weiteren Zwei-Hochschwellen-Modellen. Zwei-Hochschwellen-Modelle wurden in anderen Bereichen der Gedächtnisforschung beispielsweise dazu genutzt, um Daten aus Untersuchungen, die der Prozess-Dissoziations-Prozedur folgen, zu analysieren. Auch einige Modelle zum Quellengedächtnis sind von ihrer mathematischen Struktur her dem 4-States-Modell sehr ähnlich. Um die theoretische Basis des 4-States-Modells einer eingehenderen Prüfung zu unterziehen, ist es sinnvoll, dieses mit den verwandten Modellen der genannten Themengebiete zu vergleichen.

Betrachtet man das Zwei-Hochschwellen-Modell zur Prozess-Dissoziations-Prozedur, das Erdfelder und Buchner (1995) vorgeschlagen haben, so fällt zunächst auf, dass dieses mit dem 4-States-Modell nicht nur die Annahmen teilt, die allen Zwei-Hochschwellen-Modellen gemeinsam sind, sondern ihm auch dahingehend ähnlich ist, dass beide Modelle zwei verschiedene Gedächtnisprozesse annehmen. Entsprechend erscheint es auf den ersten Blick reizvoll, der Frage nachzugehen, inwiefern sich die in den unterschiedlichen Modellen angenommenen Gedächtnisprozesse entsprechen.

Die Antwort auf diese Frage lässt sich meiner Ansicht nach nur empirisch beantworten. Allerdings werfen die Annahmen, die der Remember-Know-Prozedur einerseits und der Prozess-Dissoziations-Prozedur andererseits zugrunde liegen, einen Schatten auf die skizzierte Fragestellung. Da die Trennung zwischen automatischen und kontrollierten Prozessen in der Prozess-Dissoziations-Prozedur darauf abzielt, implizite Gedächtnisprozesse von expliziten zu unterscheiden (vgl. Jacoby, 1991), muss bei einer Übereinstimmung zwischen Remembering und kontrollierter Erinne-

rung auch dem Zustand des Rememberings der Status der Explizitheit zugeschrieben werden. Während diese Annahme unproblematisch erscheint, ist strittig, ob sich der Zustand des Wissens als implizit charakterisieren lässt. Gegen eine solche Annahme, die sich beispielsweise in einer Arbeit von Yonelinas, Kroll, Dobbins, Lazzara und Knight (1998) wieder finden lässt, argumentierte Gardiner (2002; s. a. Gardiner, 2000; Gardiner & Richardson-Klavehn, 2000; Richardson-Klavehn, Gardiner & Java, 1996). Dabei stützte Gardiner sich auf Befunde, die belegen, dass sowohl Remember- als auch Know-Antworten mit dem Grad an kognitiver Kontrolle ansteigen, wie es Ergebnisse von Untersuchungen nahe legen, in denen die Probanden gehalten waren, ihr Rekognitionsurteil in einer Bedingung sehr schnell, d. h. innerhalb der ersten 500 ms nach der Reizpräsentation, abzugeben, während ihnen in einer zweiten Bedingung mit 1500 ms deutlich mehr Zeit für ihre Reaktion zur Verfügung standen (z. B. Toth, 1996; Gardiner et al., 1999, Exp. 1).

Ein weiterer Hinweis darauf, dass in den beiden Aufgaben tatsächlich unterschiedliche Prozesse wirksam werden, kann darin gesehen werden, dass sich das Yonelinas-Modell, das sich im Rahmen der Analyse von Prozess-Dissoziations-Untersuchungen bewährte (vgl. Yonelinas, 1994), in der vorliegenden Untersuchung als nicht adäquat erwiesen hat.

Eine weniger umstrittene Parallele lässt sich zwischen Remember-Know-Modellen einerseits und Modellen des Quellengedächtnisses andererseits ziehen. Dass ein Zusammenhang zwischen dem Quellengedächtnis und dem Abruferleben bestehen sollte, deutete sich bereits in der von Tulving (1985a) gegebenen theoretischen Einführung zum Abruferleben an. Das episodische Gedächtnissystem soll, wie in Kapitel 3 dargelegt, dazu befähigen, Episoden des täglichen Lebens als Ganzes abzurufen, so dass neben der zentralen Information auch spezifische Detailinformationen über die Enkodierung erinnert werden können. Zu diesen spezifischen Detailinformationen ist auch die Quelleninformation zu zählen. Dieser Argumentation zufolge sollten im Zustand des Remembering Quelleninformationen häufiger verfügbar sein als bei einem anderen Abruferleben.

Einen eindrucklichen Beleg für den Zusammenhang zwischen bewusstem Abruferleben und dem Quellengedächtnis lieferten Meiser und Bröder (2002). Die Autoren konnten zeigen, dass im Falle von Remember-Antworten häufiger alle dargebotenen Quelleninformationen wiedergegeben werden konnten als das bei Know-Antworten möglich war. Aus diesen Ergebnissen schlossen die Autoren darauf, dass im Falle von Remember-Antworten komplette Episoden abgerufen wurden, in denen Inhalt und Kontext eine Einheit bildeten. Neben bei sei bemerkt, dass diese Schlussfolgerung an die Annahmen des SAC-Modells erinnert, denen zufolge die theoretische Basis für Remember-Antworten in der überschwelliger Aktivierung eines Episodenknotens beruht.

Gleichzeitig verwendeten Meiser und Bröder (2002) ein Zwei-Hochschwellen-Modell für multidimensionale Quelleninformationen zur Auswertung ihrer Daten. Allerdings blieb in diesem Modell die Trennung zwischen Remembering und Knowing unberücksichtigt. Für die beiden berücksichtigten Abruferlebensarten führten die Autoren vielmehr einzelne Parameterschätzungen durch. Für weitere Arbeiten in diesem

Bereich scheint es unerlässlich, Modelle zu entwickeln, die zusätzlich in der Lage sind, jene Prozesse adäquat zu erfassen, die dem Abruferleben zugrunde liegen. Prädestiniert für diesen Einsatz erscheint das 4-States-Modell, da es nicht nur in dem berichteten Modellvergleich gut abgeschnitten hat, sondern zusätzlich, wie oben erläutert, eine große Ähnlichkeit mit dem Modell von Meiser und Bröder (2002) aufweist.

Neben diesen spezifischen Schlussfolgerungen, die sich vor dem Hintergrund der dargelegten Ergebnisse ziehen lassen, und jenen Fragen, mit denen man sich im weiteren Verlauf der Erforschung des Abruferlebens auseinandersetzen muss, sollte ein besonders wichtiger Punkt nicht aus dem Auge verloren werden. Dieser besteht darin, dass in den vorgelegten Analysen einige Zwei-Prozess-Modelle deutlich besser abschnitten als das einfachere Ein-Prozess-Modell. Damit scheint Dunns (2004) Behauptung, dass die Ein-Prozess-Sicht ausreicht, um das Abruferleben erklären zu können, widerlegt. Diese Erkenntnis verdankt die vorliegende Arbeit dem Modellierungsansatz, durch den jene spezifischen Probleme behoben werden konnten, unter denen frühere Versuche, das Ein-Prozess-Modell zu falsifizieren, wie z. B. die Metaanalyse von Gardiner et al. (2002) noch litten.

Dass das erhaltene Resultat nicht der Methodenwahl an sich zuzuschreiben ist, belegen die Analysen der ROC-Daten, die für die gleichen Modelle positiv ausgefallen sind wie die Maximum-Likelihood-basierten Analysen. Dementsprechend kann davon ausgegangen werden, dass das bessere empirische Abschneiden der beiden Zwei-Prozess-Modelle SAC und 4-States ein robustes Phänomen ist, auf welches zukünftige Forschung aufbauen sollte.

Teil III
Anhang

Anhang A

Häufigkeitstabellen

In der folgenden Tabelle werden die absoluten Antworthäufigkeiten wiedergegeben, mit denen Probanden in den einzelnen Untersuchungen auf alte respektive neue Reize Remember (R), Know (K), Guess (G) oder Neu (N) antworteten. Die absoluten Häufigkeiten wurden nach dem in Kapitel 6 beschriebenen Vorgehen aus den in der Datenbasis enthaltenen Arbeiten ermittelt. Die Antworthäufigkeiten werden nach dem Experiment, in dem sie erhoben wurden, und nach der Bedingung, zu der sie gehörten, unterteilt berichtet. Bei zweifaktoriellen Designs sind die einzelnen Faktorstufen separat nummeriert. Die den ersten Faktor betreffende Ordnungsnummer wird durch einen Punkt von der den zweiten Faktor betreffenden Nummer getrennt. Zusätzlich wird die erwartete Auswirkung der experimentellen Manipulation (Effekt) aufgeführt, die entweder für die Sensitivität des Gedächtnisses (S) oder die Antworttendenz (A) erwartet wurde. Bei zweifaktoriellen Designs werden zwei Angaben so durch einen Punkt getrennt, dass die Zuordnung des erwarteten Effektes zu der auslösenden Manipulation direkt entnommen werden kann.

| Experiment | Effekt | Bedingung | Alt | | | | Neu | | | |
|------------|--------|-----------|------|-----|-----|-----|-----|-----|-----|------|
| | | | R | K | G | N | R | K | G | N |
| A1 | S | 1 | 115 | 29 | 8 | 8 | 10 | 36 | 76 | 198 |
| | | 2 | 24 | 32 | 38 | 66 | | | | |
| A2 | S | 1 | 91 | 173 | 38 | 178 | 67 | 106 | 96 | 691 |
| | | 2 | 84 | 131 | 44 | 221 | | | | |
| A3 | S | 1 | 203 | 126 | 63 | 58 | 26 | 36 | 72 | 766 |
| | | 2 | 243 | 95 | 49 | 63 | | | | |
| A4 | S | 1 | 29 | 104 | 75 | 368 | 69 | 104 | 138 | 841 |
| | | 2 | 40 | 92 | 58 | 386 | | | | |
| A5 | S | 1 | 46 | 144 | 52 | 334 | 35 | 69 | 69 | 979 |
| | | 2 | 52 | 127 | 29 | 368 | | | | |
| A6 | S | 1 | 127 | 161 | 69 | 219 | 69 | 81 | 115 | 887 |
| | | 2 | 109 | 104 | 63 | 300 | | | | |
| A7 | S | 1 | 242 | 81 | 63 | 190 | 35 | 46 | 69 | 1002 |
| | | 2 | 184 | 81 | 69 | 242 | | | | |
| B1 | S | 1 | 231 | 167 | 7 | 161 | 23 | 63 | 12 | 478 |
| | | 2 | 179 | 190 | 0 | 207 | | | | |
| B2 | S | 1 | 264 | 102 | 48 | 186 | 60 | 84 | 78 | 378 |
| | | 2 | 306 | 66 | 24 | 204 | | | | |
| B3 | S | 1 | 550 | 190 | 50 | 210 | 60 | 120 | 90 | 730 |
| | | 2 | 510 | 180 | 50 | 260 | | | | |
| B4 | S | 1 | 475 | 274 | 201 | 490 | 209 | 98 | 146 | 987 |
| | | 2 | 677 | 259 | 144 | 360 | | | | |
| B5 | S | 1 | 310 | 122 | 79 | 209 | 43 | 123 | 86 | 468 |
| | | 2 | 288 | 130 | 108 | 194 | | | | |
| B6 | S | 1 | 297 | 270 | 72 | 261 | 63 | 153 | 72 | 612 |
| | | 2 | 378 | 234 | 63 | 225 | | | | |
| B7 | S | 1 | 473 | 355 | 22 | 266 | 11 | 22 | 57 | 1026 |
| | | 2 | 490 | 148 | 2 | 188 | | | | |
| B8 | S | 1 | 153 | 104 | 5 | 138 | 12 | 35 | 32 | 321 |
| | | 2 | 100 | 150 | 4 | 146 | | | | |
| B9 | S | 1 | 310 | 52 | 12 | 26 | 6 | 2 | 5 | 387 |
| | | 2 | 248 | 114 | 14 | 24 | | | | |
| C1 | S | 1 | 143 | 48 | 5 | 24 | 2 | 0 | 2 | 216 |
| | | 2 | 110 | 37 | 16 | 57 | | | | |
| | | 3 | 72 | 42 | 20 | 86 | | | | |
| C2 | S | 1 | 357 | 81 | 52 | 86 | 84 | 28 | 55 | 410 |
| | | 2 | 282 | 115 | 69 | 110 | | | | |
| | | 3 | 228 | 113 | 79 | 156 | | | | |
| C3 | S | 1 | 1474 | 396 | 110 | 220 | 132 | 66 | 308 | 1694 |
| | | 2 | 792 | 506 | 506 | 396 | | | | |
| | | 3 | 550 | 616 | 572 | 462 | | | | |

| Experiment | Effekt | Bedingung | Alt | | | | Neu | | | |
|------------|--------|-----------|------|------|-----|-----|-----|----|-----|------|
| | | | R | K | G | N | R | K | G | N |
| D1 | S.S | 1.1 | 2177 | 1532 | 81 | 242 | 5 | 18 | 53 | 1688 |
| | | 1.2 | 1331 | 2137 | 121 | 443 | 35 | 53 | 88 | 1588 |
| | | 2.1 | 1371 | 2218 | 161 | 282 | 3 | 18 | 106 | 1637 |
| | | 2.2 | 897 | 2429 | 252 | 454 | 18 | 88 | 194 | 1464 |
| D2 | S.S | 1.1 | 290 | 80 | 25 | 105 | 15 | 35 | 35 | 415 |
| | | 1.2 | 180 | 105 | 20 | 195 | 5 | 20 | 10 | 465 |
| | | 2.1 | 215 | 105 | 20 | 160 | 30 | 40 | 30 | 400 |
| | | 2.2 | 145 | 110 | 20 | 225 | 0 | 30 | 20 | 450 |
| D3 | S.S | 1.1 | 62 | 57 | 84 | 67 | 11 | 32 | 89 | 138 |
| | | 1.2 | 100 | 92 | 51 | 27 | 11 | 35 | 86 | 138 |
| | | 2.1 | 57 | 89 | 62 | 62 | 11 | 38 | 73 | 148 |
| | | 2.2 | 121 | 84 | 38 | 27 | 16 | 35 | 57 | 162 |
| D4 | S.S | 1.1 | 173 | 127 | 69 | 207 | 17 | 52 | 104 | 403 |
| | | 1.2 | 150 | 121 | 69 | 236 | 12 | 63 | 58 | 443 |
| | | 2.1 | 236 | 132 | 52 | 156 | 12 | 52 | 63 | 449 |
| | | 2.2 | 213 | 127 | 52 | 184 | 17 | 69 | 64 | 426 |
| D5 | S.S | 1.1 | 517 | 376 | 24 | 259 | 24 | 18 | 3 | 123 |
| | | 1.2 | 400 | 306 | 23 | 447 | 35 | 30 | 5 | 98 |
| | | 2.1 | 306 | 294 | 12 | 564 | 67 | 45 | 7 | 49 |
| | | 2.2 | 306 | 259 | 23 | 588 | 71 | 55 | 8 | 34 |
| E1 | S.S | 1.1 | 52 | 36 | 16 | 88 | 0 | 8 | 16 | 360 |
| | | 1.2 | 44 | 38 | 10 | 100 | | | | |
| | | 2.1 | 100 | 42 | 10 | 40 | 8 | 24 | 32 | 320 |
| | | 2.2 | 68 | 40 | 16 | 68 | | | | |
| E2 | S.S | 1.1 | 79 | 46 | 8 | 59 | 12 | 38 | 12 | 322 |
| | | 1.2 | 90 | 40 | 8 | 54 | | | | |
| | | 2.1 | 100 | 44 | 10 | 38 | 12 | 34 | 20 | 318 |
| | | 2.2 | 84 | 42 | 8 | 58 | | | | |
| E3 | S.S | 1.1 | 104 | 38 | 5 | 45 | 30 | 42 | 10 | 301 |
| | | 1.2 | 85 | 48 | 2 | 57 | | | | |
| | | 2.1 | 82 | 39 | 6 | 65 | 56 | 64 | 12 | 252 |
| | | 2.2 | 71 | 47 | 4 | 70 | | | | |
| E4 | S.S | 1.1 | 123 | 86 | 36 | 15 | 29 | 43 | 58 | 590 |
| | | 1.2 | 97 | 79 | 36 | 148 | | | | |
| | | 2.1 | 234 | 79 | 14 | 33 | 58 | 58 | 50 | 554 |
| | | 2.2 | 187 | 79 | 18 | 76 | | | | |
| E5 | S.S | 1.1 | 165 | 35 | 15 | 169 | 8 | 38 | 46 | 676 |
| | | 1.2 | 108 | 38 | 38 | 200 | | | | |
| | | 2.1 | 227 | 58 | 11 | 88 | 16 | 54 | 30 | 668 |
| | | 2.2 | 123 | 81 | 23 | 157 | | | | |

| Experiment | Effekt | Bedingung | Alt | | | | Neu | | | |
|------------|--------|-----------|-----|-----|----|-----|-----|-----|----|------|
| | | | R | K | G | N | R | K | G | N |
| E6 | S.S | 1.1 | 76 | 56 | 12 | 256 | 24 | 72 | 24 | 680 |
| | | 1.2 | 32 | 56 | 16 | 296 | | | | |
| | | 2.1 | 92 | 88 | 12 | 192 | 16 | 64 | 32 | |
| | | 2.2 | 52 | 76 | 20 | 252 | | | | |
| E7 | S.S | 1.1 | 403 | 115 | 51 | 151 | 8 | 8 | 8 | 696 |
| | | 1.2 | 101 | 50 | 43 | 526 | 6 | 8 | 36 | |
| | | 2.1 | 533 | 79 | 29 | 79 | | | | |
| | | 2.2 | 122 | 72 | 94 | 432 | | | | |
| E8 | S.S | 1.1 | 154 | 51 | 22 | 93 | | | | 26 |
| | | 1.2 | 93 | 45 | 25 | 157 | | | | |
| | | 2.1 | 202 | 86 | 10 | 22 | 32 | 57 | 45 | |
| | | 2.2 | 131 | 86 | 29 | 74 | | | | |
| E9 | S.S | 1.1 | 320 | 272 | 16 | 192 | 32 | 160 | 16 | 1392 |
| | | 1.2 | 120 | 200 | 8 | 472 | | | | |
| | | 2.1 | 448 | 184 | 8 | 168 | 32 | 176 | 16 | |
| | | 2.2 | 136 | 176 | 8 | 480 | | | | |
| E10 | S.S | 1.1 | 181 | 71 | 45 | 207 | 10 | 20 | 71 | 907 |
| | | 1.2 | 126 | 56 | 50 | 207 | | | | |
| | | 2.1 | 131 | 96 | 35 | 242 | 40 | 51 | 40 | |
| | | 2.2 | 121 | 71 | 35 | 277 | | | | |
| E11 | S.S | 1.1 | 146 | 128 | 22 | 94 | 4 | 11 | 12 | 373 |
| | | 1.2 | 124 | 126 | 50 | 90 | | | | |
| | | 2.1 | 154 | 126 | 28 | 92 | | | | |
| | | 2.2 | 204 | 127 | 18 | 51 | | | | |
| E12 | S.S | 1.1 | 280 | 95 | 14 | 11 | 10 | 12 | 8 | 370 |
| | | 1.2 | 260 | 100 | 14 | 26 | | | | |
| | | 2.1 | 155 | 125 | 28 | 92 | 18 | 87 | | |
| | | 2.2 | 176 | 119 | | | | | | |
| F1 | S.S | 1.1 | 180 | 84 | 32 | 104 | 20 | 32 | 28 | 320 |
| | | 1.2 | 108 | 112 | 40 | 140 | 16 | 40 | 28 | |
| | | 1.3 | 112 | 96 | 44 | 148 | 4 | 24 | 28 | |
| | | 2.1 | 320 | 44 | 12 | 24 | 4 | 12 | 12 | |
| | | 2.2 | 200 | 116 | 36 | 48 | 4 | 28 | 20 | |
| | | 2.3 | 164 | 120 | 36 | 80 | 0 | 8 | 16 | |
| F2 | S.S | 1.1 | 173 | 35 | 11 | 69 | 1 | 7 | 10 | 274 |
| | | 1.2 | 196 | 32 | 8 | 52 | 3 | 6 | 11 | |
| | | 1.3 | 207 | 29 | 9 | 43 | 6 | 6 | 14 | |
| | | 2.1 | 130 | 49 | 17 | 92 | 1 | 13 | 12 | |
| | | 2.2 | 213 | 37 | 9 | 29 | 3 | 17 | 12 | |
| | | 2.3 | 219 | 34 | 3 | 32 | 6 | 6 | 14 | |
| G1 | A | 1.1 | 155 | 65 | 36 | 104 | 43 | 61 | 47 | 209 |
| | | 1.2 | 151 | 58 | 32 | 119 | 18 | 39 | 29 | |

| Experiment | Effekt | Bedingung | Alt | | | | Neu | | | |
|------------|--------|-----------|------|-----|-----|-----|-----|-----|-----|------|
| | | | R | K | G | N | R | K | G | N |
| G2 | A | 1.1 | 343 | 117 | 71 | 117 | 26 | 58 | 117 | 447 |
| | | 1.2 | 298 | 123 | 52 | 175 | 13 | 13 | 45 | 577 |
| H1 | A | 1.1 | 121 | 55 | 6 | 106 | 6 | 23 | 14 | 245 |
| | | 1.2 | 112 | 69 | 6 | 101 | 20 | 55 | 23 | 190 |
| | | 1.3 | 112 | 81 | 9 | 86 | 26 | 69 | 20 | 173 |
| H2 | A | 1.1 | 400 | 160 | 150 | 290 | 400 | 360 | 800 | 2440 |
| | | 1.2 | 1150 | 450 | 450 | 450 | 375 | 300 | 725 | 1100 |
| | | 1.3 | 2120 | 800 | 680 | 400 | 180 | 120 | 300 | 400 |
| H3 | A | 1.1 | 91 | 25 | 11 | 161 | 3 | 20 | 9 | 256 |
| | | 1.2 | 112 | 35 | 9 | 132 | 19 | 19 | 14 | 236 |
| | | 1.3 | 102 | 50 | 15 | 121 | 26 | 32 | 17 | 213 |
| I1 | S.A | 1.1 | 57 | 42 | 9 | 36 | 8 | 24 | 14 | 242 |
| | | 2.1 | 55 | 37 | 7 | 45 | | | | |
| | | 1.2 | 83 | 20 | 3 | 26 | 14 | 28 | 14 | 208 |
| | | 2.2 | 59 | 26 | 7 | 40 | | | | |
| I2 | S.A | 1.1 | 64 | 32 | 6 | 58 | 10 | 22 | 6 | 282 |
| | | 2.1 | 19 | 35 | 10 | 96 | | | | |
| | | 1.2 | 100 | 27 | 3 | 30 | 19 | 26 | 6 | 269 |
| | | 2.2 | 43 | 31 | 3 | 83 | | | | |
| J1 | S.A | 1.1 | 104 | 43 | 14 | 127 | 14 | 12 | 14 | 248 |
| | | 1.2 | 101 | 46 | 23 | 118 | 20 | 40 | 32 | 196 |
| | | 1.3 | 89 | 75 | 23 | 101 | 26 | 57 | 29 | 176 |
| | | 2.1 | 95 | 52 | 20 | 121 | 1 | 21 | 18 | 252 |
| | | 2.2 | 86 | 63 | 32 | 107 | 11 | 35 | 17 | 225 |
| | | 2.3 | 89 | 78 | 23 | 98 | 17 | 38 | 40 | 193 |

Anhang B

ROC-Daten

Im Folgenden werden die in Kapitel 8 näher analysierten empirischen ROC-Parameter aufgeführt. Für jede einzelne Untersuchungsbedingung ist das Krümmungsmaß φ , die Steigung der abrufenerlebensbasierten ROC-Geraden (s_R) sowie die Steigung der konfidenzbasierten ROC-Geraden (s_{RK}) aufgeführt. In der Tabelle wird zusätzlich festgehalten, in welchem Experiment (Exp.) und in welcher Bedingung (Bed.) die jeweiligen ROC-Parameter erhoben wurden.

| Exp. | Bed. | ROC-Parameter | | | Exp. | Bed. | ROC-Parameter | | |
|------|------|---------------|-------|----------|------|------|---------------|-------|----------|
| | | φ | s_R | s_{RK} | | | φ | s_R | s_{RK} |
| A1 | 1 | 4.41 | 0.68 | 0.48 | | 2.1 | 17.42 | 1.29 | 0.31 |
| | 2 | 0.16 | 0.81 | 0.80 | | 2.2 | 15.24 | 1.45 | 0.46 |
| A2 | 1 | 10.97 | 1.35 | 0.61 | D2 | 1.1 | 3.32 | 0.65 | 0.50 |
| | 2 | 7.20 | 1.16 | 0.69 | | 1.2 | 2.76 | 0.75 | 0.61 |
| A3 | 1 | 3.18 | 1.47 | 1.16 | | 2.1 | 8.75 | 0.90 | 0.46 |
| | 2 | 3.30 | 1.14 | 0.91 | | 2.2 | 1.22 | 0.42 | 0.37 |
| A4 | 1 | 5.89 | 1.36 | 0.90 | D3 | 1.1 | -0.45 | 0.83 | 0.86 |
| | 2 | 6.16 | 1.10 | 0.71 | | 1.2 | 2.58 | 0.94 | 0.78 |
| A5 | 1 | 8.46 | 1.43 | 0.79 | | 2.1 | 2.31 | 0.95 | 0.81 |
| | 2 | 12.51 | 1.17 | 0.46 | | 2.2 | 2.31 | 1.08 | 0.92 |
| A6 | 1 | 7.26 | 1.32 | 0.79 | D4 | 1.1 | 3.83 | 0.65 | 0.47 |
| | 2 | 4.74 | 1.01 | 0.72 | | 1.2 | -2.02 | 0.67 | 0.78 |
| A7 | 1 | 0.51 | 0.86 | 0.83 | | 2.1 | 2.00 | 0.66 | 0.57 |
| | 2 | 0.44 | 0.89 | 0.87 | | 2.2 | 0.76 | 0.64 | 0.61 |
| B1 | 1 | 11.17 | 1.01 | 0.42 | D5 | 1.1 | 6.76 | 2.06 | 1.22 |
| | 2 | 21.12 | 1.22 | 0.15 | | 1.2 | 8.24 | 1.19 | 0.66 |
| B2 | 1 | 2.11 | 0.68 | 0.58 | | 2.1 | 14.25 | 0.86 | 0.22 |
| | 2 | -5.76 | 0.35 | 0.60 | | 2.2 | 7.41 | 0.63 | 0.31 |
| B3 | 1 | 3.77 | 0.72 | 0.54 | E1 | 1.1 | 2.19 | 0.54 | 0.45 |
| | 2 | 0.81 | 0.58 | 0.54 | | 1.2 | 5.95 | 0.52 | 0.28 |
| B4 | 1 | 3.43 | 1.48 | 1.16 | | 2.1 | 6.57 | 0.71 | 0.41 |
| | 2 | 5.39 | 1.30 | 0.89 | | 2.2 | 3.71 | 0.70 | 0.52 |
| B5 | 1 | -4.28 | 0.62 | 0.85 | E2 | 1.1 | -0.18 | 0.83 | 0.84 |
| | 2 | -1.21 | 0.94 | 1.02 | | 1.2 | -2.09 | 0.75 | 0.87 |
| B6 | 1 | 0.85 | 0.99 | 0.93 | | 2.1 | 1.86 | 0.87 | 0.76 |
| | 2 | 3.17 | 0.92 | 0.74 | | 2.2 | 4.77 | 0.74 | 0.51 |
| B7 | 1 | 18.42 | 0.97 | 0.13 | E3 | 1.1 | 1.63 | 0.99 | 0.89 |
| | 2 | 15.16 | 0.61 | 0.02 | | 1.2 | 14.78 | 1.09 | 0.32 |
| B8 | 1 | 14.20 | 0.68 | 0.10 | | 2.1 | -0.77 | 0.92 | 0.97 |
| | 2 | 19.41 | 1.05 | 0.13 | | 2.2 | 6.86 | 1.04 | 0.64 |
| B9 | 1 | 11.26 | 2.34 | 0.98 | E4 | 1.1 | 0.10 | 1.97 | 1.95 |
| | 2 | 15.97 | 1.96 | 0.60 | | 1.2 | 5.37 | 1.01 | 0.69 |
| C1 | 1 | 20.93 | 2.40 | 0.48 | | 2.1 | 7.81 | 1.42 | 0.82 |
| | 2 | 8.22 | 1.06 | 0.59 | | 2.2 | 7.93 | 1.13 | 0.64 |
| | 3 | 9.92 | 2.06 | 0.97 | | E5 | 1.1 | 0.79 | 0.29 |
| C2 | 1 | 4.29 | 1.47 | 1.08 | 1.2 | | -4.37 | 0.46 | 0.67 |
| | 2 | 3.68 | 1.27 | 0.98 | 2.1 | 2.79 | 0.56 | 0.44 | |
| | 3 | 6.11 | 1.43 | 0.93 | 2.2 | 0.64 | 0.77 | 0.73 | |
| C3 | 1 | 11.87 | 1.03 | 0.41 | E6 | 1.1 | 0.57 | 0.62 | 0.59 |
| | 2 | 2.18 | 0.83 | 0.71 | | 1.2 | -0.43 | 0.90 | 0.93 |
| | 3 | 2.56 | 0.90 | 0.75 | | 2.1 | 7.36 | 0.73 | 0.39 |
| D1 | 1.1 | 18.86 | 1.38 | 0.30 | | 2.2 | 2.60 | 0.82 | 0.68 |
| | 1.2 | 21.57 | 2.15 | 0.40 | | E7 | 1.1 | 1.66 | 1.45 |

| Exp. | Bed. | ROC-Parameter | | | Exp. | Bed. | ROC-Parameter | | |
|------|------|---------------|-------|----------|------|-------|---------------|-------|----------|
| | | φ | s_R | s_{RK} | | | φ | s_R | s_{RK} |
| E8 | 1.2 | -0.94 | 1.03 | 1.10 | G1 | 1.3 | 7.80 | 0.66 | 0.32 |
| | 2.1 | 7.22 | 0.64 | 0.33 | | 2.1 | -1.55 | 0.44 | 0.50 |
| | 2.2 | 2.95 | 0.77 | 0.62 | | 2.2 | -0.75 | 0.58 | 0.62 |
| | 1.1 | -5.55 | 0.90 | 1.33 | | 2.3 | 14.30 | 0.74 | 0.14 |
| | 1.2 | -6.49 | 0.86 | 1.37 | | 1.1 | -0.47 | 0.75 | 0.78 |
| E9 | 2.1 | 8.78 | 1.38 | 0.74 | 1.2 | -2.21 | 0.69 | 0.81 | |
| | 2.2 | 2.22 | 1.15 | 0.99 | G2 | 1.1 | 8.61 | 0.87 | 0.44 |
| | 1.1 | -3.18 | 1.03 | 1.30 | 1.2 | 2.08 | 0.67 | 0.57 | |
| | 1.2 | 6.57 | 0.87 | 0.53 | H1 | 1.1 | 7.71 | 0.54 | 0.23 |
| E10 | 2.1 | -0.84 | 0.70 | 0.74 | 1.2 | 9.18 | 0.62 | 0.24 | |
| | 2.2 | 3.17 | 0.72 | 0.56 | 1.3 | 5.67 | 0.75 | 0.48 | |
| | 1.1 | 4.36 | 0.56 | 0.37 | H2 | 1.1 | 2.46 | 0.80 | 0.67 |
| | 1.2 | 2.85 | 0.60 | 0.48 | 1.2 | 2.22 | 0.86 | 0.73 | |
| | 2.1 | 4.23 | 1.10 | 0.82 | 1.3 | 2.61 | 1.03 | 0.86 | |
| E11 | 2.2 | 1.40 | 0.93 | 0.84 | H3 | 1.1 | -5.46 | 0.30 | 0.53 |
| | 1.1 | 9.99 | 1.23 | 0.60 | 1.2 | 6.03 | 0.65 | 0.38 | |
| | 1.2 | 1.40 | 1.45 | 1.31 | 1.3 | 2.72 | 0.83 | 0.68 | |
| | 2.1 | 7.12 | 1.24 | 0.75 | I1 | 1.1 | 3.04 | 1.02 | 0.83 |
| E12 | 2.2 | 9.54 | 1.34 | 0.68 | 2.1 | 5.01 | 0.86 | 0.59 | |
| | 1.1 | 0.97 | 2.68 | 2.43 | 1.2 | 5.48 | 0.64 | 0.40 | |
| | 1.2 | 4.78 | 2.17 | 1.47 | 2.2 | 1.04 | 0.79 | 0.74 | |
| | 2.1 | 4.79 | 1.97 | 1.35 | I2 | 1.1 | -1.43 | 0.89 | 0.98 |
| | 2.2 | 9.19 | 1.79 | 0.91 | 2.1 | -2.57 | 1.36 | 1.65 | |
| F1 | 1.1 | 2.36 | 0.96 | 0.81 | 1.2 | 2.62 | 1.01 | 0.84 | |
| | 1.2 | 1.56 | 1.06 | 0.95 | 2.2 | 7.49 | 1.01 | 0.59 | |
| | 1.3 | 0.41 | 0.73 | 0.71 | J1 | 1.1 | 7.79 | 0.88 | 0.48 |
| | 2.1 | 1.04 | 0.84 | 0.78 | 1.2 | 0.33 | 0.61 | 0.59 | |
| F2 | 2.2 | -1.33 | 0.66 | 0.73 | 1.3 | 1.41 | 0.83 | 0.75 | |
| | 2.3 | 3.69 | 0.82 | 0.62 | 2.1 | -2.67 | 0.40 | 0.51 | |
| | 1.1 | 1.98 | 0.39 | 0.31 | 2.2 | -5.84 | 0.86 | 1.30 | |
| | 1.2 | 6.66 | 0.53 | 0.26 | 2.3 | 6.60 | 0.81 | 0.48 | |

Anhang C

Parameterschätzwerte

In diesem Abschnitt werden für jedes Modell die Parameterwerte wiedergegeben, die anhand der in den einzelnen experimentellen Bedingungen beobachteten Daten geschätzt wurden. In Klammern stehen jene Parametersätze, die wegen mangelnder Datenanpassungsgüte von den in Kapitel 7 beschriebenen Analysen ausgeschlossen wurden. Für alle Parametervektoren wird festgehalten, aus welchem Experiment und welcher Bedingung die Daten stammten, auf denen die Schätzungen basierten. Bei zweifaktoriellen Designs sind die einzelnen Faktorstufen separat nummeriert. Die den ersten Faktor betreffende Ordnungsnummer wird durch einen Punkt von der den zweiten Faktor betreffenden Nummer getrennt. Zusätzlich wird die erwartete Auswirkung der experimentellen Manipulation (Effekt) aufgeführt, die entweder für die Sensitivität des Gedächtnisses (S) oder die Antworttendenz (A) erwartet wurde. Bei zweifaktoriellen Designs werden zwei Angaben so durch einen Punkt getrennt, dass die Zuordnung des erwarteten Effektes zu der auslösenden Manipulation direkt entnommen werden kann.

| Experiment | Effekt | Bedingung | Parameter | | | | |
|------------|--------|-----------|-----------|-------|-------|-------|----------------|
| | | | d' | C_R | C_K | C_G | σ_{Alt} |
| SDM-Modell | | | | | | | |
| A1 | S | 1 | 2.74 | 1.91 | 1.03 | 0.31 | 1.42 |
| | | 2 | 0.58 | 1.87 | 1.06 | 0.30 | 1.24 |
| A2 | S | 1 | (0.87 | 1.53 | 0.84 | 0.59 | 0.78) |
| | | 2 | (0.69 | 1.51 | 0.87 | 0.59 | 0.90) |
| A3 | S | 1 | 1.85 | 1.94 | 1.44 | 1.04 | 0.70 |
| | | 2 | 2.03 | 1.93 | 1.45 | 1.04 | 0.90 |
| A4 | S | 1 | (0.34 | 1.59 | 0.99 | 0.62 | 0.80) |
| | | 2 | (0.20 | 1.59 | 0.99 | 0.62 | 0.98) |
| A5 | S | 1 | (0.90 | 1.93 | 1.27 | 1.04 | 0.76) |
| | | 2 | (0.71 | 1.95 | 1.26 | 1.04 | 0.95) |
| A6 | S | 1 | (1.00 | 1.60 | 1.06 | 0.74 | 0.80) |
| | | 2 | (0.70 | 1.58 | 1.09 | 0.74 | 1.03) |
| A7 | S | 1 | 1.64 | 1.88 | 1.47 | 1.13 | 1.17 |
| | | 2 | 1.35 | 1.88 | 1.47 | 1.13 | 1.12 |
| B1 | S | 1 | 1.54 | 1.77 | 1.01 | 0.96 | 1.01 |
| | | 2 | (0.89 | 1.29 | 0.60 | 0.59 | 0.83) |
| B2 | S | 1 | 1.07 | 1.29 | 0.69 | 0.33 | 1.48 |
| | | 2 | 2.31 | 2.25 | 1.45 | 1.22 | 2.66 |
| B3 | S | 1 | (1.76 | 1.58 | 0.89 | 0.62 | 1.40) |
| | | 2 | 1.93 | 1.89 | 1.07 | 0.81 | 1.75 |
| B4 | S | 1 | (0.77 | 1.07 | 0.76 | 0.49 | 0.69) |
| | | 2 | (0.97 | 1.02 | 0.70 | 0.43 | 0.78) |
| B5 | S | 1 | (1.23 | 1.52 | 0.78 | 0.38 | 1.58) |
| | | 2 | 1.20 | 1.47 | 0.97 | 0.55 | 1.06 |
| B6 | S | 1 | 1.03 | 1.48 | 0.70 | 0.47 | 1.02 |
| | | 2 | 1.08 | 1.29 | 0.59 | 0.34 | 1.09 |
| B7 | S | 1 | (2.46 | 2.72 | 1.63 | 1.41 | 1.43) |
| | | 2 | (2.93 | 2.44 | 1.55 | 1.35 | 2.06) |
| B8 | S | 1 | (1.55 | 2.02 | 1.06 | 0.86 | 1.64) |
| | | 2 | (1.47 | 2.18 | 1.18 | 1.09 | 1.07) |
| B9 | S | 1 | 2.56 | 2.21 | 1.96 | 1.85 | 0.47 |
| | | 2 | 3.28 | 3.06 | 2.37 | 2.17 | 0.70 |
| C1 | S | 1 | (2.47 | 2.28 | 1.94 | 1.87 | 0.49) |
| | | 2 | 2.73 | 2.73 | 2.26 | 2.01 | 1.11 |
| | | 3 | 2.02 | 2.26 | 2.00 | 1.86 | 0.55 |
| C2 | S | 1 | (1.28 | 1.07 | 0.82 | 0.56 | 0.68) |
| | | 2 | (0.99 | 1.00 | 0.63 | 0.29 | 0.79) |
| | | 3 | (1.11 | 1.29 | 0.97 | 0.65 | 0.73) |
| C3 | S | 1 | (2.12 | 1.66 | 1.18 | 0.75 | 1.02) |
| | | 2 | (1.03 | 1.45 | 0.80 | -0.12 | 1.22) |
| | | 3 | (0.59 | 1.33 | 0.56 | -0.34 | 1.13) |

| Experiment | Effekt | Bedingung | Parameter | | | | |
|------------|--------|-----------|-----------|-------|-------|-------|----------------|
| | | | d' | C_R | C_K | C_G | σ_{Alt} |
| D1 | S.S | 1.1 | (3.61 | 3.49 | 1.97 | 1.73 | 1.19) |
| | | 1.2 | (2.02 | 2.28 | 1.42 | 1.30 | 0.58) |
| | | 2.1 | (3.93 | 4.60 | 1.97 | 1.48 | 1.63) |
| | | 2.2 | (2.18 | 2.91 | 1.33 | 0.98 | 0.96) |
| D2 | S.S | 1.1 | 2.23 | 1.91 | 1.25 | 0.96 | 1.57 |
| | | 1.2 | 1.86 | 2.35 | 1.63 | 1.48 | 1.37 |
| | | 2.1 | (1.39 | 1.59 | 1.02 | 0.85 | 1.15) |
| | | 2.2 | 1.57 | 2.94 | 1.52 | 1.25 | 2.49 |
| D3 | S.S | 1.1 | 0.84 | 1.73 | 1.01 | 0.03 | 1.20 |
| | | 1.2 | 1.47 | 1.82 | 0.91 | 0.04 | 1.09 |
| | | 2.1 | 0.95 | 1.80 | 0.87 | 0.13 | 1.08 |
| | | 2.2 | 1.48 | 1.60 | 0.85 | 0.26 | 0.94 |
| D4 | S.S | 1.1 | (1.14 | 1.98 | 1.11 | 0.53 | 1.64) |
| | | 1.2 | 1.06 | 2.00 | 1.15 | 0.73 | 1.45 |
| | | 2.1 | 1.73 | 2.07 | 1.20 | 0.77 | 1.55 |
| | | 2.2 | 1.38 | 1.90 | 1.03 | 0.64 | 1.57 |
| D5 | S.S | 1.1 | 1.00 | 1.07 | 0.66 | 0.62 | 0.49 |
| | | 1.2 | 0.47 | 0.82 | 0.26 | 0.21 | 0.84 |
| | | 2.1 | (-0.47 | 0.27 | -0.49 | -0.53 | 1.15) |
| | | 2.2 | -0.81 | 0.20 | -0.72 | -0.81 | 1.58 |
| E1 | S.S | 1.1 | 1.69 | 2.89 | 1.92 | 1.48 | 1.98 |
| | | 1.2 | 1.37 | 3.08 | 1.87 | 1.48 | 2.33 |
| | | 2.1 | 2.19 | 2.11 | 1.33 | 0.97 | 1.47 |
| | | 2.2 | 1.54 | 2.08 | 1.35 | 0.97 | 1.48 |
| E2 | S.S | 1.1 | 1.59 | 1.86 | 1.13 | 0.99 | 1.20 |
| | | 1.2 | 1.75 | 1.86 | 1.13 | 0.99 | 1.32 |
| | | 2.1 | 1.93 | 1.87 | 1.17 | 0.95 | 1.16 |
| | | 2.2 | 1.67 | 1.88 | 1.15 | 0.95 | 1.38 |
| E3 | S.S | 1.1 | 1.52 | 1.42 | 0.88 | 0.79 | 1.01 |
| | | 1.2 | 1.30 | 1.43 | 0.86 | 0.79 | 0.93 |
| | | 2.1 | 0.85 | 1.05 | 0.49 | 0.40 | 1.09 |
| | | 2.2 | 0.74 | 1.06 | 0.48 | 0.40 | 0.96 |
| E4 | S.S | 1.1 | 1.71 | 1.75 | 1.28 | 0.91 | 0.51 |
| | | 1.2 | (1.16 | 1.78 | 1.23 | 0.92 | 1.04) |
| | | 2.1 | (1.69 | 1.42 | 0.96 | 0.74 | 0.69) |
| | | 2.2 | (1.47 | 1.42 | 0.95 | 0.74 | 0.89) |
| E5 | S.S | 1.1 | 1.71 | 2.33 | 1.55 | 1.18 | 3.51 |
| | | 1.2 | 1.06 | 2.23 | 1.60 | 1.17 | 2.00 |
| | | 2.1 | 2.47 | 2.05 | 1.32 | 1.13 | 1.80 |
| | | 2.2 | 1.43 | 2.04 | 1.33 | 1.13 | 1.31 |

| Experiment | Effekt | Bedingung | Parameter | | | | |
|------------|--------|-----------|-----------|-------|-------|-------|----------------|
| | | | d' | C_R | C_K | C_G | σ_{Alt} |
| E6 | S.S | 1.1 | 0.45 | 1.88 | 1.17 | 1.04 | 1.63 |
| | | 1.2 | 0.32 | 1.88 | 1.18 | 1.04 | 1.11 |
| | | 2.1 | (1.09 | 2.09 | 1.25 | 1.08 | 1.43) |
| | | 2.2 | 0.67 | 2.07 | 1.27 | 1.08 | 1.25 |
| E7 | S.S | 1.1 | 2.40 | 2.30 | 1.99 | 1.83 | 0.70 |
| | | 1.2 | 1.24 | 2.28 | 2.02 | 1.83 | 0.96 |
| | | 2.1 | (3.78 | 2.58 | 1.93 | 1.48 | 1.85) |
| | | 2.2 | 1.13 | 2.47 | 2.01 | 1.48 | 1.41 |
| E8 | S.S | 1.1 | 1.67 | 1.73 | 1.25 | 1.08 | 1.09 |
| | | 1.2 | 1.10 | 1.72 | 1.26 | 1.08 | 1.12 |
| | | 2.1 | (1.92 | 1.67 | 1.05 | 0.81 | 0.71) |
| | | 2.2 | 1.46 | 1.66 | 1.07 | 0.81 | 0.87 |
| E9 | S.S | 1.1 | 1.80 | 2.05 | 1.18 | 1.13 | 0.96 |
| | | 1.2 | 0.86 | 2.06 | 1.17 | 1.13 | 1.16 |
| | | 2.1 | 2.25 | 2.05 | 1.13 | 1.08 | 1.44 |
| | | 2.2 | 0.73 | 2.06 | 1.12 | 1.08 | 1.40 |
| E10 | S.S | 1.1 | (1.74 | 2.44 | 1.81 | 1.28 | 1.98) |
| | | 1.2 | 1.42 | 2.40 | 1.84 | 1.28 | 1.77 |
| | | 2.1 | 1.18 | 1.77 | 1.31 | 1.13 | 0.93 |
| | | 2.2 | 0.99 | 1.76 | 1.33 | 1.13 | 1.09 |
| E11 | S.S | 1.1 | (2.16 | 2.45 | 1.69 | 1.50 | 0.92) |
| | | 1.2 | 2.02 | 2.35 | 1.76 | 1.50 | 0.70 |
| | | 2.1 | 2.16 | 2.42 | 1.71 | 1.50 | 0.89 |
| | | 2.2 | 2.46 | 2.44 | 1.69 | 1.50 | 0.83 |
| E12 | S.S | 1.1 | 2.16 | 1.97 | 1.59 | 1.44 | 0.37 |
| | | 1.2 | 2.16 | 1.98 | 1.56 | 1.44 | 0.47 |
| | | 2.1 | 1.83 | 1.98 | 1.56 | 1.44 | 0.53 |
| | | 2.2 | 1.91 | 1.99 | 1.54 | 1.44 | 0.59 |
| F1 | S.S | 1.1 | 1.53 | 1.66 | 1.10 | 0.84 | 1.06 |
| | | 1.2 | 1.18 | 1.76 | 1.07 | 0.81 | 0.96 |
| | | 1.3 | 1.54 | 2.34 | 1.47 | 1.08 | 1.37 |
| | | 2.1 | 3.35 | 2.34 | 1.74 | 1.48 | 1.20 |
| | | 2.2 | 1.73 | 1.74 | 0.52 | -0.01 | 1.49 |
| | | 2.3 | 2.67 | 2.98 | 1.92 | 1.50 | 1.38 |
| F2 | S.S | 1.1 | 3.49 | 2.78 | 1.89 | 1.54 | 2.74 |
| | | 1.2 | 3.42 | 2.43 | 1.77 | 1.48 | 2.10 |
| | | 1.3 | (3.08 | 2.12 | 1.63 | 1.34 | 1.65) |
| | | 2.1 | 2.38 | 2.65 | 1.68 | 1.35 | 2.20 |
| | | 2.2 | 3.40 | 2.30 | 1.49 | 1.22 | 1.70 |
| | | 2.3 | (3.25 | 2.16 | 1.58 | 1.35 | 1.53) |
| G1 | A | 1.1 | 0.94 | 1.18 | 0.56 | 0.20 | 1.33 |
| | | 1.2 | 1.33 | 1.63 | 1.02 | 0.71 | 1.44 |

| Experiment | Effekt | Bedingung | Parameter | | | | |
|------------|--------|-----------|-----------|-------|-------|-------|----------------|
| | | | d' | C_R | C_K | C_G | σ_{Att} |
| G2 | A | 1.1 | (2.07 | 2.19 | 1.61 | 1.24 | 1.33) |
| | | 1.2 | 1.90 | 1.79 | 1.10 | 0.50 | 1.52 |
| H1 | A | 1.1 | 1.72 | 2.11 | 1.22 | 1.04 | 1.97 |
| | | 1.2 | (1.06 | 1.52 | 0.59 | 0.42 | 1.64) |
| | | 1.3 | 0.98 | 1.35 | 0.42 | 0.26 | 1.35 |
| H2 | A | 1.1 | (0.99 | 1.29 | 0.86 | 0.28 | 1.25) |
| | | 1.2 | (0.95 | 1.06 | 0.57 | -0.14 | 1.17) |
| | | 1.3 | (1.03 | 0.95 | 0.45 | -0.24 | 0.98) |
| H3 | A | 1.1 | 0.76 | 2.24 | 1.45 | 1.22 | 3.07 |
| | | 1.2 | 1.08 | 1.52 | 1.08 | 0.92 | 1.58 |
| | | 1.3 | 0.89 | 1.35 | 0.82 | 0.64 | 1.22 |
| I1 | S.A | 1.1 | 1.67 | 1.93 | 1.20 | 1.00 | 0.99 |
| | | 2.1 | 1.59 | 1.94 | 1.19 | 1.00 | 1.19 |
| | | 1.2 | 2.15 | 1.63 | 0.98 | 0.80 | 1.56 |
| | | 2.2 | 1.45 | 1.62 | 0.99 | 0.80 | 1.26 |
| I2 | S.A | 1.1 | 1.57 | 1.86 | 1.29 | 1.18 | 1.12 |
| | | 2.1 | 1.00 | 1.85 | 1.29 | 1.18 | 0.72 |
| | | 1.2 | 1.88 | 1.56 | 1.07 | 1.00 | 0.99 |
| | | 2.2 | 0.95 | 1.57 | 1.06 | 1.00 | 1.00 |
| J1 | S.A | 1.1 | 1.27 | 1.69 | 1.27 | 1.09 | 1.20 |
| | | 1.2 | 0.85 | 1.48 | 0.81 | 0.47 | 1.66 |
| | | 1.3 | 0.75 | 1.35 | 0.55 | 0.28 | 1.20 |
| | | 2.1 | 1.56 | 2.59 | 1.47 | 1.09 | 2.33 |
| | | 2.2 | 1.13 | 1.72 | 1.05 | 0.77 | 1.11 |
| | | 2.3 | (1.00 | 1.62 | 0.81 | 0.45 | 1.29) |

| Experiment | Effekt | Bedingung | Parameter | | | | |
|-------------|--------|-----------|-----------|------|-------|-------|--------|
| | | | r | d' | C_R | C_K | C_G |
| HTSD-Modell | | | | | | | |
| A1 | S | 1 | .47 | 1.80 | 1.89 | 1.02 | 0.31 |
| | | 2 | .08 | 0.47 | 1.87 | 1.04 | 0.31 |
| A2 | S | 1 | (.00 | 0.86 | 1.63 | 0.85 | 0.57) |
| | | 2 | (.00 | 0.67 | 1.55 | 0.87 | 0.58) |
| A3 | S | 1 | (.00 | 2.06 | 2.13 | 1.49 | 1.02) |
| | | 2 | .00 | 2.10 | 1.98 | 1.46 | 1.04 |
| A4 | S | 1 | (.00 | 0.22 | 1.66 | 1.01 | 0.60) |
| | | 2 | (.00 | 0.18 | 1.60 | 0.99 | 0.62) |
| A5 | S | 1 | (.00 | 0.79 | 2.06 | 1.29 | 1.02) |
| | | 2 | (.00 | 0.69 | 1.97 | 1.26 | 1.04) |
| A6 | S | 1 | (.00 | 1.00 | 1.69 | 1.07 | 0.72) |
| | | 2 | (.00 | 0.70 | 1.57 | 1.08 | 0.74) |
| A7 | S | 1 | .14 | 1.43 | 1.88 | 1.46 | 1.13 |
| | | 2 | .08 | 1.24 | 1.88 | 1.46 | 1.13 |
| B1 | S | 1 | .00 | 1.53 | 1.77 | 1.01 | 0.96 |
| | | 2 | (.00 | 0.90 | 1.36 | 0.58 | 0.56) |
| B2 | S | 1 | .27 | 0.55 | 1.28 | 0.67 | 0.34 |
| | | 2 | .47 | 0.85 | 2.32 | 1.42 | 1.22 |
| B3 | S | 1 | (.32 | 1.14 | 1.56 | 0.88 | 0.62) |
| | | 2 | .40 | 1.00 | 1.89 | 1.05 | 0.81 |
| B4 | S | 1 | (.00 | 0.80 | 1.18 | 0.78 | 0.44) |
| | | 2 | (.00 | 1.05 | 1.08 | 0.71 | 0.40) |
| B5 | S | 1 | .33 | 0.54 | 1.55 | 0.76 | 0.38 |
| | | 2 | .08 | 1.09 | 1.48 | 0.97 | 0.55 |
| B6 | S | 1 | .01 | 1.02 | 1.48 | 0.70 | 0.47 |
| | | 2 | .07 | 0.97 | 1.29 | 0.59 | 0.34 |
| B7 | S | 1 | (.11 | 2.10 | 2.48 | 1.60 | 1.43) |
| | | 2 | (.44 | 1.66 | 2.27 | 1.52 | 1.36) |
| B8 | S | 1 | (.23 | 1.06 | 1.91 | 1.05 | 0.88) |
| | | 2 | (.00 | 1.46 | 2.13 | 1.18 | 1.09) |
| B9 | S | 1 | (.00 | 3.28 | 2.51 | 2.02 | 1.81) |
| | | 2 | (.00 | 3.71 | 3.40 | 2.43 | 2.16) |
| C1 | S | 1 | (.00 | 3.03 | 2.63 | 1.97 | 1.83) |
| | | 2 | .04 | 2.63 | 2.68 | 2.24 | 2.01 |
| | | 3 | .00 | 2.10 | 2.53 | 2.07 | 1.84 |
| C2 | S | 1 | (.00 | 1.50 | 1.16 | 0.84 | 0.52) |
| | | 2 | (.00 | 1.09 | 1.07 | 0.64 | 0.26) |
| | | 3 | (.00 | 1.19 | 1.41 | 1.01 | 0.62) |
| C3 | S | 1 | (.00 | 2.10 | 1.65 | 1.18 | 0.75) |
| | | 2 | (.13 | 0.78 | 1.42 | 0.77 | -0.10) |
| | | 3 | (.04 | 0.52 | 1.29 | 0.55 | -0.32) |

| Experiment | Effekt | Bedingung | Parameter | | | | |
|------------|--------|-----------|-----------|-------|-------|-------|--------|
| | | | r | d' | C_R | C_K | C_G |
| D1 | S.S | 1.1 | (.00 | 3.32 | 3.23 | 1.94 | 1.74) |
| | | 1.2 | (.00 | 2.46 | 2.88 | 1.44 | 1.25) |
| | | 2.1 | (.06 | 3.01 | 3.54 | 1.84 | 1.50) |
| | | 2.2 | (.00 | 2.23 | 2.97 | 1.33 | 0.97) |
| D2 | S.S | 1.1 | .40 | 1.36 | 1.90 | 1.23 | 0.96 |
| | | 1.2 | .19 | 1.53 | 2.34 | 1.62 | 1.48 |
| | | 2.1 | (.08 | 1.26 | 1.57 | 1.02 | 0.85) |
| | | 2.2 | (.27 | 0.99 | 2.92 | 1.46 | 1.27) |
| D3 | S.S | 1.1 | .11 | 0.63 | 1.75 | 0.98 | 0.04 |
| | | 1.2 | .03 | 1.38 | 1.77 | 0.89 | 0.05 |
| | | 2.1 | .02 | 0.90 | 1.76 | 0.86 | 0.14 |
| | | 2.2 | .00 | 1.52 | 1.64 | 0.85 | 0.25 |
| D4 | S.S | 1.1 | (.21 | 0.72 | 1.91 | 1.06 | 0.55) |
| | | 1.2 | .18 | 0.74 | 2.04 | 1.12 | 0.74 |
| | | 2.1 | (.29 | 1.12 | 2.06 | 1.16 | 0.78) |
| | | 2.2 | .26 | 0.84 | 1.90 | 1.00 | 0.65 |
| D5 | S.S | 1.1 | (.00 | 1.25 | 1.38 | 0.57 | 0.50) |
| | | 1.2 | .00 | 0.48 | 0.88 | 0.23 | 0.18 |
| | | 2.1 | (.06 | -0.54 | 0.25 | -0.48 | -0.52) |
| | | 2.2 | (.17 | -1.05 | 0.19 | -0.72 | -0.79) |
| E1 | S.S | 1.1 | .23 | 1.28 | 2.86 | 1.85 | 1.49 |
| | | 1.2 | (.20 | 1.16 | 2.88 | 1.79 | 1.50) |
| | | 2.1 | (.31 | 1.55 | 2.07 | 1.31 | 0.98) |
| | | 2.2 | .21 | 1.13 | 2.05 | 1.33 | 0.98 |
| E2 | S.S | 1.1 | .16 | 1.33 | 1.86 | 1.12 | 0.99 |
| | | 1.2 | .26 | 1.30 | 1.86 | 1.13 | 0.99 |
| | | 2.1 | .17 | 1.67 | 1.87 | 1.16 | 0.95 |
| | | 2.2 | .24 | 1.24 | 1.87 | 1.15 | 0.95 |
| E3 | S.S | 1.1 | .00 | 1.52 | 1.42 | 0.88 | 0.79 |
| | | 1.2 | .00 | 1.32 | 1.45 | 0.86 | 0.79 |
| | | 2.1 | .08 | 0.73 | 1.05 | 0.49 | 0.40 |
| | | 2.2 | .00 | 0.74 | 1.07 | 0.48 | 0.40 |
| E4 | S.S | 1.1 | (.00 | 2.08 | 2.02 | 1.31 | 0.87) |
| | | 1.2 | (.00 | 1.16 | 1.77 | 1.23 | 0.92) |
| | | 2.1 | (.00 | 1.95 | 1.51 | 0.96 | 0.72) |
| | | 2.2 | (.00 | 1.52 | 1.45 | 0.95 | 0.73) |
| E5 | S.S | 1.1 | (.41 | 0.56 | 2.32 | 1.51 | 1.18) |
| | | 1.2 | .24 | 0.68 | 2.31 | 1.57 | 1.17 |
| | | 2.1 | .46 | 1.33 | 2.04 | 1.31 | 1.13 |
| | | 2.2 | .16 | 1.17 | 2.04 | 1.32 | 1.13 |

| Experiment | Effekt | Bedingung | Parameter | | | | |
|------------|--------|-----------|-----------|------|-------|-------|-------|
| | | | r | d' | C_R | C_K | C_G |
| E6 | S.S | 1.1 | .13 | 0.42 | 1.88 | 1.17 | 1.04 |
| | | 1.2 | .02 | 0.35 | 1.88 | 1.17 | 1.04 |
| | | 2.1 | (.13 | 0.92 | 2.06 | 1.24 | 1.09) |
| | | 2.2 | .05 | 0.67 | 2.06 | 1.26 | 1.08 |
| E7 | S.S | 1.1 | .00 | 2.61 | 2.45 | 2.03 | 1.82 |
| | | 1.2 | .00 | 1.21 | 2.30 | 2.02 | 1.83 |
| | | 2.1 | (.59 | 2.17 | 2.49 | 1.88 | 1.49) |
| | | 2.2 | .09 | 1.09 | 2.43 | 1.96 | 1.49 |
| E8 | S.S | 1.1 | .13 | 1.49 | 1.74 | 1.25 | 1.07 |
| | | 1.2 | .09 | 0.96 | 1.74 | 1.25 | 1.07 |
| | | 2.1 | (.00 | 2.18 | 1.79 | 1.05 | 0.79) |
| | | 2.2 | .00 | 1.51 | 1.71 | 1.07 | 0.80 |
| E9 | S.S | 1.1 | .00 | 1.82 | 2.07 | 1.18 | 1.12 |
| | | 1.2 | .04 | 0.83 | 2.06 | 1.17 | 1.13 |
| | | 2.1 | .36 | 1.53 | 2.05 | 1.13 | 1.08 |
| | | 2.2 | .10 | 0.66 | 2.05 | 1.12 | 1.08 |
| E10 | S.S | 1.1 | (.28 | 1.16 | 2.38 | 1.75 | 1.30) |
| | | 1.2 | (.21 | 1.07 | 2.36 | 1.79 | 1.29) |
| | | 2.1 | .00 | 1.17 | 1.79 | 1.32 | 1.12 |
| | | 2.2 | .04 | 0.95 | 1.76 | 1.33 | 1.13 |
| E11 | S.S | 1.1 | (.00 | 2.20 | 2.51 | 1.69 | 1.49) |
| | | 1.2 | .00 | 2.18 | 2.63 | 1.82 | 1.47 |
| | | 2.1 | .00 | 2.23 | 2.51 | 1.72 | 1.49 |
| | | 2.2 | (.00 | 2.62 | 2.59 | 1.71 | 1.49) |
| E12 | S.S | 1.1 | (.00 | 3.03 | 2.47 | 1.62 | 1.37) |
| | | 1.2 | (.00 | 2.79 | 2.37 | 1.58 | 1.39) |
| | | 2.1 | (.00 | 2.08 | 2.33 | 1.59 | 1.39) |
| | | 2.2 | (.00 | 2.15 | 2.27 | 1.55 | 1.40) |
| F1 | S.S | 1.1 | .04 | 1.46 | 1.65 | 1.10 | 0.85 |
| | | 1.2 | .00 | 1.18 | 1.79 | 1.07 | 0.80 |
| | | 1.3 | .17 | 1.23 | 2.34 | 1.44 | 1.09 |
| | | 2.1 | .40 | 2.76 | 2.34 | 1.74 | 1.48 |
| | | 2.2 | .39 | 0.88 | 1.77 | 0.48 | 0.01 |
| | | 2.3 | .24 | 2.16 | 2.91 | 1.87 | 1.51 |
| F2 | S.S | 1.1 | .56 | 1.47 | 2.76 | 1.83 | 1.55 |
| | | 1.2 | (.57 | 1.73 | 2.36 | 1.73 | 1.49) |
| | | 1.3 | (.51 | 1.89 | 2.08 | 1.61 | 1.35) |
| | | 2.1 | .41 | 1.26 | 2.73 | 1.63 | 1.35 |
| | | 2.2 | .61 | 1.87 | 2.32 | 1.47 | 1.22 |
| | | 2.3 | (.47 | 2.21 | 2.10 | 1.56 | 1.35) |
| G1 | A | 1.1 | .23 | 0.52 | 1.18 | 0.55 | 0.21 |
| | | 1.2 | .27 | 0.83 | 1.64 | 1.01 | 0.71 |

| Experiment | Effekt | Bedingung | Parameter | | | | |
|------------|--------|-----------|-----------|------|-------|-------|--------|
| | | | r | d' | C_R | C_K | C_G |
| G2 | A | 1.1 | (.17 | 1.73 | 2.11 | 1.58 | 1.25) |
| | | 1.2 | (.36 | 1.12 | 1.77 | 1.06 | 0.51) |
| H1 | A | 1.1 | (.33 | 0.96 | 2.06 | 1.20 | 1.05) |
| | | 1.2 | (.27 | 0.52 | 1.49 | 0.59 | 0.43) |
| | | 1.3 | .21 | 0.60 | 1.34 | 0.41 | 0.27) |
| H2 | A | 1.1 | (.16 | 0.73 | 1.28 | 0.86 | 0.29) |
| | | 1.2 | (.14 | 0.72 | 1.04 | 0.56 | -0.13) |
| | | 1.3 | (.00 | 1.05 | 0.97 | 0.45 | -0.24) |
| H3 | A | 1.1 | .30 | 0.39 | 2.31 | 1.42 | 1.22) |
| | | 1.2 | .23 | 0.69 | 1.51 | 1.07 | 0.92) |
| | | 1.3 | .13 | 0.70 | 1.34 | 0.82 | 0.65) |
| I1 | S.A | 1.1 | .00 | 1.67 | 1.94 | 1.20 | 1.00) |
| | | 2.1 | .12 | 1.39 | 1.92 | 1.19 | 1.00) |
| | | 1.2 | .44 | 1.22 | 1.62 | 0.98 | 0.80) |
| | | 2.2 | .20 | 1.11 | 1.62 | 0.99 | 0.80) |
| I2 | S.A | 1.1 | .11 | 1.42 | 1.86 | 1.28 | 1.18) |
| | | 2.1 | .00 | 0.87 | 1.98 | 1.29 | 1.16) |
| | | 1.2 | .00 | 1.88 | 1.56 | 1.07 | 1.00) |
| | | 2.2 | .00 | 0.95 | 1.57 | 1.06 | 1.00) |
| J1 | S.A | 1.1 | .09 | 1.15 | 1.67 | 1.27 | 1.09) |
| | | 1.2 | .25 | 0.37 | 1.48 | 0.79 | 0.48) |
| | | 1.3 | .12 | 0.56 | 1.34 | 0.54 | 0.29) |
| | | 2.1 | .31 | 0.84 | 2.72 | 1.40 | 1.10) |
| | | 2.2 | .11 | 0.96 | 1.76 | 1.04 | 0.77) |
| | | 2.3 | (.12 | 0.79 | 1.58 | 0.79 | 0.46) |

| Experiment | Effekt | Bedingung | Parameter | | | | | |
|---------------|--------|-----------|-----------|------|-------|-------|-------|----------------|
| | | | r | d' | C_R | C_K | C_G | σ_{Alt} |
| HTUVSD-Modell | | | | | | | | |
| A1 | S | 1 | (.00 | 1.93 | 1.35 | 0.73 | 0.22 | 0.71) |
| | | 2 | (.00 | 0.46 | 1.50 | 0.85 | 0.24 | 0.80) |
| A2 | S | 1 | (.00 | 1.12 | 1.96 | 1.08 | 0.76 | 1.28) |
| | | 2 | (.00 | 0.77 | 1.67 | 0.96 | 0.65 | 1.11) |
| A3 | S | 1 | (.00 | 2.64 | 2.76 | 2.05 | 1.49 | 1.42) |
| | | 2 | (.00 | 2.26 | 2.16 | 1.62 | 1.17 | 1.12) |
| A4 | S | 1 | (.00 | 0.43 | 2.00 | 1.24 | 0.78 | 1.25) |
| | | 2 | (.00 | 0.20 | 1.63 | 1.01 | 0.63 | 1.02) |
| A5 | S | 1 | (.00 | 1.18 | 2.54 | 1.67 | 1.37 | 1.31) |
| | | 2 | (.00 | 0.75 | 2.04 | 1.32 | 1.09 | 1.05) |
| A6 | S | 1 | (.00 | 1.25 | 2.00 | 1.32 | 0.93 | 1.25) |
| | | 2 | (.00 | 0.68 | 1.54 | 1.06 | 0.72 | 0.97) |
| A7 | S | 1 | .01 | 1.40 | 1.62 | 1.26 | 0.97 | 0.86) |
| | | 2 | (.00 | 1.21 | 1.68 | 1.31 | 1.01 | 0.89) |
| B1 | S | 1 | (.00 | 1.52 | 1.75 | 1.00 | 0.94 | 0.99) |
| | | 2 | (.00 | 1.07 | 1.57 | 0.73 | 0.71 | 1.21) |
| B2 | S | 1 | (.00 | 0.72 | 0.88 | 0.47 | 0.23 | 0.68) |
| | | 2 | .50 | 1.15 | 3.04 | 1.84 | 1.60 | 1.31) |
| B3 | S | 1 | (.00 | 1.25 | 1.13 | 0.63 | 0.44 | 0.71) |
| | | 2 | .02 | 1.09 | 1.10 | 0.62 | 0.47 | 0.58) |
| B4 | S | 1 | (.00 | 1.13 | 1.56 | 1.10 | 0.71 | 1.46) |
| | | 2 | (.00 | 1.24 | 1.31 | 0.90 | 0.56 | 1.29) |
| B5 | S | 1 | .39 | 0.57 | 2.03 | 0.96 | 0.50 | 1.30) |
| | | 2 | .22 | 1.07 | 1.80 | 1.16 | 0.67 | 1.21) |
| B6 | S | 1 | (.01 | 1.01 | 1.47 | 0.69 | 0.46 | 0.99) |
| | | 2 | (.00 | 0.99 | 1.19 | 0.54 | 0.31 | 0.92) |
| B7 | S | 1 | (.00 | 1.73 | 1.91 | 1.14 | 0.99 | 0.70) |
| | | 2 | (.00 | 1.42 | 1.18 | 0.75 | 0.65 | 0.48) |
| B8 | S | 1 | (.00 | 0.94 | 1.23 | 0.65 | 0.52 | 0.61) |
| | | 2 | (.00 | 1.38 | 2.04 | 1.11 | 1.02 | 0.94) |
| B9 | S | 1 | (.02 | 5.52 | 4.78 | 4.25 | 4.01 | 2.17) |
| | | 2 | (.00 | 4.66 | 4.36 | 3.37 | 3.09 | 1.42) |
| C1 | S | 1 | (.00 | 5.06 | 4.68 | 3.97 | 3.82 | 2.05) |
| | | 2 | (.01 | 2.46 | 2.47 | 2.04 | 1.82 | 0.90) |
| | | 3 | (.00 | 3.71 | 4.15 | 3.67 | 3.43 | 1.84) |
| C2 | S | 1 | (.00 | 1.87 | 1.56 | 1.20 | 0.82 | 1.46) |
| | | 2 | (.00 | 1.25 | 1.27 | 0.80 | 0.36 | 1.27) |
| | | 3 | (.00 | 1.52 | 1.78 | 1.34 | 0.90 | 1.38) |
| C3 | S | 1 | (.00 | 2.08 | 1.63 | 1.16 | 0.74 | 0.98) |
| | | 2 | (.00 | 0.84 | 1.19 | 0.65 | -0.09 | 0.82) |
| | | 3 | (.00 | 0.53 | 1.18 | 0.50 | -0.30 | 0.89) |

| Experiment | Effekt | Bedingung | Parameter | | | | | |
|------------|--------|-----------|-----------|-------|-------|-------|-------|----------------|
| | | | r | d' | C_R | C_K | C_G | σ_{Alt} |
| D1 | S.S | 1.1 | (.00 | 3.03 | 2.93 | 1.65 | 1.45 | 0.84) |
| | | 1.2 | (.00 | 3.50 | 3.93 | 2.45 | 2.24 | .73) |
| | | 2.1 | (.00 | 2.41 | 2.82 | 1.21 | 0.91 | 0.61) |
| | | 2.2 | (.00 | 2.27 | 3.02 | 1.38 | 1.02 | 1.04) |
| D2 | S.S | 1.1 | (.00 | 1.42 | 1.22 | 0.80 | 0.61 | 0.64) |
| | | 1.2 | (.01 | 1.36 | 1.73 | 1.19 | 1.08 | 0.73) |
| | | 2.1 | (.00 | 1.21 | 1.38 | 0.89 | 0.74 | 0.87) |
| | | 2.2 | (.01 | 0.63 | 1.19 | 0.61 | 0.51 | 0.41) |
| D3 | S.S | 1.1 | .06 | 0.66 | 1.56 | 0.89 | 0.02 | 0.90) |
| | | 1.2 | (.00 | 1.35 | 1.67 | 0.83 | 0.03 | 0.92) |
| | | 2.1 | (.00 | 0.87 | 1.66 | 0.80 | 0.12 | 0.92) |
| | | 2.2 | (.00 | 1.58 | 1.71 | 0.91 | 0.28 | 1.07) |
| D4 | S.S | 1.1 | (.00 | 0.69 | 1.21 | 0.68 | 0.32 | 0.61) |
| | | 1.2 | .16 | 0.72 | 1.91 | 1.05 | 0.69 | 0.94) |
| | | 2.1 | (.00 | 1.11 | 1.34 | 0.77 | 0.50 | 0.65) |
| | | 2.2 | (.00 | 0.88 | 1.21 | 0.66 | 0.41 | 0.64) |
| D5 | S.S | 1.1 | (.03 | 2.05 | 2.25 | 1.37 | 1.30 | 2.09) |
| | | 1.2 | (.00 | 0.56 | 0.97 | 0.31 | 0.25 | 1.19) |
| | | 2.1 | (.00 | -0.41 | 0.23 | -0.43 | -0.46 | 0.87) |
| | | 2.2 | (.00 | -0.52 | 0.13 | -0.46 | -0.52 | 0.63) |
| E1 | S.S | 1.1 | (.00 | 0.86 | 1.46 | 0.97 | 0.75 | 0.51) |
| | | 1.2 | (.00 | 0.59 | 1.32 | 0.80 | 0.64 | 0.43) |
| | | 2.1 | (.00 | 1.49 | 1.43 | 0.91 | 0.66 | 0.68) |
| | | 2.2 | (.00 | 1.04 | 1.40 | 0.91 | 0.65 | 0.67) |
| E2 | S.S | 1.1 | .07 | 1.32 | 1.66 | 1.00 | 0.88 | 0.89) |
| | | 1.2 | .32 | 1.34 | 2.11 | 1.27 | 1.12 | 1.13) |
| | | 2.1 | (.00 | 1.67 | 1.62 | 1.01 | 0.82 | 0.87) |
| | | 2.2 | (.00 | 1.21 | 1.37 | 0.84 | 0.69 | 0.72) |
| E3 | S.S | 1.1 | .05 | 1.50 | 1.45 | 0.90 | 0.81 | 1.02) |
| | | 1.2 | (.00 | 1.40 | 1.54 | 0.93 | 0.86 | 1.08) |
| | | 2.1 | .17 | 0.68 | 1.18 | 0.55 | 0.45 | 1.12) |
| | | 2.2 | (.00 | 0.77 | 1.10 | 0.50 | 0.42 | 1.04) |
| E4 | S.S | 1.1 | .02 | 3.38 | 3.47 | 2.54 | 1.81 | 1.98) |
| | | 1.2 | (.00 | 1.12 | 1.72 | 1.19 | 0.88 | 0.96) |
| | | 2.1 | (.00 | 2.44 | 2.05 | 1.38 | 1.07 | 1.45) |
| | | 2.2 | (.00 | 1.66 | 1.60 | 1.07 | 0.83 | 1.13) |
| E5 | S.S | 1.1 | .01 | 0.48 | 0.67 | 0.45 | 0.34 | 0.29) |
| | | 1.2 | .26 | 0.85 | 2.71 | 1.83 | 1.38 | 1.17) |
| | | 2.1 | (.01 | 1.37 | 1.14 | 0.74 | 0.63 | 0.56) |
| | | 2.2 | (.00 | 1.09 | 1.56 | 1.02 | 0.86 | 0.76) |

| Experiment | Effekt | Bedingung | Parameter | | | | | |
|------------|--------|-----------|-----------|------|-------|-------|-------|----------------|
| | | | r | d' | C_R | C_K | C_G | σ_{Alt} |
| E6 | S.S | 1.1 | (.00 | 0.28 | 1.17 | 0.73 | 0.64 | 0.62) |
| | | 1.2 | .01 | 0.33 | 1.81 | 1.13 | 1.00 | 0.96 |
| | | 2.1 | (.00 | 0.76 | 1.46 | 0.87 | 0.75 | 0.70) |
| | | 2.2 | (.00 | 0.54 | 1.66 | 1.02 | 0.87 | 0.80) |
| E7 | S.S | 1.1 | .04 | 3.47 | 3.36 | 2.92 | 2.69 | 1.47 |
| | | 1.2 | .06 | 1.67 | 3.03 | 2.66 | 2.43 | 1.33 |
| | | 2.1 | (.00 | 2.05 | 1.40 | 1.05 | 0.81 | 0.54) |
| | | 2.2 | (.00 | 0.80 | 1.75 | 1.42 | 1.05 | 0.71) |
| E8 | S.S | 1.1 | .45 | 2.34 | 3.89 | 2.73 | 2.41 | 2.23 |
| | | 1.2 | .27 | 1.99 | 3.95 | 2.77 | 2.44 | 2.26 |
| | | 2.1 | (.00 | 2.68 | 2.34 | 1.47 | 1.14 | 1.40) |
| | | 2.2 | (.00 | 1.67 | 1.89 | 1.22 | 0.93 | 1.14) |
| E9 | S.S | 1.1 | .30 | 2.18 | 3.23 | 1.85 | 1.77 | 1.57 |
| | | 1.2 | (.00 | 0.75 | 1.78 | 1.01 | 0.97 | 0.86) |
| | | 2.1 | .27 | 1.51 | 1.79 | 0.98 | 0.94 | 0.87 |
| | | 2.2 | (.00 | 0.52 | 1.48 | 0.81 | 0.78 | 0.72) |
| E10 | S.S | 1.1 | (.00 | 0.88 | 1.23 | 0.91 | 0.65 | 0.50) |
| | | 1.2 | (.00 | 0.80 | 1.36 | 1.04 | 0.73 | 0.57) |
| | | 2.1 | (.00 | 1.27 | 1.91 | 1.41 | 1.21 | 1.08) |
| | | 2.2 | (.00 | 0.91 | 1.62 | 1.22 | 1.04 | 0.92) |
| E11 | S.S | 1.1 | (.00 | 2.34 | 2.66 | 1.83 | 1.63 | 1.08) |
| | | 1.2 | .01 | 2.88 | 3.37 | 2.53 | 2.15 | 1.44) |
| | | 2.1 | (.00 | 2.43 | 2.72 | 1.92 | 1.68 | 1.12) |
| | | 2.2 | (.00 | 2.95 | 2.93 | 2.03 | 1.80 | 1.20) |
| E12 | S.S | 1.1 | .07 | 5.81 | 5.35 | 4.32 | 3.91 | 2.72 |
| | | 1.2 | (.00 | 4.58 | 4.20 | 3.32 | 3.05 | 2.12) |
| | | 2.1 | (.00 | 3.48 | 3.77 | 2.97 | 2.74 | 1.90) |
| | | 2.2 | (.00 | 3.24 | 3.39 | 2.62 | 2.45 | 1.70) |
| F1 | S.S | 1.1 | (.00 | 1.44 | 1.57 | 1.04 | 0.80 | 0.94) |
| | | 1.2 | (.00 | 1.23 | 1.84 | 1.11 | 0.84 | 1.05) |
| | | 1.3 | .02 | 1.12 | 1.74 | 1.09 | 0.80 | 0.74 |
| | | 2.1 | .19 | 2.76 | 2.07 | 1.54 | 1.31 | 0.89 |
| | | 2.2 | .28 | 0.97 | 1.47 | 0.42 | 0.00 | 0.83 |
| | | 2.3 | (.01 | 1.94 | 2.17 | 1.40 | 1.09 | 0.73) |
| F2 | S.S | 1.1 | (.01 | 1.27 | 1.02 | 0.69 | 0.57 | 0.37) |
| | | 1.2 | (.00 | 1.62 | 1.16 | 0.84 | 0.71 | 0.48) |
| | | 1.3 | (.00 | 1.86 | 1.28 | 0.99 | 0.81 | 0.61) |
| | | 2.1 | .31 | 1.00 | 1.82 | 1.12 | 0.91 | 0.67 |
| | | 2.2 | .43 | 1.85 | 1.73 | 1.11 | 0.91 | 0.75 |
| | | 2.3 | (.00 | 2.13 | 1.42 | 1.03 | 0.88 | 0.65) |
| G1 | A | 1.1 | .12 | 0.61 | 1.00 | 0.47 | 0.17 | 0.85 |
| | | 1.2 | .31 | 0.85 | 1.83 | 1.11 | 0.79 | 1.11 |

| Experiment | Effekt | Bedingung | Parameter | | | | | |
|------------|--------|-----------|-----------|------|-------|-------|-------|----------------|
| | | | r | d' | C_R | C_K | C_G | σ_{Alt} |
| G2 | A | 1.1 | (.00 | 1.55 | 1.65 | 1.21 | 0.93 | 0.75) |
| | | 1.2 | (.00 | 1.25 | 1.17 | 0.72 | 0.33 | 0.66) |
| H1 | A | 1.1 | (.00 | 0.87 | 1.07 | 0.62 | 0.53 | 0.51) |
| | | 1.2 | (.00 | 0.65 | 0.92 | 0.36 | 0.26 | 0.61) |
| | | 1.3 | (.00 | 0.73 | 1.01 | 0.31 | 0.19 | 0.74) |
| H2 | A | 1.1 | (.00 | 0.79 | 1.04 | 0.69 | 0.23 | 0.80) |
| | | 1.2 | (.00 | 0.81 | 0.90 | 0.49 | -0.12 | 0.85) |
| | | 1.3 | (.00 | 1.05 | 0.98 | 0.46 | -0.24 | 1.02) |
| H3 | A | 1.1 | .31 | 0.59 | 2.74 | 1.67 | 1.45 | 1.19 |
| | | 1.2 | (.00 | 0.69 | 0.97 | 0.68 | 0.58 | 0.63) |
| | | 1.3 | (.00 | 0.73 | 1.10 | 0.67 | 0.53 | 0.82) |
| I1 | S.A | 1.1 | (.00 | 1.69 | 1.95 | 1.21 | 1.01 | 1.01) |
| | | 2.1 | (.00 | 1.33 | 1.63 | 1.00 | 0.84 | 0.84) |
| | | 1.2 | (.01 | 1.37 | 1.05 | 0.63 | 0.51 | 0.64) |
| | | 2.2 | .01 | 1.15 | 1.30 | 0.80 | 0.64 | 0.80) |
| I2 | S.A | 1.1 | .21 | 1.51 | 2.22 | 1.53 | 1.41 | 1.19 |
| | | 2.1 | .08 | 1.86 | 3.53 | 2.43 | 2.24 | 1.89 |
| | | 1.2 | .03 | 1.89 | 1.60 | 1.10 | 1.02 | 1.02 |
| | | 2.2 | (.00 | 0.95 | 1.57 | 1.06 | 1.00 | 1.00) |
| J1 | S.A | 1.1 | (.00 | 1.06 | 1.41 | 1.06 | 0.91 | 0.84) |
| | | 1.2 | .03 | 0.49 | 0.92 | 0.50 | 0.29 | 0.62 |
| | | 1.3 | (.00 | 0.62 | 1.12 | 0.46 | 0.24 | 0.83) |
| | | 2.1 | .26 | 0.61 | 1.95 | 1.04 | 0.79 | 0.72 |
| | | 2.2 | .27 | 1.37 | 3.20 | 1.80 | 1.40 | 1.81 |
| | | 2.3 | (.00 | 0.78 | 1.26 | 0.63 | 0.35 | 0.78) |

| Experiment | Effekt | Bedingung | Parameter | | | | | |
|-----------------|--------|-----------|-----------|----------|----------|----------------------|----------------------|----------------------|
| | | | <i>r</i> | <i>k</i> | <i>d</i> | <i>g_R</i> | <i>g_K</i> | <i>g_G</i> |
| 4-States-Modell | | | | | | | | |
| A1 | S | 1 | .71 | .55 | .38 | .05 | .19 | .50 |
| | | 2 | .12 | .10 | .21 | .04 | .15 | .37 |
| A2 | S | 1 | .13 | .32 | .25 | .09 | .16 | .18 |
| | | 2 | .11 | .19 | .22 | .09 | .15 | .17 |
| A3 | S | 1 | .40 | .35 | .78 | .13 | .21 | .52 |
| | | 2 | .50 | .31 | .75 | .11 | .18 | .44 |
| A4 | S | 1 | (.00 | .09 | .09 | .06 | .11 | .16) |
| | | 2 | (.02 | .08 | .00 | .06 | .10 | .14) |
| A5 | S | 1 | .03 | .17 | .47 | .06 | .12 | .13 |
| | | 2 | .06 | .18 | .09 | .03 | .07 | .07 |
| A6 | S | 1 | .15 | .23 | .45 | .11 | .14 | .24 |
| | | 2 | .12 | .12 | .29 | .08 | .11 | .17 |
| A7 | S | 1 | .36 | .11 | .69 | .10 | .14 | .25 |
| | | 2 | .26 | .08 | .66 | .09 | .13 | .22 |
| B1 | S | 1 | .38 | .37 | .35 | .06 | .18 | .04 |
| | | 2 | (.26 | .33 | .00 | .10 | .19 | .01) |
| B2 | S | 1 | .38 | .13 | .13 | .11 | .18 | .21 |
| | | 2 | .50 | .03 | .63 | .03 | .20 | .11 |
| B3 | S | 1 | .52 | .26 | .35 | .09 | .20 | .19 |
| | | 2 | .49 | .20 | .43 | .05 | .20 | .16 |
| B4 | S | 1 | .13 | .11 | .44 | .26 | .16 | .29 |
| | | 2 | .32 | .17 | .39 | .26 | .15 | .29 |
| B5 | S | 1 | .38 | .02 | .33 | .09 | .28 | .27 |
| | | 2 | .31 | .08 | .49 | .14 | .23 | .36 |
| B6 | S | 1 | .26 | .18 | .39 | .11 | .31 | .22 |
| | | 2 | .35 | .22 | .27 | .14 | .27 | .22 |
| B7 | S | 1 | .42 | .54 | .30 | .01 | .03 | .08 |
| | | 2 | (.59 | .42 | .00 | .01 | .02 | .05) |
| B8 | S | 1 | (.37 | .36 | .00 | .03 | .09 | .07) |
| | | 2 | (.24 | .45 | .00 | .02 | .08 | .04) |
| B9 | S | 1 | .74 | .45 | .94 | .25 | .11 | .32 |
| | | 2 | .61 | .71 | .97 | .08 | .08 | .37 |
| C1 | S | 1 | .63 | .57 | .90 | .14 | .05 | .17 |
| | | 2 | .48 | .27 | .93 | .06 | .07 | .22 |
| | | 3 | .24 | .21 | .91 | .15 | .06 | .19 |
| C2 | S | 1 | .48 | .18 | .55 | .33 | .16 | .38 |
| | | 2 | .35 | .19 | .39 | .27 | .20 | .39 |
| | | 3 | .27 | .20 | .52 | .22 | .11 | .34 |
| C3 | S | 1 | .65 | .48 | .49 | .12 | .07 | .33 |
| | | 2 | .31 | .22 | .18 | .10 | .16 | .56 |
| | | 3 | .18 | .20 | .06 | .11 | .20 | .55 |

| Experiment | Effekt | Bedingung | Parameter | | | | | |
|------------|--------|-----------|-----------|----------|----------|----------------------|----------------------|----------------------|
| | | | <i>r</i> | <i>k</i> | <i>d</i> | <i>g_R</i> | <i>g_K</i> | <i>g_G</i> |
| D1 | S.S | 1.1 | .54 | .81 | .87 | .02 | .08 | .25 |
| | | 1.2 | .32 | .75 | .72 | .07 | .11 | .21 |
| | | 2.1 | (.34 | .82 | .82 | .01 | .06 | .36) |
| | | 2.2 | .22 | .73 | .63 | .03 | .14 | .36 |
| D2 | S.S | 1.1 | .56 | .25 | .54 | .06 | .16 | .19 |
| | | 1.2 | .34 | .20 | .73 | .04 | .16 | .09 |
| | | 2.1 | .39 | .26 | .32 | .09 | .13 | .11 |
| | | 2.2 | (.29 | .23 | .43 | .00 | .11 | .08) |
| D3 | S.S | 1.1 | .19 | .12 | .25 | .05 | .17 | .56 |
| | | 1.2 | .35 | .40 | .34 | .06 | .21 | .65 |
| | | 2.1 | .18 | .26 | .28 | .06 | .21 | .50 |
| | | 2.2 | .41 | .38 | .45 | .11 | .26 | .58 |
| D4 | S.S | 1.1 | .28 | .22 | .16 | .04 | .11 | .25 |
| | | 1.2 | .24 | .10 | .42 | .04 | .20 | .23 |
| | | 2.1 | .39 | .25 | .45 | .04 | .17 | .25 |
| | | 2.2 | .35 | .19 | .35 | .05 | .19 | .22 |
| D5 | S.S | 1.1 | .28 | .27 | .54 | .31 | .34 | .08 |
| | | 1.2 | .20 | .18 | .01 | .21 | .23 | .05 |
| | | 2.1 | (.00 | .00 | .00 | .28 | .35 | .03) |
| | | 2.2 | (.00 | .00 | .00 | .28 | .32 | .05) |
| E1 | S.S | 1.1 | (.27 | .20 | .69 | .01 | .08 | .15) |
| | | 1.2 | (.23 | .22 | .49 | .01 | .05 | .09) |
| | | 2.1 | .51 | .37 | .50 | .04 | .13 | .20 |
| | | 2.2 | .33 | .22 | .48 | .04 | .12 | .19 |
| E2 | S.S | 1.1 | .37 | .17 | .61 | .08 | .27 | .12 |
| | | 1.2 | .43 | .13 | .63 | .08 | .29 | .13 |
| | | 2.1 | .49 | .28 | .63 | .08 | .26 | .21 |
| | | 2.2 | .41 | .25 | .45 | .06 | .17 | .12 |
| E3 | S.S | 1.1 | .46 | .17 | .55 | .17 | .30 | .10 |
| | | 1.2 | .41 | .35 | .04 | .08 | .12 | .03 |
| | | 2.1 | .28 | .05 | .32 | .21 | .31 | .08 |
| | | 2.2 | .27 | .18 | .11 | .16 | .22 | .05 |
| E4 | S.S | 1.1 | .40 | .38 | .79 | .19 | .34 | .71 |
| | | 1.2 | .22 | .19 | .49 | .08 | .13 | .20 |
| | | 2.1 | .60 | .44 | .61 | .20 | .26 | .30 |
| | | 2.2 | .46 | .30 | .48 | .15 | .18 | .19 |
| E5 | S.S | 1.1 | .42 | .10 | .21 | .01 | .06 | .08 |
| | | 1.2 | .26 | .02 | .56 | .02 | .12 | .16 |
| | | 2.1 | .58 | .23 | .56 | .05 | .17 | .11 |
| | | 2.2 | .29 | .14 | .60 | .05 | .19 | .13 |

| Experiment | Effekt | Bedingung | Parameter | | | | | |
|------------|--------|-----------|-----------|-----|-----|-------|-------|-------|
| | | | r | k | d | g_R | g_K | g_G |
| E6 | S.S | 1.1 | .16 | .06 | .21 | .04 | .12 | .04 |
| | | 1.2 | .04 | .02 | .29 | .04 | .13 | .05 |
| | | 2.1 | .22 | .21 | .22 | .03 | .11 | .06 |
| | | 2.2 | .11 | .10 | .36 | .03 | .13 | .07 |
| E7 | S.S | 1.1 | .49 | .17 | .93 | .17 | .20 | .25 |
| | | 1.2 | .08 | .01 | .83 | .07 | .07 | .08 |
| | | 2.1 | .73 | .38 | .79 | .04 | .06 | .27 |
| | | 2.2 | .15 | .08 | .70 | .03 | .04 | .18 |
| E8 | S.S | 1.1 | (.39) | .00 | .73 | .15 | .31 | (.19) |
| | | 1.2 | (.20) | .00 | .64 | .11 | .21 | (.13) |
| | | 2.1 | .61 | .59 | .64 | .14 | .28 | .31 |
| | | 2.2 | .34 | .24 | .61 | .13 | .26 | .28 |
| E9 | S.S | 1.1 | .36 | .22 | .75 | .08 | .43 | .08 |
| | | 1.2 | .13 | .17 | .28 | .03 | .14 | .02 |
| | | 2.1 | .53 | .26 | .65 | .06 | .33 | .05 |
| | | 2.2 | .15 | .13 | .26 | .03 | .15 | .02 |
| E10 | S.S | 1.1 | .35 | .18 | .58 | .02 | .05 | .18 |
| | | 1.2 | .27 | .13 | .61 | .03 | .05 | .19 |
| | | 2.1 | .19 | .13 | .60 | .10 | .14 | .13 |
| | | 2.2 | .17 | .06 | .56 | .09 | .13 | .11 |
| E11 | S.S | 1.1 | .36 | .43 | .80 | .05 | .15 | .19 |
| | | 1.2 | .28 | .28 | .88 | .08 | .25 | .36 |
| | | 2.1 | .36 | .39 | .83 | .06 | .18 | .23 |
| | | 2.2 | .50 | .55 | .85 | .07 | .19 | .26 |
| E12 | S.S | 1.1 | .66 | .54 | .91 | .28 | .46 | .56 |
| | | 1.2 | .61 | .50 | .89 | .22 | .34 | .35 |
| | | 2.1 | .30 | .30 | .86 | .18 | .26 | .23 |
| | | 2.2 | .38 | .37 | .83 | .15 | .20 | .17 |
| F1 | S.S | 1.1 | .39 | .20 | .57 | .12 | .21 | .24 |
| | | 1.2 | .21 | .17 | .54 | .09 | .24 | .22 |
| | | 1.3 | .26 | .20 | .62 | .03 | .16 | .23 |
| | | 2.1 | .79 | .38 | .87 | .08 | .25 | .33 |
| | | 2.2 | .48 | .32 | .24 | .05 | .38 | .43 |
| | | 2.3 | .40 | .42 | .84 | .02 | .14 | .31 |
| F2 | S.S | 1.1 | .60 | .24 | .72 | .01 | .09 | .14 |
| | | 1.2 | .67 | .29 | .68 | .03 | .07 | .13 |
| | | 1.3 | .71 | .30 | .68 | .06 | .07 | .17 |
| | | 2.1 | .45 | .19 | .69 | .01 | .14 | .16 |
| | | 2.2 | .73 | .31 | .75 | .04 | .25 | .24 |
| | | 2.3 | .76 | .47 | .39 | .03 | .04 | .09 |
| G1 | A | 1.1 | .34 | .08 | .20 | .15 | .25 | .26 |
| | | 1.2 | .36 | .07 | .46 | .09 | .22 | .21 |

| Experiment | Effekt | Bedingung | Parameter | | | | | |
|------------|--------|-----------|-----------|----------|----------|----------------------|----------------------|----------------------|
| | | | <i>r</i> | <i>k</i> | <i>d</i> | <i>g_R</i> | <i>g_K</i> | <i>g_G</i> |
| G2 | A | 1.1 | .44 | .30 | .66 | .06 | .06 | .23 |
| | | 1.2 | .50 | .26 | .39 | .07 | .16 | .38 |
| H1 | A | 1.1 | (.41 | .27 | .00 | .02 | .08 | .05) |
| | | 1.2 | (.35 | .22 | .00 | .07 | .21 | .09) |
| | | 1.3 | (.34 | .25 | .00 | .09 | .26 | .10) |
| H2 | A | 1.1 | .33 | .14 | .22 | .13 | .13 | .34 |
| | | 1.2 | .37 | .17 | .15 | .18 | .17 | .50 |
| | | 1.3 | .43 | .23 | .22 | .23 | .20 | .63 |
| H3 | A | 1.1 | (.30 | .00 | .43 | .02 | .12 | .06) |
| | | 1.2 | .35 | .12 | .11 | .07 | .08 | .06 |
| | | 1.3 | .27 | .10 | .26 | .12 | .17 | .11 |
| I1 | S.A | 1.1 | .36 | .29 | .65 | .08 | .26 | .20 |
| | | 2.1 | .35 | .27 | .53 | .06 | .19 | .13 |
| | | 1.2 | .61 | .27 | .33 | .08 | .17 | .10 |
| | | 2.2 | .39 | .15 | .48 | .10 | .23 | .15 |
| I2 | S.A | 1.1 | .34 | .09 | .70 | .10 | .26 | .09 |
| | | 2.1 | (.02 | .00 | .69 | .10 | .25 | .09) |
| | | 1.2 | .57 | .20 | .65 | .17 | .28 | .09 |
| | | 2.2 | .21 | .14 | .32 | .09 | .13 | .03 |
| J1 | S.A | 1.1 | .31 | .16 | .42 | .08 | .08 | .10 |
| | | 1.2 | .30 | .09 | .11 | .08 | .17 | .16 |
| | | 1.3 | .24 | .14 | .17 | .11 | .27 | .19 |
| | | 2.1 | (.33 | .15 | .49 | .01 | .14 | .14) |
| | | 2.2 | (.23 | .00 | .58 | .09 | .31 | .23) |
| | | 2.3 | .28 | .27 | .08 | .06 | .15 | .19 |

| Experiment | Effekt | Bedingung | Parameter | | | | | |
|---------------|--------|-----------|-----------|-------|-------|-------|-------|----------------|
| | | | d_x | d_y | C_R | C_G | C_O | σ_{Neu} |
| Streak-Modell | | | | | | | | |
| A1 | S | 1 | 1.01 | 1.91 | 0.84 | 0.51 | 0.14 | 0.48 |
| | | 2 | 0.25 | 0.59 | -0.18 | 0.85 | 0.24 | 0.80 |
| A2 | S | 1 | 0.58 | 0.42 | -0.40 | 0.56 | 0.36 | 0.61 |
| | | 2 | 0.39 | 0.33 | -0.28 | 0.63 | 0.40 | 0.69 |
| A3 | S | 1 | 1.50 | 1.88 | 0.30 | 1.73 | 1.21 | 1.16 |
| | | 2 | 1.23 | 1.78 | 0.58 | 1.35 | 0.95 | 0.91 |
| A4 | S | 1 | 0.57 | 0.10 | -0.78 | 0.93 | 0.55 | 0.90 |
| | | 2 | 0.33 | 0.00 | -0.52 | 0.74 | 0.44 | 0.71 |
| A5 | S | 1 | 0.62 | 0.35 | -0.70 | 1.05 | 0.82 | 0.79 |
| | | 2 | 0.37 | 0.06 | -0.55 | 0.61 | 0.47 | 0.46 |
| A6 | S | 1 | 0.65 | 0.60 | -0.15 | 0.89 | 0.58 | 0.79 |
| | | 2 | 0.31 | 0.38 | 0.03 | 0.82 | 0.54 | 0.72 |
| A7 | S | 1 | 0.79 | 1.38 | 0.67 | 1.22 | 0.93 | 0.83 |
| | | 2 | 0.68 | 1.16 | 0.51 | 1.28 | 0.98 | 0.87 |
| B1 | S | 1 | 0.57 | 0.91 | 0.20 | 0.44 | 0.40 | 0.42 |
| | | 2 | 0.32 | 0.33 | -0.04 | 0.10 | 0.09 | 0.16 |
| B2 | S | 1 | 0.37 | 0.91 | 0.59 | 0.41 | 0.19 | 0.58 |
| | | 2 | 0.60 | 1.89 | 0.93 | 0.84 | 0.73 | 0.60 |
| B3 | S | 1 | 0.63 | 1.29 | 0.65 | 0.49 | 0.33 | 0.54 |
| | | 2 | 0.58 | 1.40 | 0.64 | 0.58 | 0.43 | 0.54 |
| B4 | S | 1 | 0.77 | 0.63 | 0.34 | 0.92 | 0.56 | 1.16 |
| | | 2 | 0.71 | 0.80 | 0.59 | 0.67 | 0.38 | 0.89 |
| B5 | S | 1 | 0.47 | 1.35 | 0.58 | 0.63 | 0.33 | 0.85 |
| | | 2 | 0.67 | 1.21 | 0.49 | 0.97 | 0.56 | 1.02 |
| B6 | S | 1 | 0.57 | 0.99 | 0.06 | 0.66 | 0.43 | 0.93 |
| | | 2 | 0.53 | 0.93 | 0.30 | 0.45 | 0.24 | 0.74 |
| B7 | S | 1 | 0.56 | 0.73 | 0.18 | 0.24 | 0.18 | 0.13 |
| | | 2 | 0.42 | 0.99 | 0.73 | 0.04 | 0.03 | 0.02 |
| B8 | S | 1 | 0.28 | 0.50 | 0.24 | 0.12 | 0.09 | 0.10 |
| | | 2 | 0.41 | 0.30 | -0.25 | 0.17 | 0.14 | 0.13 |
| B9 | S | 1 | 2.21 | 2.50 | 1.06 | 2.00 | 1.80 | 0.98 |
| | | 2 | 1.87 | 2.22 | 0.48 | 1.55 | 1.31 | 0.60 |
| C1 | S | 1 | 1.39 | 1.64 | 0.67 | 1.00 | 0.89 | 0.48 |
| | | 2 | 1.11 | 1.59 | 0.67 | 1.39 | 1.18 | 0.59 |
| | | 3 | 1.61 | 1.38 | 0.34 | 2.05 | 1.82 | 0.97 |
| C2 | S | 1 | 1.10 | 1.22 | 0.90 | 0.93 | 0.60 | 1.08 |
| | | 2 | 0.75 | 0.89 | 0.55 | 0.66 | 0.28 | 0.98 |
| | | 3 | 0.89 | 0.83 | 0.44 | 0.98 | 0.60 | 0.93 |
| C3 | S | 1 | 0.96 | 1.41 | 0.80 | 0.55 | 0.30 | 0.41 |
| | | 2 | 0.48 | 0.82 | 0.28 | 0.60 | -0.09 | 0.71 |
| | | 3 | 0.33 | 0.46 | -0.07 | 0.46 | -0.27 | 0.75 |

| Experiment | Effekt | Bedingung | Parameter | | | | | |
|------------|--------|-----------|-----------|-------|-------|-------|-------|----------------|
| | | | d_x | d_y | C_R | C_G | C_O | σ_{Neu} |
| D1 | S.S | 1.1 | 1.32 | 1.64 | 0.22 | 0.66 | 0.51 | 0.30 |
| | | 1.2 | 1.31 | 1.17 | -0.30 | 0.66 | 0.52 | 0.40 |
| | | 2.1 | 1.35 | 1.38 | -0.30 | 0.70 | 0.46 | 0.31 |
| | | 2.2 | 1.24 | 1.12 | -0.61 | 0.72 | 0.44 | 0.46 |
| D2 | S.S | 1.1 | 0.71 | 1.49 | 0.79 | 0.64 | 0.48 | 0.50 |
| | | 1.2 | 0.66 | 1.29 | 0.34 | 1.00 | 0.90 | 0.61 |
| | | 2.1 | 0.49 | 0.88 | 0.45 | 0.49 | 0.39 | 0.46 |
| | | 2.2 | 0.31 | 0.99 | 0.17 | 0.56 | 0.46 | 0.37 |
| D3 | S.S | 1.1 | 0.39 | 0.85 | 0.05 | 0.85 | 0.02 | 0.86 |
| | | 1.2 | 0.77 | 1.21 | 0.05 | 0.75 | 0.02 | 0.78 |
| | | 2.1 | 0.51 | 0.75 | -0.28 | 0.74 | 0.10 | 0.81 |
| | | 2.2 | 0.90 | 1.39 | 0.23 | 0.81 | 0.23 | 0.92 |
| D4 | S.S | 1.1 | 0.33 | 0.72 | 0.19 | 0.55 | 0.25 | 0.47 |
| | | 1.2 | 0.43 | 1.13 | 0.13 | 0.87 | 0.57 | 0.78 |
| | | 2.1 | 0.58 | 1.23 | 0.36 | 0.69 | 0.44 | 0.57 |
| | | 2.2 | 0.47 | 1.11 | 0.32 | 0.63 | 0.39 | 0.61 |
| D5 | S.S | 1.1 | 1.09 | 1.07 | 0.20 | 0.82 | 0.75 | 1.22 |
| | | 1.2 | 0.28 | 0.36 | 0.17 | 0.19 | 0.14 | 0.66 |
| | | 2.1 | -0.06 | 0.04 | 0.03 | -0.09 | -0.12 | 0.22 |
| | | 2.2 | 0.17 | -0.20 | 0.10 | -0.21 | -0.26 | 0.31 |
| E1 | S.S | 1.1 | 0.41 | 1.03 | 0.23 | 0.87 | 0.66 | 0.45 |
| | | 1.2 | 0.19 | 0.55 | 0.09 | 0.55 | 0.42 | 0.28 |
| | | 2.1 | 0.68 | 1.29 | 0.54 | 0.57 | 0.40 | 0.41 |
| | | 2.2 | 0.49 | 1.00 | 0.33 | 0.72 | 0.50 | 0.52 |
| E2 | S.S | 1.1 | 0.75 | 1.45 | 0.34 | 0.95 | 0.83 | 0.84 |
| | | 1.2 | 0.80 | 1.64 | 0.50 | 0.98 | 0.86 | 0.87 |
| | | 2.1 | 0.90 | 1.63 | 0.51 | 0.90 | 0.72 | 0.76 |
| | | 2.2 | 0.56 | 1.12 | 0.43 | 0.60 | 0.48 | 0.51 |
| E3 | S.S | 1.1 | 0.83 | 1.42 | 0.62 | 0.78 | 0.70 | 0.89 |
| | | 1.2 | 0.46 | 0.77 | 0.36 | 0.28 | 0.25 | 0.32 |
| | | 2.1 | 0.46 | 0.86 | 0.46 | 0.47 | 0.39 | 0.97 |
| | | 2.2 | 0.35 | 0.58 | 0.26 | 0.31 | 0.26 | 0.64 |
| E4 | S.S | 1.1 | 2.16 | 2.66 | 0.22 | 2.50 | 1.78 | 1.95 |
| | | 1.2 | 0.52 | 0.74 | 0.13 | 0.88 | 0.63 | 0.69 |
| | | 2.1 | 1.19 | 1.67 | 0.67 | 0.81 | 0.60 | 0.82 |
| | | 2.2 | 0.77 | 1.15 | 0.53 | 0.64 | 0.47 | 0.64 |
| E5 | S.S | 1.1 | 0.23 | 1.24 | 0.93 | 0.40 | 0.31 | 0.26 |
| | | 1.2 | 0.38 | 1.41 | 0.64 | 1.04 | 0.78 | 0.67 |
| | | 2.1 | 0.67 | 1.55 | 0.83 | 0.59 | 0.49 | 0.44 |
| | | 2.2 | 0.59 | 1.19 | 0.26 | 0.97 | 0.82 | 0.73 |

| Experiment | Effekt | Bedingung | Parameter | | | | | |
|------------|--------|-----------|-----------|-------|-------|-------|-------|----------------|
| | | | d_x | d_y | C_R | C_G | C_O | σ_{Neu} |
| E6 | S.S | 1.1 | 0.13 | 0.63 | 0.19 | 0.69 | 0.61 | 0.59 |
| | | 1.2 | 0.18 | 0.39 | -0.35 | 1.09 | 0.96 | 0.93 |
| | | 2.1 | 0.23 | 0.50 | 0.03 | 0.50 | 0.42 | 0.39 |
| | | 2.2 | 0.22 | 0.47 | -0.24 | 0.87 | 0.73 | 0.68 |
| E7 | S.S | 1.1 | 2.01 | 2.56 | 0.77 | 2.58 | 2.36 | 1.29 |
| | | 1.2 | 0.87 | 1.18 | 0.44 | 2.20 | 2.01 | 1.10 |
| | | 2.1 | 0.97 | 1.85 | 1.13 | 0.68 | 0.48 | 0.33 |
| | | 2.2 | 0.37 | 0.70 | 0.33 | 1.27 | 0.91 | 0.62 |
| E8 | S.S | 1.1 | 1.15 | 1.98 | 0.68 | 1.62 | 1.43 | 1.33 |
| | | 1.2 | 0.86 | 1.53 | 0.45 | 1.67 | 1.47 | 1.37 |
| | | 2.1 | 1.26 | 1.83 | 0.53 | 0.80 | 0.60 | 0.74 |
| | | 2.2 | 0.92 | 1.37 | 0.26 | 1.07 | 0.80 | 0.99 |
| E9 | S.S | 1.1 | 1.24 | 2.24 | 0.10 | 1.52 | 1.46 | 1.30 |
| | | 1.2 | 0.22 | 0.36 | -0.32 | 0.62 | 0.60 | 0.53 |
| | | 2.1 | 0.89 | 1.87 | 0.55 | 0.83 | 0.80 | 0.74 |
| | | 2.2 | 0.19 | 0.51 | -0.16 | 0.64 | 0.61 | 0.56 |
| E10 | S.S | 1.1 | 0.38 | 0.95 | 0.58 | 0.70 | 0.48 | 0.37 |
| | | 1.2 | 0.37 | 0.91 | 0.50 | 0.90 | 0.61 | 0.48 |
| | | 2.1 | 0.60 | 0.83 | 0.19 | 1.10 | 0.92 | 0.82 |
| | | 2.2 | 0.48 | 0.81 | 0.33 | 1.12 | 0.95 | 0.84 |
| E11 | S.S | 1.1 | 1.00 | 1.33 | 0.08 | 1.07 | 0.90 | 0.60 |
| | | 1.2 | 1.68 | 2.26 | -0.01 | 2.34 | 1.96 | 1.31 |
| | | 2.1 | 1.15 | 1.58 | 0.13 | 1.34 | 1.12 | 0.75 |
| | | 2.2 | 1.33 | 1.84 | 0.30 | 1.21 | 1.01 | 0.68 |
| E12 | S.S | 1.1 | 3.53 | 4.20 | 0.66 | 3.88 | 3.49 | 2.42 |
| | | 1.2 | 2.34 | 2.87 | 0.59 | 2.34 | 2.11 | 1.47 |
| | | 2.1 | 1.80 | 2.01 | 0.13 | 2.16 | 1.95 | 1.35 |
| | | 2.2 | 1.37 | 1.62 | 0.24 | 1.46 | 1.32 | 0.91 |
| F1 | S.S | 1.1 | 0.77 | 1.29 | 0.47 | 0.91 | 0.68 | 0.81 |
| | | 1.2 | 0.69 | 1.06 | -0.02 | 1.02 | 0.76 | 0.95 |
| | | 1.3 | 0.61 | 1.25 | 0.10 | 1.05 | 0.77 | 0.71 |
| | | 2.1 | 1.54 | 2.79 | 1.17 | 1.36 | 1.15 | 0.78 |
| | | 2.2 | 0.64 | 1.54 | 0.34 | 0.37 | 0.00 | 0.73 |
| | | 2.3 | 1.03 | 1.75 | 0.20 | 1.22 | 0.93 | 0.62 |
| F2 | S.S | 1.1 | 0.64 | 1.66 | 0.96 | 0.60 | 0.48 | 0.31 |
| | | 1.2 | 0.71 | 1.62 | 1.08 | 0.49 | 0.39 | 0.26 |
| | | 1.3 | 0.82 | 1.69 | 1.16 | 0.56 | 0.43 | 0.32 |
| | | 2.1 | 0.61 | 1.65 | 0.60 | 0.84 | 0.68 | 0.50 |
| | | 2.2 | 1.12 | 2.39 | 1.05 | 0.92 | 0.75 | 0.62 |
| | | 2.3 | 0.78 | 1.61 | 1.11 | 0.24 | 0.18 | 0.14 |
| G1 | A | 1.1 | 0.39 | 0.93 | 0.54 | 0.43 | 0.16 | 0.78 |
| | | 1.2 | 0.55 | 1.29 | 0.59 | 0.81 | 0.57 | 0.81 |

| Experiment | Effekt | Bedingung | Parameter | | | | | |
|------------|--------|-----------|-----------|-------|-------|-------|-------|----------------|
| | | | d_x | d_y | C_R | C_G | C_O | σ_{Neu} |
| G2 | A | 1.1 | 0.68 | 1.08 | 0.55 | 0.77 | 0.54 | 0.44 |
| | | 1.2 | 0.66 | 1.37 | 0.66 | 0.64 | 0.28 | 0.57 |
| H1 | A | 1.1 | 0.31 | 0.83 | 0.49 | 0.29 | 0.24 | 0.23 |
| | | 1.2 | 0.26 | 0.61 | 0.30 | 0.16 | 0.10 | 0.24 |
| | | 1.3 | 0.36 | 0.73 | 0.20 | 0.21 | 0.12 | 0.48 |
| H2 | A | 1.1 | 0.42 | 0.80 | 0.57 | 0.59 | 0.19 | 0.67 |
| | | 1.2 | 0.46 | 0.81 | 0.58 | 0.45 | -0.11 | 0.73 |
| | | 1.3 | 0.65 | 0.92 | 0.60 | 0.45 | -0.22 | 0.86 |
| H3 | A | 1.1 | 0.25 | 1.44 | 0.79 | 0.74 | 0.64 | 0.53 |
| | | 1.2 | 0.24 | 0.81 | 0.71 | 0.43 | 0.35 | 0.38 |
| | | 1.3 | 0.35 | 0.75 | 0.44 | 0.57 | 0.44 | 0.68 |
| I1 | S.A | 1.1 | 0.88 | 1.42 | 0.19 | 1.01 | 0.82 | 0.83 |
| | | 2.1 | 0.62 | 1.09 | 0.25 | 0.72 | 0.59 | 0.59 |
| | | 1.2 | 0.64 | 1.42 | 0.86 | 0.40 | 0.32 | 0.40 |
| | | 2.2 | 0.62 | 1.23 | 0.51 | 0.74 | 0.59 | 0.74 |
| I2 | S.A | 1.1 | 0.87 | 1.54 | 0.43 | 1.26 | 1.16 | 0.98 |
| | | 2.1 | 1.08 | 1.38 | -0.38 | 2.12 | 1.95 | 1.65 |
| | | 1.2 | 1.00 | 1.70 | 0.80 | 0.91 | 0.84 | 0.84 |
| | | 2.2 | 0.31 | 0.54 | 0.20 | 0.63 | 0.58 | 0.59 |
| J1 | S.A | 1.1 | 0.38 | 0.75 | 0.55 | 0.65 | 0.52 | 0.48 |
| | | 1.2 | 0.26 | 0.86 | 0.49 | 0.48 | 0.28 | 0.59 |
| | | 1.3 | 0.33 | 0.69 | 0.11 | 0.42 | 0.21 | 0.75 |
| | | 2.1 | 0.40 | 1.40 | 0.38 | 0.73 | 0.56 | 0.51 |
| | | 2.2 | 0.74 | 1.58 | 0.19 | 1.29 | 1.01 | 1.30 |
| | | 2.3 | 0.37 | 0.60 | 0.08 | 0.42 | 0.21 | 0.48 |

| Experiment | Effekt | Bedingung | Parameter | | | | | |
|------------|--------|-----------|-----------|-------|-------|-------|-------|----------------|
| | | | d_x | d_y | C_R | C_K | C_O | σ_{Neu} |
| SAC-Modell | | | | | | | | |
| A1 | S | 1 | 1.16 | 1.81 | 1.23 | 0.79 | 0.23 | 0.66 |
| | | 2 | 0.36 | 0.65 | 1.68 | 1.08 | 0.32 | 0.90 |
| A2 | S | 1 | 0.54 | -0.04 | 0.84 | 0.67 | 0.43 | 0.57 |
| | | 2 | 0.36 | 0.07 | 1.00 | 0.80 | 0.51 | 0.68 |
| A3 | S | 1 | 2.12 | 2.16 | 2.28 | 2.09 | 1.39 | 1.20 |
| | | 2 | 1.74 | 2.11 | 2.01 | 1.84 | 1.23 | 1.06 |
| A4 | S | 1 | 0.15 | -0.41 | 1.23 | 1.03 | 0.60 | 0.79 |
| | | 2 | -0.07 | -0.44 | 1.04 | 0.87 | 0.51 | 0.67 |
| A5 | S | 1 | 0.50 | -0.06 | 1.35 | 1.11 | 0.83 | 0.72 |
| | | 2 | -0.02 | -0.52 | 0.82 | 0.68 | 0.51 | 0.44 |
| A6 | S | 1 | 0.71 | 0.38 | 1.15 | 1.07 | 0.68 | 0.74 |
| | | 2 | 0.32 | 0.29 | 1.17 | 1.08 | 0.69 | 0.75 |
| A7 | S | 1 | 1.23 | 1.88 | 2.08 | 1.93 | 1.41 | 1.11 |
| | | 2 | 1.09 | 1.59 | 2.06 | 1.91 | 1.39 | 1.10 |
| B1 | S | 1 | 0.59 | 0.64 | 0.87 | 0.60 | 0.55 | 0.50 |
| | | 2 | 0.08 | -0.29 | 0.20 | 0.14 | 0.13 | 0.16 |
| B2 | S | 1 | 0.27 | 0.84 | 1.00 | 0.79 | 0.41 | 0.78 |
| | | 2 | 1.13 | 3.01 | 2.99 | 1.89 | 1.64 | 1.29 |
| B3 | S | 1 | 0.65 | 1.28 | 1.15 | 0.84 | 0.57 | 0.74 |
| | | 2 | 0.67 | 1.59 | 1.57 | 1.01 | 0.75 | 0.83 |
| B4 | S | 1 | 0.82 | 0.60 | 1.04 | 1.39 | 0.84 | 0.99 |
| | | 2 | 0.73 | 0.75 | 0.82 | 1.15 | 0.66 | 0.82 |
| B5 | S | 1 | 0.59 | 1.75 | 1.93 | 1.13 | 0.62 | 1.24 |
| | | 2 | 1.01 | 1.58 | 1.83 | 1.53 | 0.88 | 1.24 |
| B6 | S | 1 | 0.81 | 1.10 | 1.54 | 0.94 | 0.64 | 1.04 |
| | | 2 | 0.62 | 0.89 | 1.09 | 0.75 | 0.45 | 0.85 |
| B7 | S | 1 | 0.43 | 0.15 | 0.34 | 0.30 | 0.22 | 0.15 |
| | | 2 | -0.10 | 0.30 | 0.06 | 0.06 | 0.04 | 0.03 |
| B8 | S | 1 | -0.02 | -0.05 | 0.24 | 0.17 | 0.12 | 0.13 |
| | | 2 | 0.20 | -0.38 | 0.30 | 0.20 | 0.17 | 0.14 |
| B9 | S | 1 | 2.17 | 2.43 | 1.67 | 1.98 | 1.62 | 0.77 |
| | | 2 | 2.30 | 1.94 | 1.63 | 1.63 | 1.30 | 0.58 |
| C1 | S | 1 | 1.21 | 1.14 | 0.76 | 0.89 | 0.72 | 0.34 |
| | | 2 | 1.49 | 1.91 | 1.91 | 1.91 | 1.54 | 0.73 |
| | | 3 | 1.29 | 1.13 | 1.58 | 1.86 | 1.49 | 0.71 |
| C2 | S | 1 | 1.21 | 1.33 | 1.03 | 1.54 | 0.93 | 0.97 |
| | | 2 | 0.90 | 0.91 | 0.94 | 1.17 | 0.58 | 0.95 |
| | | 3 | 0.88 | 0.74 | 1.00 | 1.34 | 0.75 | 0.80 |
| C3 | S | 1 | 0.90 | 1.10 | 0.66 | 0.79 | 0.39 | 0.43 |
| | | 2 | 0.56 | 0.79 | 1.15 | 0.92 | -0.02 | 0.82 |
| | | 3 | 0.38 | 0.35 | 1.02 | 0.70 | -0.20 | 0.80 |

| Experiment | Effekt | Bedingung | Parameter | | | | | |
|------------|--------|-----------|-----------|-------|-------|-------|-------|----------------|
| | | | d_x | d_y | C_R | C_K | C_O | σ_{Neu} |
| D1 | S.S | 1.1 | 1.69 | 1.00 | 0.90 | 0.76 | 0.57 | 0.33 |
| | | 1.2 | 1.47 | 0.28 | 0.72 | 0.66 | 0.49 | 0.35 |
| | | 2.1 | 1.74 | 0.56 | 0.97 | 0.77 | 0.49 | 0.33 |
| | | 2.2 | 1.52 | 0.32 | 1.08 | 0.76 | 0.46 | 0.47 |
| D2 | S.S | 1.1 | 0.81 | 1.63 | 1.43 | 1.11 | 0.81 | 0.76 |
| | | 1.2 | 1.04 | 1.63 | 1.98 | 1.49 | 1.32 | 0.85 |
| | | 2.1 | 0.42 | 0.68 | 0.85 | 0.75 | 0.57 | 0.55 |
| | | 2.2 | 0.36 | 1.04 | 1.59 | 0.85 | 0.70 | 0.55 |
| D3 | S.S | 1.1 | 0.54 | 0.98 | 1.72 | 1.14 | 0.08 | 0.99 |
| | | 1.2 | 1.07 | 1.20 | 1.53 | 0.97 | 0.07 | 0.88 |
| | | 2.1 | 0.71 | 0.71 | 1.52 | 0.91 | 0.16 | 0.87 |
| | | 2.2 | 1.27 | 1.46 | 1.59 | 1.11 | 0.36 | 1.02 |
| D4 | S.S | 1.1 | 0.32 | 0.62 | 1.15 | 0.80 | 0.36 | 0.61 |
| | | 1.2 | 0.67 | 1.43 | 2.08 | 1.24 | 0.81 | 1.02 |
| | | 2.1 | 0.75 | 1.35 | 1.58 | 1.03 | 0.64 | 0.78 |
| | | 2.2 | 0.57 | 1.23 | 1.56 | 0.96 | 0.59 | 0.83 |
| D5 | S.S | 1.1 | 1.30 | 0.89 | 1.04 | 1.12 | 1.03 | 0.98 |
| | | 1.2 | 0.21 | 0.10 | 0.52 | 0.48 | 0.40 | 0.64 |
| | | 2.1 | -0.39 | -0.59 | 0.06 | 0.03 | -0.01 | 0.22 |
| | | 2.2 | -0.59 | -0.57 | 0.07 | -0.06 | -0.13 | 0.35 |
| E1 | S.S | 1.1 | 0.65 | 1.23 | 1.84 | 1.31 | 0.98 | 0.66 |
| | | 1.2 | 0.14 | 0.38 | 1.12 | 0.80 | 0.60 | 0.40 |
| | | 2.1 | 0.75 | 1.21 | 1.15 | 0.86 | 0.59 | 0.57 |
| | | 2.2 | 0.61 | 1.06 | 1.43 | 1.07 | 0.73 | 0.70 |
| E2 | S.S | 1.1 | 1.16 | 1.82 | 2.04 | 1.39 | 1.21 | 1.10 |
| | | 1.2 | 1.28 | 2.20 | 2.28 | 1.55 | 1.35 | 1.22 |
| | | 2.1 | 1.28 | 1.91 | 1.86 | 1.33 | 1.06 | 1.00 |
| | | 2.2 | 0.64 | 1.13 | 1.28 | 0.92 | 0.73 | 0.69 |
| E3 | S.S | 1.1 | 1.11 | 1.64 | 1.54 | 1.28 | 1.14 | 1.08 |
| | | 1.2 | 0.29 | 0.36 | 0.51 | 0.42 | 0.37 | 0.36 |
| | | 2.1 | 0.60 | 1.02 | 1.20 | 0.98 | 0.83 | 1.14 |
| | | 2.2 | 0.30 | 0.38 | 0.71 | 0.58 | 0.50 | 0.68 |
| E4 | S.S | 1.1 | 3.20 | 3.20 | 3.27 | 2.87 | 1.97 | 1.87 |
| | | 1.2 | 0.64 | 0.71 | 1.32 | 1.16 | 0.80 | 0.76 |
| | | 2.1 | 1.46 | 1.56 | 1.18 | 1.14 | 0.82 | 0.84 |
| | | 2.2 | 0.84 | 1.03 | 0.98 | 0.95 | 0.69 | 0.70 |
| E5 | S.S | 1.1 | -0.02 | 1.20 | 1.38 | 0.98 | 0.73 | 0.60 |
| | | 1.2 | 0.84 | 2.13 | 2.71 | 1.93 | 1.44 | 1.17 |
| | | 2.1 | 0.75 | 1.74 | 1.51 | 1.08 | 0.90 | 0.74 |
| | | 2.2 | 0.92 | 1.51 | 1.98 | 1.42 | 1.18 | 0.97 |

| Experiment | Effekt | Bedingung | Parameter | | | | | |
|------------|--------|-----------|-----------|-------|-------|-------|-------|----------------|
| | | | d_x | d_y | C_R | C_K | C_O | σ_{Neu} |
| E6 | S.S | 1.1 | 0.14 | 0.65 | 1.53 | 1.08 | 0.94 | 0.81 |
| | | 1.2 | 0.32 | 0.51 | 1.91 | 1.35 | 1.18 | 1.02 |
| | | 2.1 | 0.17 | 0.31 | 1.02 | 0.69 | 0.58 | 0.50 |
| | | 2.2 | 0.32 | 0.49 | 1.62 | 1.10 | 0.92 | 0.79 |
| E7 | S.S | 1.1 | 3.03 | 3.54 | 3.39 | 3.38 | 2.97 | 1.48 |
| | | 1.2 | 1.60 | 1.93 | 3.01 | 3.00 | 2.64 | 1.32 |
| | | 2.1 | 1.01 | 1.91 | 1.26 | 1.20 | 0.81 | 0.53 |
| | | 2.2 | 0.62 | 0.92 | 1.88 | 1.79 | 1.21 | 0.79 |
| E8 | S.S | 1.1 | 2.13 | 3.12 | 3.17 | 2.64 | 2.28 | 1.82 |
| | | 1.2 | 1.75 | 2.57 | 3.12 | 2.60 | 2.25 | 1.79 |
| | | 2.1 | 1.66 | 1.64 | 1.31 | 1.05 | 0.76 | 0.79 |
| | | 2.2 | 1.33 | 1.57 | 1.80 | 1.44 | 1.05 | 1.09 |
| E9 | S.S | 1.1 | 2.13 | 2.92 | 3.17 | 1.96 | 1.88 | 1.54 |
| | | 1.2 | 0.24 | 0.22 | 1.26 | 0.78 | 0.74 | 0.61 |
| | | 2.1 | 1.34 | 2.35 | 2.21 | 1.31 | 1.25 | 1.08 |
| | | 2.2 | 0.23 | 0.49 | 1.44 | 0.85 | 0.82 | 0.70 |
| E10 | S.S | 1.1 | 0.40 | 0.97 | 1.33 | 1.18 | 0.76 | 0.57 |
| | | 1.2 | 0.52 | 1.07 | 1.63 | 1.44 | 0.93 | 0.70 |
| | | 2.1 | 0.79 | 0.92 | 1.56 | 1.44 | 1.17 | 0.89 |
| | | 2.2 | 0.72 | 1.05 | 1.75 | 1.62 | 1.32 | 1.00 |
| E11 | S.S | 1.1 | 1.34 | 1.24 | 1.56 | 1.28 | 1.05 | 0.67 |
| | | 1.2 | 2.63 | 2.80 | 3.27 | 2.69 | 2.21 | 1.41 |
| | | 2.1 | 1.65 | 1.68 | 1.97 | 1.62 | 1.33 | 0.85 |
| | | 2.2 | 1.85 | 1.81 | 1.78 | 1.47 | 1.20 | 0.77 |
| E12 | S.S | 1.1 | 4.90 | 4.81 | 4.29 | 4.09 | 3.57 | 2.19 |
| | | 1.2 | 3.15 | 3.10 | 2.71 | 2.59 | 2.26 | 1.38 |
| | | 2.1 | 2.33 | 2.13 | 2.41 | 2.30 | 2.01 | 1.23 |
| | | 2.2 | 1.70 | 1.54 | 1.70 | 1.62 | 1.41 | 0.87 |
| F1 | S.S | 1.1 | 1.06 | 1.50 | 1.62 | 1.36 | 0.99 | 0.99 |
| | | 1.2 | 1.02 | 1.22 | 1.84 | 1.32 | 0.97 | 1.05 |
| | | 1.3 | 1.00 | 1.56 | 2.14 | 1.43 | 1.03 | 0.92 |
| | | 2.1 | 2.42 | 3.68 | 2.84 | 2.29 | 1.89 | 1.22 |
| | | 2.2 | 0.75 | 1.67 | 1.67 | 0.55 | 0.05 | 0.95 |
| | | 2.3 | 1.64 | 2.03 | 2.26 | 1.61 | 1.22 | 0.80 |
| F2 | S.S | 1.1 | 0.74 | 1.97 | 1.71 | 1.25 | 0.99 | 0.63 |
| | | 1.2 | 0.58 | 1.57 | 1.10 | 0.97 | 0.74 | 0.48 |
| | | 1.3 | 0.67 | 1.62 | 1.04 | 1.04 | 0.75 | 0.51 |
| | | 2.1 | 0.97 | 2.22 | 2.34 | 1.47 | 1.18 | 0.87 |
| | | 2.2 | 1.64 | 3.10 | 2.46 | 1.66 | 1.35 | 1.07 |
| | | 2.3 | 0.38 | 1.11 | 0.40 | 0.40 | 0.29 | 0.20 |
| G1 | A | 1.1 | 0.39 | 1.00 | 1.18 | 0.87 | 0.41 | 1.00 |
| | | 1.2 | 0.80 | 1.70 | 1.90 | 1.39 | 0.98 | 1.16 |

| Experiment | Effekt | Bedingung | Parameter | | | | | |
|------------|--------|-----------|-----------|-------|-------|-------|-------|----------------|
| | | | d_x | d_y | C_R | C_K | C_O | σ_{Neu} |
| G2 | A | 1.1 | 0.71 | 1.00 | 1.10 | 1.10 | 0.71 | 0.54 |
| | | 1.2 | 0.76 | 1.47 | 1.40 | 1.05 | 0.46 | 0.80 |
| H1 | A | 1.1 | 0.05 | 0.52 | 0.72 | 0.50 | 0.40 | 0.36 |
| | | 1.2 | -0.01 | 0.19 | 0.48 | 0.26 | 0.18 | 0.32 |
| | | 1.3 | 0.27 | 0.50 | 0.78 | 0.37 | 0.24 | 0.58 |
| H2 | A | 1.1 | 0.42 | 0.79 | 1.04 | 1.04 | 0.37 | 0.81 |
| | | 1.2 | 0.47 | 0.77 | 0.87 | 0.90 | 0.04 | 0.84 |
| | | 1.3 | 0.77 | 0.91 | 0.83 | 0.96 | -0.03 | 0.91 |
| H3 | A | 1.1 | 0.58 | 2.21 | 2.69 | 1.72 | 1.48 | 1.16 |
| | | 1.2 | -0.03 | 0.55 | 0.83 | 0.81 | 0.64 | 0.55 |
| | | 1.3 | 0.35 | 0.74 | 1.12 | 0.97 | 0.74 | 0.83 |
| I1 | S.A | 1.1 | 1.29 | 1.60 | 1.87 | 1.33 | 1.07 | 0.97 |
| | | 2.1 | 0.80 | 1.12 | 1.42 | 1.01 | 0.82 | 0.74 |
| | | 1.2 | 0.51 | 1.32 | 0.99 | 0.74 | 0.59 | 0.61 |
| | | 2.2 | 0.82 | 1.45 | 1.58 | 1.19 | 0.94 | 0.98 |
| I2 | S.A | 1.1 | 1.45 | 2.14 | 2.39 | 1.88 | 1.72 | 1.28 |
| | | 2.1 | 1.70 | 1.83 | 3.01 | 2.38 | 2.17 | 1.62 |
| | | 1.2 | 1.34 | 1.99 | 1.68 | 1.46 | 1.34 | 1.07 |
| | | 2.2 | 0.26 | 0.40 | 1.02 | 0.89 | 0.82 | 0.65 |
| J1 | S.A | 1.1 | 0.26 | 0.61 | 0.96 | 0.99 | 0.76 | 0.58 |
| | | 1.2 | 0.18 | 0.85 | 1.23 | 0.87 | 0.51 | 0.83 |
| | | 1.3 | 0.37 | 0.68 | 1.18 | 0.69 | 0.39 | 0.88 |
| | | 2.1 | 0.59 | 1.79 | 2.23 | 1.20 | 0.91 | 0.83 |
| | | 2.2 | 1.36 | 2.33 | 2.86 | 1.85 | 1.43 | 1.62 |
| | | 2.3 | 0.33 | 0.38 | 0.88 | 0.61 | 0.32 | 0.56 |

Anhang D

Datenanpassung hypothesekonformer Modellvarianten

Im Folgenden werden die Datenanpassungsindizes berichtet, die bei der Anpassung hypothesekonform spezifizierter Modellvarianten an die Daten der einzelnen Experimente erreicht wurden. Neben der logarithmierten Likelihood (\mathcal{LL}) und dem Bayesianischen Informationskriterium (BIC) werden G^2 -Werte mit Freiheitsgraden (df), AIC_c -Werte und AIC_c -Differenzwerte (ΔAIC_c) berichtet. In der Tabelle wird gesondert festgehalten, für welches Modell und Experiment die jeweiligen Datenanpassungsindizes ermittelt wurden.

| Experiment | Modell | \mathcal{LL} | BIC | G^2 | df | AIC_c | ΔAIC_c |
|------------|----------|----------------|---------|-------|------|---------|----------------|
| A1 | SDM | -664.27 | 1373.76 | 2.21 | 5 | 1342.71 | 0.00 |
| | HTSD | -664.99 | 1375.21 | 3.66 | 5 | 1344.16 | 1.45 |
| | HTUVSD | -664.41 | 1386.97 | 2.50 | 3 | 1347.10 | 4.39 |
| | SAC | -663.59 | 1385.34 | 0.87 | 3 | 1345.47 | 2.76 |
| | Streak | -676.70 | 1411.54 | 27.07 | 3 | 1371.68 | 28.97 |
| | 4-States | -663.68 | 1385.51 | 1.04 | 3 | 1345.65 | 2.94 |
| A2 | SDM | -2064.68 | 4182.28 | 25.87 | 5 | 4143.42 | 21.75 |
| | HTSD | -2068.70 | 4190.31 | 33.90 | 5 | 4151.45 | 29.78 |
| | HTUVSD | -2065.39 | 4198.82 | 27.28 | 3 | 4148.87 | 27.20 |
| | SAC | -2052.08 | 4172.20 | 0.67 | 3 | 4122.26 | 0.59 |
| | Streak | -2052.45 | 4172.94 | 1.41 | 3 | 4122.99 | 1.32 |
| | 4-States | -2051.79 | 4171.62 | 0.08 | 3 | 4121.67 | 0.00 |
| A3 | SDM | -1609.83 | 3272.13 | 3.51 | 5 | 3233.72 | 0.00 |
| | HTSD | -1615.11 | 3282.68 | 14.07 | 5 | 3244.28 | 10.56 |
| | HTUVSD | -1610.49 | 3288.43 | 4.83 | 3 | 3239.07 | 5.36 |
| | SAC | -1608.35 | 3284.15 | 0.55 | 3 | 3234.79 | 1.07 |
| | Streak | -1612.97 | 3293.40 | 9.80 | 3 | 3244.04 | 10.32 |
| | 4-States | -1608.88 | 3285.22 | 1.61 | 3 | 3235.86 | 2.14 |
| A4 | SDM | -2158.00 | 4370.20 | 16.25 | 5 | 4330.05 | 11.20 |
| | HTSD | -2161.89 | 4377.98 | 24.04 | 5 | 4337.84 | 18.98 |
| | HTUVSD | -2159.79 | 4389.26 | 19.83 | 3 | 4337.65 | 18.80 |
| | SAC | -2150.39 | 4370.46 | 1.02 | 3 | 4318.85 | 0.00 |
| | Streak | -2152.13 | 4373.94 | 4.51 | 3 | 4322.34 | 3.49 |
| | 4-States | -2151.35 | 4372.39 | 2.96 | 3 | 4320.79 | 1.94 |
| A5 | SDM | -1872.37 | 3798.94 | 25.43 | 5 | 3758.80 | 16.25 |
| | HTSD | -1877.14 | 3808.47 | 34.96 | 5 | 3768.32 | 25.78 |
| | HTUVSD | -1874.52 | 3818.72 | 29.72 | 3 | 3767.11 | 24.57 |
| | SAC | -1862.23 | 3794.14 | 5.15 | 3 | 3742.54 | 0.00 |
| | Streak | -1863.32 | 3796.32 | 7.32 | 3 | 3744.72 | 2.17 |
| | 4-States | -1863.71 | 3797.11 | 8.11 | 3 | 3745.51 | 2.97 |
| A6 | SDM | -2366.78 | 4787.76 | 17.47 | 5 | 4747.61 | 13.43 |
| | HTSD | -2370.21 | 4794.61 | 24.31 | 5 | 4754.46 | 20.28 |
| | HTUVSD | -2367.77 | 4805.23 | 19.45 | 3 | 4753.63 | 19.45 |
| | SAC | -2358.05 | 4785.79 | 0.01 | 3 | 4734.18 | 0.00 |
| | Streak | -2359.41 | 4788.50 | 2.72 | 3 | 4736.90 | 2.72 |
| | 4-States | -2360.20 | 4790.09 | 4.31 | 3 | 4738.48 | 4.30 |
| A7 | SDM | -2049.19 | 4152.57 | 0.07 | 5 | 4112.43 | 0.00 |
| | HTSD | -2049.28 | 4152.76 | 0.26 | 5 | 4112.62 | 0.19 |
| | HTUVSD | -2049.21 | 4168.09 | 0.11 | 3 | 4116.49 | 4.06 |
| | SAC | -2049.16 | 4167.99 | 0.00 | 3 | 4116.39 | 3.96 |
| | Streak | -2050.29 | 4170.27 | 2.28 | 3 | 4118.67 | 6.24 |
| | 4-States | -2049.44 | 4168.56 | 0.57 | 3 | 4116.96 | 4.53 |

| Experiment | Modell | \mathcal{LL} | BIC | G^2 | df | AIC_c | ΔAIC_c |
|------------|----------|----------------|----------|--------|------|----------|----------------|
| B1 | SDM | -2126.26 | 4306.70 | 43.75 | 5 | 4266.57 | 31.76 |
| | HTSD | -2127.73 | 4309.65 | 46.69 | 5 | 4269.52 | 34.71 |
| | HTUVSD | -2114.76 | 4299.18 | 20.74 | 3 | 4247.60 | 12.79 |
| | SAC | -2108.37 | 4286.39 | 7.96 | 3 | 4234.81 | 0.00 |
| | Streak | -2111.31 | 4292.28 | 13.85 | 3 | 4240.70 | 5.89 |
| | 4-States | -2112.39 | 4294.44 | 16.01 | 3 | 4242.86 | 8.05 |
| B2 | SDM | -2354.25 | 4762.98 | 133.76 | 5 | 4722.54 | 103.94 |
| | HTSD | -2355.93 | 4766.35 | 137.13 | 5 | 4725.91 | 107.31 |
| | HTUVSD | -2315.74 | 4701.53 | 56.75 | 3 | 4649.56 | 30.95 |
| | SAC | -2313.24 | 4696.53 | 51.75 | 3 | 4644.56 | 25.95 |
| | Streak | -2323.98 | 4718.02 | 73.23 | 3 | 4666.04 | 47.43 |
| | 4-States | -2300.27 | 4670.58 | 25.79 | 3 | 4618.61 | 0.00 |
| B3 | SDM | -3873.37 | 7804.80 | 18.70 | 5 | 7760.77 | 10.51 |
| | HTSD | -3876.43 | 7810.91 | 24.81 | 5 | 7766.88 | 16.62 |
| | HTUVSD | -3867.67 | 7809.99 | 7.29 | 3 | 7753.38 | 3.12 |
| | SAC | -3866.75 | 7808.15 | 5.46 | 3 | 7751.55 | 1.29 |
| | Streak | -3866.11 | 7806.86 | 4.17 | 3 | 7750.26 | 0.00 |
| | 4-States | -3867.30 | 7809.25 | 6.55 | 3 | 7752.64 | 2.38 |
| B4 | SDM | -6501.28 | 13063.17 | 30.86 | 5 | 13016.58 | 26.55 |
| | HTSD | -6524.04 | 13108.69 | 76.38 | 5 | 13062.10 | 72.07 |
| | HTUVSD | -6502.23 | 13082.40 | 32.77 | 3 | 13022.50 | 32.48 |
| | SAC | -6487.11 | 13052.15 | 2.53 | 3 | 12992.26 | 2.23 |
| | Streak | -6496.26 | 13070.44 | 20.82 | 3 | 13010.54 | 20.52 |
| | 4-States | -6486.00 | 13049.92 | 0.30 | 3 | 12990.02 | 0.00 |
| B5 | SDM | -3250.07 | 6555.89 | 24.72 | 5 | 6514.17 | 13.65 |
| | HTSD | -3248.46 | 6552.69 | 21.51 | 5 | 6510.97 | 10.45 |
| | HTUVSD | -3247.23 | 6566.16 | 19.05 | 3 | 6512.53 | 12.01 |
| | SAC | -3245.51 | 6562.70 | 15.60 | 3 | 6509.07 | 8.56 |
| | Streak | -3241.23 | 6554.14 | 7.04 | 3 | 6500.52 | 0.00 |
| | 4-States | -3245.08 | 6561.86 | 14.75 | 3 | 6508.23 | 7.71 |
| B6 | SDM | -4091.42 | 8240.16 | 10.97 | 5 | 8196.87 | 3.42 |
| | HTSD | -4091.68 | 8240.67 | 11.49 | 5 | 8197.38 | 3.93 |
| | HTUVSD | -4089.70 | 8253.10 | 7.53 | 3 | 8197.45 | 4.00 |
| | SAC | -4088.38 | 8250.46 | 4.90 | 3 | 8194.81 | 1.36 |
| | Streak | -4093.55 | 8260.81 | 15.24 | 3 | 8205.16 | 11.71 |
| | 4-States | -4087.70 | 8249.10 | 3.54 | 3 | 8193.45 | 0.00 |
| B7 | SDM | -2878.05 | 5813.95 | 155.90 | 5 | 5770.12 | 142.38 |
| | HTSD | -2887.41 | 5832.68 | 174.62 | 5 | 5788.85 | 161.10 |
| | HTUVSD | -2881.62 | 5837.63 | 163.04 | 3 | 5781.28 | 153.54 |
| | SAC | -2804.85 | 5684.09 | 9.50 | 3 | 5627.74 | 0.00 |
| | Streak | -2838.60 | 5751.59 | 77.01 | 3 | 5695.25 | 67.50 |
| | 4-States | -2806.98 | 5688.35 | 13.76 | 3 | 5632.00 | 4.26 |

| Experiment | Modell | \mathcal{L} | BIC | G^2 | df | AIC_c | ΔAIC_c |
|------------|----------|---------------|----------|---------|------|----------|----------------|
| B8 | SDM | -1430.74 | 2913.12 | 58.42 | 5 | 2875.54 | 48.54 |
| | HTSD | -1432.85 | 2917.35 | 62.66 | 5 | 2879.78 | 52.77 |
| | HTUVSD | -1431.27 | 2928.93 | 59.49 | 3 | 2880.65 | 53.64 |
| | SAC | -1404.45 | 2875.29 | 5.84 | 3 | 2827.00 | 0.00 |
| | Streak | -1414.04 | 2894.48 | 25.03 | 3 | 2846.19 | 19.19 |
| | 4-States | -1407.10 | 2880.59 | 11.14 | 3 | 2832.30 | 5.30 |
| B9 | SDM | -788.33 | 1628.31 | 14.24 | 5 | 1590.73 | 8.83 |
| | HTSD | -797.02 | 1645.69 | 31.62 | 5 | 1608.12 | 26.21 |
| | HTUVSD | -786.36 | 1639.12 | 10.29 | 3 | 1590.83 | 8.92 |
| | SAC | -781.90 | 1630.19 | 1.37 | 3 | 1581.91 | 0.00 |
| | Streak | -796.91 | 1660.22 | 31.40 | 3 | 1611.94 | 30.03 |
| | 4-States | -782.22 | 1630.83 | 2.01 | 3 | 1582.55 | 0.64 |
| C1 | SDM | -858.67 | 1782.09 | 14 | 9 | 1735.47 | 6.63 |
| | HTSD | -861.17 | 1787.08 | 18.99 | 9 | 1740.47 | 11.62 |
| | HTUVSD | -857.94 | 1802.21 | 12.54 | 6 | 1740.12 | 11.27 |
| | SAC | -853.69 | 1793.71 | 4.03 | 6 | 1731.61 | 2.77 |
| | Streak | -855.85 | 1798.03 | 8.35 | 6 | 1735.93 | 7.09 |
| | 4-States | -852.30 | 1790.94 | 1.27 | 6 | 1728.84 | 0.00 |
| C2 | SDM | -3738.84 | 7551.02 | 53.55 | 9 | 7495.74 | 39.73 |
| | HTSD | -3753.41 | 7580.16 | 82.69 | 9 | 7524.88 | 68.87 |
| | HTUVSD | -3734.70 | 7567.19 | 45.27 | 6 | 7493.50 | 37.49 |
| | SAC | -3715.96 | 7529.69 | 7.78 | 6 | 7456.01 | 0.00 |
| | Streak | -3730.48 | 7558.73 | 36.81 | 6 | 7485.04 | 29.04 |
| | 4-States | -3716.67 | 7531.11 | 9.19 | 6 | 7457.42 | 1.42 |
| C3 | SDM | -15711.76 | 31508.92 | 1101.19 | 9 | 31441.54 | 973.39 |
| | HTSD | -15743.07 | 31571.53 | 1163.81 | 9 | 31504.16 | 1036.01 |
| | HTUVSD | -15738.28 | 31590.41 | 1154.22 | 6 | 31500.58 | 1032.43 |
| | SAC | -15603.95 | 31321.76 | 885.57 | 6 | 31231.93 | 763.78 |
| | Streak | -15835.75 | 31785.35 | 1349.15 | 6 | 31695.51 | 1227.36 |
| | 4-States | -15222.06 | 30557.98 | 121.79 | 6 | 30468.15 | 0.00 |
| D1 | SDM | -19442.15 | 38994.86 | 696.87 | 13 | 38906.31 | 591.75 |
| | HTSD | -19473.95 | 39058.46 | 760.46 | 13 | 38969.91 | 655.35 |
| | HTUVSD | -19373.34 | 38897.46 | 559.26 | 9 | 38776.71 | 462.15 |
| | SAC | -19149.00 | 38448.76 | 110.56 | 9 | 38328.01 | 13.46 |
| | Streak | -19532.80 | 39216.36 | 878.16 | 9 | 39095.61 | 781.05 |
| | 4-States | -19142.27 | 38435.31 | 97.10 | 9 | 38314.56 | 0.00 |
| D2 | SDM | -3411.73 | 6914.71 | 91.18 | 13 | 6845.53 | 52.50 |
| | HTSD | -3414.54 | 6920.33 | 96.80 | 13 | 6851.15 | 58.13 |
| | HTUVSD | -3383.55 | 6891.52 | 34.81 | 9 | 6797.21 | 4.19 |
| | SAC | -3381.45 | 6887.33 | 30.62 | 9 | 6793.03 | 0.00 |
| | Streak | -3391.64 | 6907.71 | 51.00 | 9 | 6813.41 | 20.38 |
| | 4-States | -3384.15 | 6892.73 | 36.02 | 9 | 6798.43 | 5.40 |

| Experiment | Modell | \mathcal{LL} | BIC | G^2 | df | AIC_c | ΔAIC_c |
|------------|----------|----------------|----------|--------|------|----------|----------------|
| D3 | SDM | -2607.71 | 5299.88 | 28.43 | 13 | 5237.55 | 8.07 |
| | HTSD | -2606.88 | 5298.22 | 26.77 | 13 | 5235.89 | 6.41 |
| | HTUVSD | -2604.23 | 5323.62 | 21.46 | 9 | 5238.68 | 9.20 |
| | SAC | -2602.09 | 5319.36 | 17.19 | 9 | 5234.41 | 4.93 |
| | Streak | -2606.76 | 5328.68 | 26.51 | 9 | 5243.74 | 14.25 |
| | 4-States | -2599.63 | 5314.43 | 12.26 | 9 | 5229.48 | 0.00 |
| D4 | SDM | -4834.19 | 9761.17 | 29.93 | 13 | 9690.43 | 6.64 |
| | HTSD | -4841.00 | 9774.79 | 43.55 | 13 | 9704.05 | 20.26 |
| | HTUVSD | -4832.98 | 9792.49 | 27.51 | 9 | 9696.06 | 12.27 |
| | SAC | -4831.15 | 9788.83 | 23.85 | 9 | 9692.40 | 8.61 |
| | Streak | -4834.02 | 9794.57 | 29.58 | 9 | 9698.14 | 14.34 |
| | 4-States | -4826.85 | 9780.23 | 15.25 | 9 | 9683.80 | 0.00 |
| D5 | SDM | -6087.75 | 12269.99 | 145.38 | 13 | 12197.56 | 89.16 |
| | HTSD | -6097.30 | 12289.08 | 164.46 | 13 | 12216.64 | 108.24 |
| | HTUVSD | -6084.71 | 12298.26 | 139.28 | 9 | 12199.50 | 91.10 |
| | SAC | -6039.16 | 12207.16 | 48.18 | 9 | 12108.40 | 0.00 |
| | Streak | -6057.17 | 12243.19 | 84.21 | 9 | 12144.43 | 36.03 |
| | 4-States | -6065.81 | 12260.46 | 101.48 | 9 | 12161.70 | 53.30 |
| E1 | SDM | -1301.73 | 2684.23 | 24.09 | 13 | 2625.64 | 12.20 |
| | HTSD | -1304.47 | 2689.71 | 29.58 | 13 | 2631.12 | 17.69 |
| | HTUVSD | -1294.33 | 2698.79 | 9.29 | 9 | 2618.97 | 5.53 |
| | SAC | -1291.56 | 2693.25 | 3.75 | 9 | 2613.44 | 0.00 |
| | Streak | -1294.80 | 2699.73 | 10.23 | 9 | 2619.91 | 6.48 |
| | 4-States | -1292.95 | 2696.04 | 6.54 | 9 | 2616.22 | 2.79 |
| E2 | SDM | -1377.52 | 2835.75 | 3.17 | 13 | 2777.22 | 0.00 |
| | HTSD | -1377.85 | 2836.41 | 3.83 | 13 | 2777.88 | 0.66 |
| | HTUVSD | -1377.66 | 2865.38 | 3.45 | 9 | 2785.64 | 8.42 |
| | SAC | -1377.58 | 2865.22 | 3.29 | 9 | 2785.48 | 8.26 |
| | Streak | -1378.31 | 2866.68 | 4.75 | 9 | 2786.94 | 9.72 |
| | 4-States | -1378.08 | 2866.21 | 4.29 | 9 | 2786.48 | 9.26 |
| E3 | SDM | -1528.70 | 3138.11 | 22.02 | 13 | 3079.58 | 10.42 |
| | HTSD | -1528.85 | 3138.41 | 22.31 | 13 | 3079.87 | 10.71 |
| | HTUVSD | -1523.62 | 3157.30 | 11.86 | 9 | 3077.57 | 8.41 |
| | SAC | -1519.42 | 3148.90 | 3.45 | 9 | 3069.16 | 0.00 |
| | Streak | -1522.46 | 3154.98 | 9.53 | 9 | 3075.24 | 6.08 |
| | 4-States | -1519.46 | 3148.97 | 3.52 | 9 | 3069.23 | 0.07 |
| E4 | SDM | -2588.36 | 5263.95 | 36.26 | 13 | 5198.82 | 0.00 |
| | HTSD | -2613.11 | 5313.45 | 85.76 | 13 | 5248.32 | 49.50 |
| | HTUVSD | -2590.26 | 5299.48 | 40.07 | 9 | 5210.70 | 11.88 |
| | SAC | -2584.99 | 5288.92 | 29.51 | 9 | 5200.14 | 1.33 |
| | Streak | -2601.52 | 5322.00 | 62.59 | 9 | 5233.22 | 34.40 |
| | 4-States | -2599.99 | 5318.93 | 59.52 | 9 | 5230.15 | 31.33 |

| Experiment | Modell | $\mathcal{L}\mathcal{L}$ | BIC | G^2 | df | AIC_c | ΔAIC_c |
|------------|----------|--------------------------|----------|--------|------|---------|----------------|
| E5 | SDM | -2499.29 | 5086.92 | 28.54 | 13 | 5020.67 | 5.71 |
| | HTSD | -2498.67 | 5085.67 | 27.29 | 13 | 5019.43 | 4.46 |
| | HTUVSD | -2496.34 | 5113.14 | 22.64 | 9 | 5022.85 | 7.88 |
| | SAC | -2494.96 | 5110.38 | 19.88 | 9 | 5020.09 | 5.12 |
| | Streak | -2511.18 | 5142.81 | 52.31 | 9 | 5052.51 | 37.55 |
| | 4-States | -2492.41 | 5105.26 | 14.76 | 9 | 5014.97 | 0.00 |
| E6 | SDM | -2455.23 | 4999.19 | 6.78 | 13 | 4932.55 | 0.00 |
| | HTSD | -2456.42 | 5001.57 | 9.16 | 13 | 4934.93 | 2.38 |
| | HTUVSD | -2454.81 | 5030.61 | 5.93 | 9 | 4939.77 | 7.22 |
| | SAC | -2453.71 | 5028.41 | 3.73 | 9 | 4937.57 | 5.02 |
| | Streak | -2461.13 | 5043.24 | 18.57 | 9 | 4952.40 | 19.85 |
| | 4-States | -2455.01 | 5031.01 | 6.33 | 9 | 4940.17 | 7.62 |
| E7 | SDM | -3202.33 | 6496.75 | 38.40 | 13 | 6426.73 | 3.54 |
| | HTSD | -3203.69 | 6499.45 | 41.11 | 13 | 6429.43 | 6.24 |
| | HTUVSD | -3202.13 | 6529.82 | 38.00 | 9 | 6434.37 | 11.18 |
| | SAC | -3196.54 | 6518.64 | 26.82 | 9 | 6423.19 | 0.00 |
| | Streak | -3233.56 | 6592.69 | 100.86 | 9 | 6497.24 | 74.05 |
| | 4-States | -3216.61 | 6558.78 | 66.95 | 9 | 6463.32 | 40.14 |
| E8 | SDM | -2312.52 | 4711.39 | 20.05 | 13 | 4647.15 | 0.68 |
| | HTSD | -2318.73 | 4723.81 | 32.47 | 13 | 4659.57 | 13.11 |
| | HTUVSD | -2310.11 | 4737.96 | 15.22 | 9 | 4650.40 | 3.94 |
| | SAC | -2308.14 | 4734.02 | 11.28 | 9 | 4646.47 | 0.00 |
| | Streak | -2314.59 | 4746.91 | 24.18 | 9 | 4659.36 | 12.90 |
| | 4-States | -2310.52 | 4738.77 | 16.03 | 9 | 4651.22 | 4.75 |
| E9 | SDM | -4899.14 | 9894.69 | 4.64 | 13 | 9820.32 | 0.00 |
| | HTSD | -4899.93 | 9896.28 | 6.22 | 13 | 9821.90 | 1.58 |
| | HTUVSD | -4899.78 | 9931.05 | 5.93 | 9 | 9829.64 | 9.32 |
| | SAC | -4900.22 | 9931.92 | 6.80 | 9 | 9830.52 | 10.20 |
| | Streak | -4961.49 | 10054.46 | 129.34 | 9 | 9953.05 | 132.74 |
| | 4-States | -4906.67 | 9944.81 | 19.69 | 9 | 9843.41 | 23.09 |
| E10 | SDM | -3310.42 | 6711.98 | 64.34 | 13 | 6642.90 | 22.99 |
| | HTSD | -3311.86 | 6714.86 | 67.22 | 13 | 6645.79 | 25.87 |
| | HTUVSD | -3298.59 | 6721.46 | 40.68 | 9 | 6627.29 | 7.38 |
| | SAC | -3294.89 | 6714.08 | 33.29 | 9 | 6619.91 | 0.00 |
| | Streak | -3298.51 | 6721.30 | 40.52 | 9 | 6627.14 | 7.23 |
| | 4-States | -3303.11 | 6730.51 | 49.73 | 9 | 6636.34 | 16.43 |
| E11 | SDM | -2084.44 | 4252.37 | 13.06 | 13 | 4191.01 | 0.00 |
| | HTSD | -2085.85 | 4255.21 | 15.90 | 13 | 4193.84 | 2.83 |
| | HTUVSD | -2085.27 | 4284.41 | 14.74 | 9 | 4200.79 | 9.78 |
| | SAC | -2084.19 | 4282.25 | 12.58 | 9 | 4198.63 | 7.62 |
| | Streak | -2089.44 | 4292.74 | 23.07 | 9 | 4209.12 | 18.12 |
| | 4-States | -2082.90 | 4279.67 | 10.00 | 9 | 4196.05 | 5.05 |

| Experiment | Modell | \mathcal{LL} | BIC | G^2 | df | AIC_c | ΔAIC_c |
|------------|----------|----------------|----------|--------|------|----------|----------------|
| E12 | SDM | -1804.87 | 3693.36 | 4.71 | 13 | 3631.88 | 0.00 |
| | HTSD | -1812.65 | 3708.91 | 20.27 | 13 | 3647.43 | 15.55 |
| | HTUVSD | -1806.07 | 3726.16 | 7.11 | 9 | 3642.39 | 10.51 |
| | SAC | -1806.14 | 3726.29 | 7.24 | 9 | 3642.52 | 10.64 |
| | Streak | -1825.43 | 3764.86 | 45.81 | 9 | 3681.09 | 49.21 |
| | 4-States | -1810.88 | 3735.77 | 16.72 | 9 | 3652.00 | 20.12 |
| F1 | SDM | -4039.13 | 8204.47 | 163.11 | 21 | 8108.38 | 112.73 |
| | HTSD | -4041.03 | 8208.26 | 166.90 | 21 | 8112.16 | 116.52 |
| | HTUVSD | -3989.36 | 8155.40 | 63.56 | 15 | 8020.92 | 25.28 |
| | SAC | -3981.48 | 8139.64 | 47.80 | 15 | 8005.16 | 9.52 |
| | Streak | -4047.63 | 8271.94 | 180.10 | 15 | 8137.46 | 141.82 |
| | 4-States | -3976.72 | 8130.12 | 38.28 | 15 | 7995.64 | 0.00 |
| F2 | SDM | -2264.05 | 4650.36 | 33.70 | 21 | 4558.24 | 3.99 |
| | HTSD | -2267.59 | 4657.44 | 40.78 | 21 | 4565.33 | 11.08 |
| | HTUVSD | -2261.06 | 4693.26 | 27.70 | 15 | 4564.38 | 10.13 |
| | SAC | -2255.99 | 4683.14 | 17.58 | 15 | 4554.25 | 0.00 |
| | Streak | -2260.06 | 4691.28 | 25.72 | 15 | 4562.39 | 8.14 |
| | 4-States | -2256.57 | 4684.29 | 18.73 | 15 | 4555.41 | 1.16 |
| G1 | SDM | -1600.89 | 3259.96 | 6.55 | 4 | 3217.88 | 0.29 |
| | HTSD | -1600.75 | 3259.67 | 6.27 | 4 | 3217.59 | 0.00 |
| | HTUVSD | -1600.68 | 3266.81 | 6.14 | 3 | 3219.49 | 1.90 |
| | SAC | -1602.51 | 3270.48 | 9.80 | 3 | 3223.15 | 5.56 |
| | Streak | -1600.80 | 3267.05 | 6.37 | 3 | 3219.72 | 2.13 |
| | 4-States | -1600.65 | 3266.74 | 6.07 | 3 | 3219.42 | 1.83 |
| G2 | SDM | -2461.23 | 4985.33 | 21.75 | 4 | 4938.51 | 12.29 |
| | HTSD | -2467.84 | 4998.56 | 34.97 | 4 | 4951.73 | 25.52 |
| | HTUVSD | -2461.23 | 4993.19 | 21.75 | 3 | 4940.52 | 14.31 |
| | SAC | -2454.83 | 4980.40 | 8.96 | 3 | 4927.73 | 1.51 |
| | Streak | -2454.07 | 4978.89 | 7.44 | 3 | 4926.22 | 0.00 |
| | 4-States | -2458.09 | 4986.92 | 15.48 | 3 | 4934.25 | 8.03 |
| H1 | SDM | -1762.86 | 3607.72 | 26.19 | 7 | 3547.87 | 18.41 |
| | HTSD | -1765.88 | 3613.76 | 32.23 | 7 | 3553.91 | 24.44 |
| | HTUVSD | -1762.86 | 3615.18 | 26.19 | 6 | 3549.90 | 20.44 |
| | SAC | -1754.92 | 3599.29 | 10.31 | 6 | 3534.02 | 4.56 |
| | Streak | -1753.91 | 3597.27 | 8.29 | 6 | 3532.00 | 2.54 |
| | 4-States | -1752.64 | 3594.74 | 5.75 | 6 | 3529.46 | 0.00 |
| H2 | SDM | -18028.21 | 36162.19 | 74.23 | 7 | 36078.43 | 41.76 |
| | HTSD | -18032.05 | 36169.87 | 81.92 | 7 | 36086.12 | 49.44 |
| | HTUVSD | -18028.21 | 36171.80 | 74.23 | 6 | 36080.43 | 43.76 |
| | SAC | -18006.33 | 36128.04 | 30.47 | 6 | 36036.67 | 0.00 |
| | Streak | -18010.52 | 36136.43 | 38.85 | 6 | 36045.06 | 8.38 |
| | 4-States | -18008.64 | 36132.67 | 35.10 | 6 | 36041.31 | 4.63 |

| Experiment | Modell | \mathcal{LL} | BIC | G^2 | df | AIC_c | ΔAIC_c |
|------------|----------|----------------|---------|-------|------|---------|----------------|
| H3 | SDM | -1525.54 | 3133.07 | 14.88 | 7 | 3073.22 | 8.15 |
| | HTSD | -1522.70 | 3127.39 | 9.20 | 7 | 3067.55 | 2.47 |
| | HTUVSD | -1521.94 | 3133.34 | 7.70 | 6 | 3068.07 | 2.99 |
| | SAC | -1525.05 | 3139.56 | 13.91 | 6 | 3074.28 | 9.20 |
| | Streak | -1525.16 | 3139.78 | 14.13 | 6 | 3074.50 | 9.43 |
| | 4-States | -1520.45 | 3130.35 | 4.71 | 6 | 3065.08 | 0.00 |
| I1 | SDM | -1014.74 | 2099.54 | 6.52 | 14 | 2049.67 | 0.00 |
| | HTSD | -1015.11 | 2100.30 | 7.28 | 14 | 2050.43 | 0.76 |
| | HTUVSD | -1014.71 | 2113.49 | 6.47 | 12 | 2053.70 | 4.03 |
| | SAC | -1015.47 | 2115.02 | 7.99 | 12 | 2055.23 | 5.56 |
| | Streak | -1015.09 | 2114.25 | 7.23 | 12 | 2054.46 | 4.79 |
| | 4-States | -1019.99 | 2124.06 | 17.03 | 12 | 2064.26 | 14.59 |
| I2 | SDM | -1033.33 | 2138.21 | 6.69 | 14 | 2086.83 | 0.00 |
| | HTSD | -1033.84 | 2139.22 | 7.71 | 14 | 2087.85 | 1.02 |
| | HTUVSD | -1033.09 | 2152.05 | 6.22 | 12 | 2090.44 | 3.60 |
| | SAC | -1031.70 | 2149.27 | 3.44 | 12 | 2087.66 | 0.82 |
| | Streak | -1044.35 | 2174.55 | 28.73 | 12 | 2112.95 | 26.11 |
| | 4-States | -1036.90 | 2159.65 | 13.82 | 12 | 2098.04 | 11.20 |
| J1 | SDM | -3575.76 | 7257.45 | 62.08 | 23 | 7177.62 | 9.59 |
| | HTSD | -3578.16 | 7262.25 | 66.89 | 23 | 7182.42 | 14.39 |
| | HTUVSD | -3569.78 | 7261.80 | 50.14 | 21 | 7169.71 | 1.68 |
| | SAC | -3569.77 | 7261.77 | 50.11 | 21 | 7169.68 | 1.65 |
| | Streak | -3569.84 | 7261.91 | 50.25 | 21 | 7169.81 | 1.79 |
| | 4-States | -3568.94 | 7260.12 | 48.46 | 21 | 7168.02 | 0.00 |

Anhang E

Datenanpassung hypothesendiskonformer Modellvarianten

Im Folgenden werden die Datenanpassungsindizes berichtet, die bei der Anpassung hypotheseninkonform spezifizierter Modellvarianten an die Daten der einzelnen Experimente erreicht wurden. Neben der logarithmierten Likelihood (\mathcal{LL}) und dem Bayesianischen Informationskriterium (BIC) werden G^2 -Werte mit Freiheitsgraden (df), AIC_c -Werte und AIC_c -Differenzwerte (ΔAIC_c) berichtet. In der Tabelle wird gesondert festgehalten für welches Modell und Experiment die jeweiligen Datenanpassungsindizes ermittelt wurden.

| Experiment | Modell | \mathcal{LL} | BIC | G^2 | df | AIC_c | ΔAIC_c |
|------------|----------|----------------|---------|--------|------|---------|----------------|
| A1 | SDM | -700.88 | 1453.45 | 75.44 | 4 | 1417.99 | 34.66 |
| | HTSD | -700.93 | 1453.55 | 75.54 | 4 | 1418.09 | 34.77 |
| | HTUVSD | -700.88 | 1459.91 | 75.44 | 3 | 1420.05 | 36.72 |
| | SAC | -682.52 | 1423.19 | 38.72 | 3 | 1383.32 | 0.00 |
| | Streak | -696.74 | 1451.64 | 67.16 | 3 | 1411.77 | 28.45 |
| | 4-States | -715.55 | 1489.25 | 104.77 | 3 | 1449.38 | 66.06 |
| A2 | SDM | -2065.75 | 4191.97 | 28.00 | 4 | 4147.57 | 19.00 |
| | HTSD | -2068.39 | 4197.26 | 33.28 | 4 | 4152.85 | 24.29 |
| | HTUVSD | -2065.75 | 4199.53 | 28.00 | 3 | 4149.58 | 21.02 |
| | SAC | -2055.41 | 4178.87 | 7.34 | 3 | 4128.92 | 0.36 |
| | Streak | -2055.23 | 4178.51 | 6.98 | 3 | 4128.56 | 0.00 |
| | 4-States | -2055.64 | 4179.33 | 7.79 | 3 | 4129.38 | 0.82 |
| A3 | SDM | -1610.55 | 3281.06 | 4.95 | 4 | 3237.18 | 1.52 |
| | HTSD | -1613.22 | 3286.41 | 10.30 | 4 | 3242.53 | 6.87 |
| | HTUVSD | -1610.55 | 3288.56 | 4.95 | 3 | 3239.20 | 3.54 |
| | SAC | -1609.33 | 3286.12 | 2.51 | 3 | 3236.76 | 1.10 |
| | Streak | -1608.78 | 3285.01 | 1.41 | 3 | 3235.65 | 0.00 |
| | 4-States | -1611.63 | 3290.73 | 7.13 | 3 | 3241.37 | 5.72 |
| A4 | SDM | -2159.00 | 4379.95 | 18.26 | 4 | 4334.07 | 13.52 |
| | HTSD | -2160.69 | 4383.33 | 21.64 | 4 | 4337.45 | 16.90 |
| | HTUVSD | -2159.00 | 4387.69 | 18.26 | 3 | 4336.09 | 15.54 |
| | SAC | -2151.46 | 4372.61 | 3.18 | 3 | 4321.00 | 0.46 |
| | Streak | -2151.24 | 4372.17 | 2.74 | 3 | 4320.57 | 0.02 |
| | 4-States | -2151.24 | 4372.15 | 2.72 | 3 | 4320.55 | 0.00 |
| A5 | SDM | -1873.06 | 3808.06 | 26.81 | 4 | 3762.18 | 20.07 |
| | HTSD | -1875.16 | 3812.26 | 31.00 | 4 | 3766.38 | 24.27 |
| | HTUVSD | -1873.06 | 3815.80 | 26.81 | 3 | 3764.20 | 22.08 |
| | SAC | -1862.51 | 3794.71 | 5.71 | 3 | 3743.11 | 0.99 |
| | Streak | -1862.22 | 3794.12 | 5.12 | 3 | 3742.51 | 0.40 |
| | 4-States | -1862.02 | 3793.72 | 4.72 | 3 | 3742.12 | 0.00 |
| A6 | SDM | -2371.14 | 4804.23 | 26.19 | 4 | 4758.35 | 9.63 |
| | HTSD | -2372.24 | 4806.41 | 28.38 | 4 | 4760.54 | 11.81 |
| | HTUVSD | -2371.14 | 4811.97 | 26.19 | 3 | 4760.37 | 11.65 |
| | SAC | -2365.67 | 4801.01 | 15.23 | 3 | 4749.41 | 0.69 |
| | Streak | -2365.32 | 4800.33 | 14.55 | 3 | 4748.72 | 0.00 |
| | 4-States | -2365.49 | 4800.66 | 14.88 | 3 | 4749.06 | 0.34 |
| A7 | SDM | -2051.82 | 4165.57 | 5.32 | 4 | 4119.70 | 0.44 |
| | HTSD | -2051.88 | 4165.69 | 5.45 | 4 | 4119.82 | 0.56 |
| | HTUVSD | -2051.82 | 4173.33 | 5.34 | 3 | 4121.73 | 2.47 |
| | SAC | -2050.59 | 4170.86 | 2.87 | 3 | 4119.25 | 0.00 |
| | Streak | -2051.84 | 4173.37 | 5.38 | 3 | 4121.77 | 2.51 |
| | 4-States | -2053.36 | 4176.40 | 8.42 | 3 | 4124.80 | 5.55 |

| Experiment | Modell | \mathcal{LL} | BIC | G^2 | df | AIC_c | ΔAIC_c |
|------------|----------|----------------|----------|--------|------|----------|----------------|
| B1 | SDM | -2127.44 | 4316.80 | 46.11 | 4 | 4270.95 | 20.74 |
| | HTSD | -2128.53 | 4318.98 | 48.29 | 4 | 4273.13 | 22.92 |
| | HTUVSD | -2127.44 | 4324.54 | 46.11 | 3 | 4272.96 | 22.75 |
| | SAC | -2117.01 | 4303.69 | 25.25 | 3 | 4252.11 | 1.90 |
| | Streak | -2116.07 | 4301.79 | 23.36 | 3 | 4250.21 | 0.00 |
| | 4-States | -2116.43 | 4302.52 | 24.08 | 3 | 4250.94 | 0.73 |
| B2 | SDM | -2311.91 | 4686.09 | 49.09 | 4 | 4639.89 | 19.58 |
| | HTSD | -2309.17 | 4680.60 | 43.60 | 4 | 4634.39 | 14.08 |
| | HTUVSD | -2309.16 | 4688.37 | 43.58 | 3 | 4636.39 | 16.08 |
| | SAC | -2316.45 | 4702.95 | 58.17 | 3 | 4650.98 | 30.67 |
| | Streak | -2310.04 | 4690.13 | 45.34 | 3 | 4638.15 | 17.84 |
| | 4-States | -2301.12 | 4672.29 | 27.50 | 3 | 4620.31 | 0.00 |
| B3 | SDM | -3867.58 | 7801.52 | 7.12 | 4 | 7751.20 | 3.32 |
| | HTSD | -3869.48 | 7805.31 | 10.91 | 4 | 7754.99 | 7.11 |
| | HTUVSD | -3867.60 | 7809.84 | 7.15 | 3 | 7753.24 | 5.36 |
| | SAC | -3865.66 | 7805.97 | 3.28 | 3 | 7749.37 | 1.48 |
| | Streak | -3865.79 | 7806.23 | 3.54 | 3 | 7749.63 | 1.74 |
| | 4-States | -3864.92 | 7804.49 | 1.79 | 3 | 7747.89 | 0.00 |
| B4 | SDM | -6505.77 | 13080.81 | 39.84 | 4 | 13027.56 | 19.65 |
| | HTSD | -6529.22 | 13127.70 | 86.73 | 4 | 13074.46 | 66.55 |
| | HTUVSD | -6505.77 | 13089.47 | 39.84 | 3 | 13029.57 | 21.66 |
| | SAC | -6496.76 | 13071.44 | 21.82 | 3 | 13011.55 | 3.64 |
| | Streak | -6494.94 | 13067.80 | 18.18 | 3 | 13007.91 | 0.00 |
| | 4-States | -6500.44 | 13078.81 | 29.18 | 3 | 13018.91 | 11.00 |
| B5 | SDM | -3246.52 | 6556.77 | 17.63 | 4 | 6509.09 | 2.73 |
| | HTSD | -3245.23 | 6554.19 | 15.05 | 4 | 6506.52 | 0.16 |
| | HTUVSD | -3244.85 | 6561.40 | 14.29 | 3 | 6507.77 | 1.41 |
| | SAC | -3244.65 | 6560.98 | 13.88 | 3 | 6507.35 | 0.99 |
| | Streak | -3244.15 | 6559.99 | 12.88 | 3 | 6506.36 | 0.00 |
| | 4-States | -3244.68 | 6561.06 | 13.96 | 3 | 6507.43 | 1.07 |
| B6 | SDM | -4087.57 | 8240.65 | 3.28 | 4 | 8191.18 | 0.00 |
| | HTSD | -4087.84 | 8241.19 | 3.82 | 4 | 8191.72 | 0.54 |
| | HTUVSD | -4087.57 | 8248.84 | 3.28 | 3 | 8193.20 | 2.01 |
| | SAC | -4086.69 | 8247.07 | 1.51 | 3 | 8191.43 | 0.24 |
| | Streak | -4086.67 | 8247.05 | 1.48 | 3 | 8191.40 | 0.21 |
| | 4-States | -4090.26 | 8254.22 | 8.66 | 3 | 8198.58 | 7.39 |
| B7 | SDM | -2878.93 | 5823.99 | 157.67 | 4 | 5773.90 | 98.01 |
| | HTSD | -2889.92 | 5845.96 | 179.64 | 4 | 5795.87 | 119.97 |
| | HTUVSD | -2878.93 | 5832.26 | 157.67 | 3 | 5775.91 | 100.01 |
| | SAC | -2833.55 | 5741.48 | 66.90 | 3 | 5685.14 | 9.24 |
| | Streak | -2828.93 | 5732.24 | 57.66 | 3 | 5675.90 | 0.00 |
| | 4-States | -2838.29 | 5750.97 | 76.38 | 3 | 5694.62 | 18.72 |

| Experiment | Modell | \mathcal{LL} | BIC | G^2 | df | AIC_c | ΔAIC_c |
|------------|----------|----------------|----------|---------|------|----------|----------------|
| B8 | SDM | -1426.63 | 2912.29 | 50.22 | 4 | 2869.35 | 32.40 |
| | HTSD | -1429.81 | 2918.65 | 56.58 | 4 | 2875.72 | 38.76 |
| | HTUVSD | -1426.63 | 2919.66 | 50.22 | 3 | 2871.38 | 34.42 |
| | SAC | -1409.84 | 2886.09 | 16.64 | 3 | 2837.80 | 0.84 |
| | Streak | -1409.42 | 2885.25 | 15.80 | 3 | 2836.96 | 0.00 |
| | 4-States | -1414.08 | 2894.56 | 25.12 | 3 | 2846.28 | 9.32 |
| B9 | SDM | -786.68 | 1632.39 | 10.94 | 4 | 1589.46 | 5.98 |
| | HTSD | -789.53 | 1638.08 | 16.63 | 4 | 1595.14 | 11.67 |
| | HTUVSD | -786.68 | 1639.77 | 10.94 | 3 | 1591.48 | 8.00 |
| | SAC | -783.90 | 1634.19 | 5.37 | 3 | 1585.91 | 2.43 |
| | Streak | -782.68 | 1631.77 | 2.94 | 3 | 1583.48 | 0.00 |
| | 4-States | -789.51 | 1645.42 | 16.60 | 3 | 1597.14 | 13.66 |
| C1 | SDM | -858.70 | 1796.54 | 14.06 | 7 | 1739.60 | 0.21 |
| | HTSD | -864.25 | 1807.63 | 25.15 | 7 | 1750.69 | 11.31 |
| | HTUVSD | -858.70 | 1803.74 | 14.06 | 6 | 1741.64 | 2.26 |
| | SAC | -857.89 | 1802.11 | 12.43 | 6 | 1740.01 | 0.63 |
| | Streak | -857.57 | 1801.48 | 11.81 | 6 | 1739.39 | 0.00 |
| | 4-States | -859.59 | 1805.52 | 15.85 | 6 | 1743.43 | 4.04 |
| C2 | SDM | -3731.86 | 7553.35 | 39.58 | 7 | 7485.79 | 17.56 |
| | HTSD | -3743.77 | 7577.18 | 63.41 | 7 | 7509.62 | 41.39 |
| | HTUVSD | -3731.86 | 7561.49 | 39.58 | 6 | 7487.81 | 19.58 |
| | SAC | -3722.60 | 7542.98 | 21.07 | 6 | 7469.30 | 1.07 |
| | Streak | -3722.07 | 7541.92 | 20.00 | 6 | 7468.23 | 0.00 |
| | 4-States | -3729.51 | 7556.81 | 34.89 | 6 | 7483.12 | 14.89 |
| C3 | SDM | -15705.11 | 31514.59 | 1087.89 | 7 | 31432.24 | 239.55 |
| | HTSD | -15722.71 | 31549.78 | 1123.07 | 7 | 31467.43 | 274.74 |
| | HTUVSD | -15705.11 | 31524.08 | 1087.89 | 6 | 31434.25 | 241.56 |
| | SAC | -15584.33 | 31282.52 | 846.33 | 6 | 31192.69 | 0.00 |
| | Streak | -15613.39 | 31340.64 | 904.45 | 6 | 31250.81 | 58.12 |
| | 4-States | -15707.15 | 31528.15 | 1091.96 | 6 | 31438.32 | 245.63 |
| D1 | SDM | -19467.16 | 39075.03 | 746.88 | 10 | 38962.33 | 429.40 |
| | HTSD | -19481.86 | 39104.44 | 776.29 | 10 | 38991.74 | 458.81 |
| | HTUVSD | -19467.16 | 39085.08 | 746.88 | 9 | 38964.33 | 431.40 |
| | SAC | -19288.79 | 38728.34 | 390.14 | 9 | 38607.59 | 74.66 |
| | Streak | -19251.46 | 38653.68 | 315.48 | 9 | 38532.93 | 0.00 |
| | 4-States | -19582.26 | 39315.29 | 977.09 | 9 | 39194.54 | 661.61 |
| D2 | SDM | -3391.26 | 6898.64 | 50.23 | 10 | 6810.62 | 17.04 |
| | HTSD | -3391.10 | 6898.34 | 49.92 | 10 | 6810.31 | 16.74 |
| | HTUVSD | -3390.88 | 6906.18 | 49.47 | 9 | 6811.87 | 18.30 |
| | SAC | -3381.73 | 6887.88 | 31.17 | 9 | 6793.58 | 0.00 |
| | Streak | -3390.06 | 6904.55 | 47.84 | 9 | 6810.24 | 16.67 |
| | 4-States | -3403.98 | 6932.39 | 75.68 | 9 | 6838.09 | 44.51 |

| Experiment | Modell | \mathcal{LL} | BIC | G^2 | df | AIC_c | ΔAIC_c |
|------------|----------|----------------|----------|--------|------|----------|----------------|
| D3 | SDM | -2616.36 | 5340.20 | 45.71 | 10 | 5260.91 | 9.06 |
| | HTSD | -2616.78 | 5341.06 | 46.57 | 10 | 5261.76 | 9.91 |
| | HTUVSD | -2616.36 | 5347.88 | 45.71 | 9 | 5262.94 | 11.09 |
| | SAC | -2610.81 | 5336.79 | 34.63 | 9 | 5251.85 | 0.00 |
| | Streak | -2615.76 | 5346.69 | 44.53 | 9 | 5261.75 | 9.90 |
| | 4-States | -2626.17 | 5367.52 | 65.35 | 9 | 5282.57 | 30.72 |
| D4 | SDM | -4838.82 | 9795.74 | 39.20 | 10 | 9705.74 | 10.93 |
| | HTSD | -4846.28 | 9810.65 | 54.11 | 10 | 9720.65 | 25.84 |
| | HTUVSD | -4838.82 | 9804.18 | 39.20 | 9 | 9707.75 | 12.94 |
| | SAC | -4832.35 | 9791.23 | 26.25 | 9 | 9694.81 | 0.00 |
| | Streak | -4837.90 | 9802.34 | 37.35 | 9 | 9705.91 | 11.10 |
| | 4-States | -4844.12 | 9814.77 | 49.79 | 9 | 9718.34 | 23.54 |
| D5 | SDM | -6139.28 | 12398.81 | 248.42 | 10 | 12306.63 | 29.47 |
| | HTSD | -6153.68 | 12427.61 | 277.22 | 10 | 12335.43 | 58.27 |
| | HTUVSD | -6130.68 | 12390.21 | 231.24 | 9 | 12291.46 | 14.30 |
| | SAC | -6123.59 | 12376.03 | 217.05 | 9 | 12277.27 | 0.12 |
| | Streak | -6123.53 | 12375.91 | 216.94 | 9 | 12277.16 | 0.00 |
| | 4-States | -6141.88 | 12412.61 | 253.64 | 9 | 12313.86 | 36.70 |
| E1 | SDM | -1299.55 | 2701.89 | 19.73 | 10 | 2627.38 | 9.06 |
| | HTSD | -1300.62 | 2704.04 | 21.88 | 10 | 2629.52 | 11.20 |
| | HTUVSD | -1299.55 | 2709.24 | 19.74 | 9 | 2629.42 | 11.10 |
| | SAC | -1294.00 | 2698.14 | 8.64 | 9 | 2618.32 | 0.00 |
| | Streak | -1299.20 | 2708.53 | 19.03 | 9 | 2628.71 | 10.39 |
| | 4-States | -1306.20 | 2722.53 | 33.03 | 9 | 2642.71 | 24.39 |
| E2 | SDM | -1378.62 | 2859.96 | 5.37 | 10 | 2785.52 | 0.00 |
| | HTSD | -1378.74 | 2860.20 | 5.61 | 10 | 2785.76 | 0.24 |
| | HTUVSD | -1378.63 | 2867.32 | 5.39 | 9 | 2787.58 | 2.06 |
| | SAC | -1377.80 | 2865.65 | 3.72 | 9 | 2785.91 | 0.39 |
| | Streak | -1378.58 | 2867.22 | 5.29 | 9 | 2787.48 | 1.96 |
| | 4-States | -1379.24 | 2868.54 | 6.62 | 9 | 2788.81 | 3.29 |
| E3 | SDM | -1532.66 | 3168.03 | 29.92 | 10 | 3093.59 | 3.13 |
| | HTSD | -1532.66 | 3168.04 | 29.93 | 10 | 3093.59 | 3.14 |
| | HTUVSD | -1532.66 | 3175.37 | 29.92 | 9 | 3095.63 | 5.17 |
| | SAC | -1531.55 | 3173.15 | 27.71 | 9 | 3093.42 | 2.96 |
| | Streak | -1530.07 | 3170.20 | 24.75 | 9 | 3090.46 | 0.00 |
| | 4-States | -1531.14 | 3172.34 | 26.89 | 9 | 3092.60 | 2.14 |
| E4 | SDM | -2603.37 | 5317.76 | 66.28 | 10 | 5234.89 | 3.35 |
| | HTSD | -2617.65 | 5346.33 | 94.85 | 10 | 5263.46 | 31.91 |
| | HTUVSD | -2602.90 | 5324.75 | 65.33 | 9 | 5235.97 | 4.42 |
| | SAC | -2600.68 | 5320.32 | 60.91 | 9 | 5231.54 | 0.00 |
| | Streak | -2603.66 | 5326.27 | 66.86 | 9 | 5237.50 | 5.95 |
| | 4-States | -2613.43 | 5345.80 | 86.39 | 9 | 5257.02 | 25.48 |

| Experiment | Modell | \mathcal{LL} | BIC | G^2 | df | AIC_c | ΔAIC_c |
|------------|----------|----------------|---------|--------|------|---------|----------------|
| E5 | SDM | -2519.80 | 5152.03 | 69.56 | 10 | 5067.74 | 40.28 |
| | HTSD | -2519.11 | 5150.64 | 68.17 | 10 | 5066.35 | 38.88 |
| | HTUVSD | -2519.05 | 5158.56 | 68.06 | 9 | 5068.27 | 40.80 |
| | SAC | -2498.66 | 5117.76 | 27.26 | 9 | 5027.47 | 0.00 |
| | Streak | -2514.03 | 5148.52 | 58.02 | 9 | 5058.22 | 30.75 |
| | 4-States | -2524.59 | 5169.63 | 79.13 | 9 | 5079.34 | 51.87 |
| E6 | SDM | -2472.58 | 5058.09 | 41.48 | 10 | 4973.30 | 7.98 |
| | HTSD | -2473.35 | 5059.62 | 43.01 | 10 | 4974.83 | 9.51 |
| | HTUVSD | -2472.58 | 5066.15 | 41.48 | 9 | 4975.31 | 10.00 |
| | SAC | -2467.58 | 5056.15 | 31.48 | 9 | 4965.31 | 0.00 |
| | Streak | -2471.65 | 5064.28 | 39.61 | 9 | 4973.44 | 8.13 |
| | 4-States | -2476.87 | 5074.72 | 50.05 | 9 | 4983.89 | 18.57 |
| E7 | SDM | -3231.02 | 6579.23 | 95.78 | 10 | 6490.14 | 31.49 |
| | HTSD | -3260.75 | 6638.69 | 155.24 | 10 | 6549.59 | 90.95 |
| | HTUVSD | -3228.94 | 6583.45 | 91.63 | 9 | 6488.00 | 29.36 |
| | SAC | -3214.27 | 6554.10 | 62.27 | 9 | 6458.64 | 0.00 |
| | Streak | -3229.13 | 6583.82 | 91.99 | 9 | 6488.36 | 29.72 |
| | 4-States | -3241.89 | 6609.35 | 117.53 | 9 | 6513.90 | 55.26 |
| E8 | SDM | -2328.86 | 4767.60 | 52.72 | 10 | 4685.88 | 20.83 |
| | HTSD | -2334.82 | 4779.53 | 64.65 | 10 | 4697.81 | 32.76 |
| | HTUVSD | -2325.98 | 4769.70 | 46.97 | 9 | 4682.15 | 17.10 |
| | SAC | -2317.43 | 4752.60 | 29.87 | 9 | 4665.05 | 0.00 |
| | Streak | -2321.80 | 4761.33 | 38.60 | 9 | 4673.78 | 8.73 |
| | 4-States | -2334.26 | 4786.27 | 63.53 | 9 | 4698.71 | 33.66 |
| E9 | SDM | -786.68 | 1632.39 | 10.94 | 4 | 1589.46 | 156.77 |
| | HTSD | -789.53 | 1638.08 | 16.63 | 4 | 1595.14 | 170.90 |
| | HTUVSD | -786.68 | 1639.77 | 10.94 | 3 | 1591.48 | 143.46 |
| | SAC | -783.90 | 1634.19 | 5.37 | 3 | 1585.91 | 0.00 |
| | Streak | -782.68 | 1631.77 | 2.94 | 3 | 1583.48 | 31.63 |
| | 4-States | -789.51 | 1645.42 | 16.60 | 3 | 1597.14 | 154.84 |
| E10 | SDM | -3304.94 | 6725.88 | 53.38 | 10 | 6637.98 | 10.84 |
| | HTSD | -3306.69 | 6729.39 | 56.89 | 10 | 6641.50 | 14.36 |
| | HTUVSD | -3304.94 | 6734.16 | 53.38 | 9 | 6640.00 | 12.86 |
| | SAC | -3299.93 | 6724.15 | 43.37 | 9 | 6629.99 | 2.85 |
| | Streak | -3298.58 | 6721.44 | 40.66 | 9 | 6627.28 | 0.14 |
| | 4-States | -3298.51 | 6721.30 | 40.52 | 9 | 6627.14 | 0.00 |
| E11 | SDM | -2080.44 | 4267.16 | 5.08 | 10 | 4189.10 | 0.00 |
| | HTSD | -2081.35 | 4268.98 | 6.90 | 10 | 4190.92 | 1.82 |
| | HTUVSD | -2080.45 | 4274.77 | 5.10 | 9 | 4191.15 | 2.05 |
| | SAC | -2079.58 | 4273.02 | 3.35 | 9 | 4189.40 | 0.30 |
| | Streak | -2080.27 | 4274.41 | 4.74 | 9 | 4190.79 | 1.70 |
| | 4-States | -2082.65 | 4279.16 | 9.49 | 9 | 4195.54 | 6.44 |

| Experiment | Modell | \mathcal{LL} | BIC | G^2 | df | AIC_c | ΔAIC_c |
|------------|----------|----------------|----------|--------|------|----------|----------------|
| E12 | SDM | -1806.14 | 3718.70 | 7.25 | 10 | 3640.50 | 0.00 |
| | HTSD | -1818.76 | 3743.93 | 32.48 | 10 | 3665.73 | 25.23 |
| | HTUVSD | -1805.95 | 3725.92 | 6.87 | 9 | 3642.14 | 1.64 |
| | SAC | -1805.64 | 3725.30 | 6.25 | 9 | 3641.53 | 1.03 |
| | Streak | -1806.29 | 3726.60 | 7.55 | 9 | 3642.83 | 2.33 |
| | 4-States | -1809.03 | 3732.08 | 13.03 | 9 | 3648.31 | 7.80 |
| F1 | SDM | -4075.49 | 8319.26 | 235.83 | 16 | 8191.18 | 49.17 |
| | HTSD | -4075.51 | 8319.28 | 235.86 | 16 | 8191.20 | 49.19 |
| | HTUVSD | -4075.49 | 8327.67 | 235.83 | 15 | 8193.19 | 51.18 |
| | SAC | -4049.90 | 8276.49 | 184.65 | 15 | 8142.01 | 0.00 |
| | Streak | -4075.42 | 8327.53 | 235.69 | 15 | 8193.05 | 51.04 |
| | 4-States | -4106.63 | 8389.95 | 298.11 | 15 | 8255.47 | 113.46 |
| F2 | SDM | -2277.78 | 4718.56 | 61.15 | 16 | 4595.80 | 40.42 |
| | HTSD | -2278.42 | 4719.84 | 62.43 | 16 | 4597.08 | 41.71 |
| | HTUVSD | -2277.74 | 4726.63 | 61.07 | 15 | 4597.75 | 42.38 |
| | SAC | -2256.55 | 4684.26 | 18.70 | 15 | 4555.37 | 0.00 |
| | Streak | -2277.35 | 4725.86 | 60.30 | 15 | 4596.97 | 41.60 |
| | 4-States | -2289.75 | 4750.65 | 85.09 | 15 | 4621.76 | 66.39 |
| G1 | SDM | -1612.07 | 3275.04 | 28.91 | 5 | 3238.21 | 22.35 |
| | HTSD | -1611.92 | 3274.75 | 28.62 | 5 | 3237.92 | 22.06 |
| | HTUVSD | -1606.12 | 3277.69 | 17.01 | 3 | 3230.36 | 14.50 |
| | SAC | -1602.39 | 3270.24 | 9.56 | 3 | 3222.91 | 7.05 |
| | Streak | -1605.16 | 3275.77 | 15.09 | 3 | 3228.44 | 12.58 |
| | 4-States | -1598.87 | 3263.19 | 2.51 | 3 | 3215.86 | 0.00 |
| G2 | SDM | -2501.31 | 5057.65 | 101.93 | 5 | 5016.67 | 83.38 |
| | HTSD | -2506.62 | 5068.27 | 112.54 | 5 | 5027.29 | 94.00 |
| | HTUVSD | -2484.88 | 5040.50 | 69.06 | 3 | 4987.83 | 54.54 |
| | SAC | -2470.34 | 5011.42 | 39.97 | 3 | 4958.74 | 25.45 |
| | Streak | -2468.72 | 5008.18 | 36.73 | 3 | 4955.51 | 22.21 |
| | 4-States | -2457.61 | 4985.96 | 14.52 | 3 | 4933.29 | 0.00 |
| H1 | SDM | -1782.35 | 3631.78 | 65.16 | 9 | 3582.80 | 49.20 |
| | HTSD | -1785.05 | 3637.18 | 70.56 | 9 | 3588.20 | 54.60 |
| | HTUVSD | -1767.39 | 3624.24 | 35.26 | 6 | 3558.97 | 25.38 |
| | SAC | -1754.70 | 3598.87 | 9.88 | 6 | 3533.59 | 0.00 |
| | Streak | -1756.76 | 3602.97 | 13.99 | 6 | 3537.70 | 4.11 |
| | 4-States | -1756.46 | 3602.37 | 13.39 | 6 | 3537.10 | 3.51 |
| H2 | SDM | -18138.11 | 36362.76 | 294.03 | 9 | 36294.23 | 171.98 |
| | HTSD | -18153.91 | 36394.36 | 325.64 | 9 | 36325.83 | 203.58 |
| | HTUVSD | -18146.28 | 36407.94 | 310.37 | 6 | 36316.57 | 194.32 |
| | SAC | -18105.21 | 36325.82 | 228.25 | 6 | 36234.45 | 112.20 |
| | Streak | -18138.23 | 36391.86 | 294.29 | 6 | 36300.49 | 178.24 |
| | 4-States | -18049.12 | 36213.62 | 116.05 | 6 | 36122.25 | 0.00 |

| Experiment | Modell | \mathcal{LL} | BIC | G^2 | df | AIC_c | ΔAIC_c |
|------------|----------|----------------|---------|--------|------|---------|----------------|
| H3 | SDM | -1535.07 | 3137.23 | 33.95 | 9 | 3088.24 | 20.21 |
| | HTSD | -1535.55 | 3138.18 | 34.90 | 9 | 3089.20 | 21.17 |
| | HTUVSD | -1521.92 | 3133.30 | 7.66 | 6 | 3068.03 | 0.00 |
| | SAC | -1522.88 | 3135.22 | 9.58 | 6 | 3069.95 | 1.92 |
| | Streak | -1522.75 | 3134.95 | 9.31 | 6 | 3069.68 | 1.65 |
| | 4-States | -1524.81 | 3139.07 | 13.42 | 6 | 3073.79 | 5.76 |
| I1 | SDM | -1018.09 | 2106.25 | 13.23 | 14 | 2056.38 | 1.98 |
| | HTSD | -1018.49 | 2107.05 | 14.03 | 14 | 2057.18 | 2.78 |
| | HTUVSD | -1016.78 | 2117.64 | 10.62 | 12 | 2057.85 | 3.45 |
| | SAC | -1015.06 | 2114.19 | 7.16 | 12 | 2054.40 | 0.00 |
| | Streak | -1022.16 | 2128.41 | 21.38 | 12 | 2068.61 | 14.22 |
| | 4-States | -1017.10 | 2118.29 | 11.26 | 12 | 2058.49 | 4.09 |
| I2 | SDM | -1045.68 | 2162.91 | 31.40 | 14 | 2111.54 | 13.47 |
| | HTSD | -1049.93 | 2171.41 | 39.90 | 14 | 2120.04 | 21.97 |
| | HTUVSD | -1044.10 | 2174.05 | 28.23 | 12 | 2112.44 | 14.37 |
| | SAC | -1036.91 | 2159.68 | 13.85 | 12 | 2098.07 | 0.00 |
| | Streak | -1044.75 | 2175.35 | 29.52 | 12 | 2113.74 | 15.67 |
| | 4-States | -1047.23 | 2180.32 | 34.50 | 12 | 2118.72 | 20.65 |
| J1 | SDM | -3603.05 | 7303.89 | 116.67 | 24 | 7230.19 | 71.75 |
| | HTSD | -3605.03 | 7307.85 | 120.63 | 24 | 7234.15 | 75.71 |
| | HTUVSD | -3584.31 | 7290.86 | 79.20 | 21 | 7198.76 | 40.32 |
| | SAC | -3573.65 | 7269.54 | 57.88 | 21 | 7177.45 | 19.01 |
| | Streak | -3579.44 | 7281.11 | 69.45 | 21 | 7189.01 | 30.57 |
| | 4-States | -3564.15 | 7250.54 | 38.88 | 21 | 7158.44 | 0.00 |

Anhang F

Datenanpassung unrestringierter Modellvarianten

Im Folgenden werden die Datenanpassungsindizes berichtet, die bei der freien Anpassung der Modelle an die Daten der einzelnen Experimente erreicht wurden. Neben der logarithmierten Likelihood wird der Korrekturfaktor für den Vorhersagefehler *Optim* berichtet. Außerdem werden G^2 -Werte mit Freiheitsgraden (*df*), der AIC_c -Wert und AIC_c -Differenzwerte (ΔAIC_c) berichtet. In der Tabelle wird gesondert festgehalten, für welches Modell und Experiment die jeweiligen Datenanpassungsindizes ermittelt wurden.

| Experiment | Modell | \mathcal{LL} | $Optim$ | BIC | G^2 | df | AIC_c | ΔAIC_c |
|------------|----------|----------------|---------|---------|-------|------|---------|----------------|
| A1 | SDM | -664.27 | 6.96 | 1373.76 | 2.21 | 2 | 1342.71 | 1.34 |
| | HTSD | -664.99 | 6.71 | 1375.21 | 3.66 | 2 | 1344.16 | 2.79 |
| | HTUVSD | -664.41 | 7.20 | 1380.51 | 2.50 | 1 | 1345.05 | 1.63 |
| | SAC | -663.59 | 7.85 | 1378.88 | 0.87 | 1 | 1343.42 | 0.00 |
| | Streak | -676.70 | 6.88 | 1405.08 | 27.07 | 1 | 1369.62 | 26.20 |
| | 4-States | -663.68 | 7.62 | 1379.05 | 1.04 | 1 | 1343.59 | 0.17 |
| A2 | SDM | -2067.13 | 7.79 | 4187.19 | 26.09 | 2 | 4148.33 | 26.01 |
| | HTSD | -2071.32 | 5.95 | 4195.56 | 34.47 | 2 | 4156.70 | 34.38 |
| | HTUVSD | -2067.85 | 7.77 | 4196.18 | 27.52 | 1 | 4151.77 | 27.44 |
| | SAC | -2054.42 | 8.14 | 4169.33 | 0.67 | 1 | 4124.92 | 0.59 |
| | Streak | -2054.79 | 7.05 | 4170.06 | 1.41 | 1 | 4125.66 | 1.32 |
| | 4-States | -2054.13 | 8.13 | 4168.74 | 0.08 | 1 | 4124.33 | 0.00 |
| A3 | SDM | -1609.83 | 8.58 | 3272.13 | 3.51 | 2 | 3233.72 | 2.97 |
| | HTSD | -1615.11 | 7.66 | 3282.68 | 14.07 | 2 | 3244.28 | 13.52 |
| | HTUVSD | -1610.49 | 8.65 | 3280.94 | 4.83 | 1 | 3237.05 | 4.28 |
| | SAC | -1608.35 | 9.63 | 3276.65 | 0.55 | 1 | 3232.77 | 0.00 |
| | Streak | -1612.97 | 8.95 | 3285.90 | 9.80 | 1 | 3242.02 | 9.25 |
| | 4-States | -1608.88 | 9.74 | 3277.72 | 1.61 | 1 | 3233.84 | 1.06 |
| A4 | SDM | -2155.41 | 11.60 | 4365.01 | 16.05 | 2 | 4324.86 | 15.02 |
| | HTSD | -2159.18 | 9.96 | 4372.57 | 23.61 | 2 | 4332.42 | 22.58 |
| | HTUVSD | -2157.19 | 11.49 | 4376.32 | 19.61 | 1 | 4330.44 | 18.59 |
| | SAC | -2147.89 | 13.13 | 4357.73 | 1.02 | 1 | 4311.85 | 0.00 |
| | Streak | -2149.64 | 11.97 | 4361.21 | 4.51 | 1 | 4315.34 | 3.49 |
| | 4-States | -2148.83 | 11.83 | 4359.60 | 2.89 | 1 | 4313.72 | 1.87 |
| A5 | SDM | -1874.24 | 11.45 | 3802.67 | 25.15 | 2 | 3762.52 | 20.00 |
| | HTSD | -1878.57 | 10.03 | 3811.33 | 33.81 | 2 | 3771.18 | 28.66 |
| | HTUVSD | -1876.38 | 11.41 | 3814.70 | 29.44 | 1 | 3768.82 | 24.29 |
| | SAC | -1864.23 | 12.23 | 3790.41 | 5.15 | 1 | 3744.53 | 0.00 |
| | Streak | -1865.32 | 11.19 | 3792.58 | 7.32 | 1 | 3746.71 | 2.17 |
| | 4-States | -1865.72 | 12.10 | 3793.37 | 8.11 | 1 | 3747.50 | 2.97 |
| A6 | SDM | -2365.08 | 3.34 | 4784.36 | 17.79 | 2 | 4744.21 | 17.79 |
| | HTSD | -2368.83 | 2.06 | 4791.86 | 25.29 | 2 | 4751.71 | 25.28 |
| | HTUVSD | -2366.08 | 3.42 | 4794.11 | 19.80 | 1 | 4748.23 | 19.79 |
| | SAC | -2356.19 | 4.44 | 4774.31 | 0.01 | 1 | 4728.44 | 0.00 |
| | Streak | -2357.55 | 3.37 | 4777.03 | 2.72 | 1 | 4731.16 | 2.72 |
| | 4-States | -2358.34 | 4.48 | 4778.61 | 4.31 | 1 | 4732.74 | 4.30 |
| A7 | SDM | -2048.54 | 9.15 | 4151.27 | 0.08 | 2 | 4111.12 | 0.08 |
| | HTSD | -2048.65 | 8.88 | 4151.51 | 0.32 | 2 | 4111.36 | 0.31 |
| | HTUVSD | -2048.56 | 9.54 | 4159.05 | 0.12 | 1 | 4113.18 | 0.12 |
| | SAC | -2048.50 | 10.31 | 4158.94 | 0.00 | 1 | 4113.06 | 0.00 |
| | Streak | -2049.64 | 9.21 | 4161.21 | 2.28 | 1 | 4115.33 | 2.28 |
| | 4-States | -2048.78 | 10.34 | 4159.50 | 0.57 | 1 | 4113.63 | 0.57 |

| Experiment | Modell | \mathcal{LL} | $Optim$ | BIC | G^2 | df | AIC_c | ΔAIC_c |
|------------|----------|----------------|---------|----------|--------|------|----------|----------------|
| B1 | SDM | -2109.18 | 10.14 | 4295.76 | 9.59 | 2 | 4238.46 | 9.58 |
| | HTSD | -2110.77 | 8.51 | 4298.94 | 12.77 | 2 | 4241.64 | 12.77 |
| | HTUVSD | -2109.18 | 10.09 | 4311.24 | 9.59 | 0 | 4242.50 | 9.59 |
| | SAC | -2104.39 | 11.94 | 4301.66 | 0.00 | 0 | 4232.91 | 0.00 |
| | Streak | -2104.39 | 11.89 | 4301.66 | 0.01 | 0 | 4232.92 | 0.00 |
| | 4-States | -2105.05 | 11.04 | 4302.97 | 1.31 | 0 | 4234.23 | 1.31 |
| B2 | SDM | -2289.65 | 14.58 | 4657.14 | 4.57 | 2 | 4599.40 | 4.57 |
| | HTSD | -2289.35 | 28.10 | 4656.53 | 3.96 | 2 | 4598.79 | 3.96 |
| | HTUVSD | -2287.99 | 16.20 | 4669.38 | 1.24 | 0 | 4600.11 | 1.24 |
| | SAC | -2287.37 | 16.92 | 4668.14 | 0.00 | 0 | 4598.87 | 0.00 |
| | Streak | -2287.37 | 17.07 | 4668.14 | 0.00 | 0 | 4598.87 | 0.00 |
| | 4-States | -2287.37 | 16.45 | 4668.14 | 0.00 | 0 | 4598.87 | 0.00 |
| B3 | SDM | -3866.03 | 8.95 | 7815.00 | 4.02 | 2 | 7752.12 | 4.02 |
| | HTSD | -3868.76 | 8.90 | 7820.46 | 9.47 | 2 | 7757.57 | 9.47 |
| | HTUVSD | -3866.05 | 9.31 | 7831.63 | 4.05 | 0 | 7756.18 | 4.05 |
| | SAC | -3864.02 | 11.15 | 7827.57 | 0.00 | 0 | 7752.12 | 0.00 |
| | Streak | -3864.02 | 11.15 | 7827.57 | 0.00 | 0 | 7752.12 | 0.00 |
| | 4-States | -3864.02 | 11.15 | 7827.57 | 0.00 | 0 | 7752.12 | 0.00 |
| B4 | SDM | -6500.52 | 7.22 | 13087.63 | 29.35 | 2 | 13021.09 | 29.35 |
| | HTSD | -6521.76 | 5.14 | 13130.10 | 71.81 | 2 | 13063.55 | 71.81 |
| | HTUVSD | -6500.52 | 7.22 | 13104.95 | 29.35 | 0 | 13025.10 | 29.35 |
| | SAC | -6485.85 | 9.31 | 13075.60 | 0.00 | 0 | 12995.75 | 0.00 |
| | Streak | -6485.85 | 9.31 | 13075.60 | 0.00 | 0 | 12995.75 | 0.00 |
| | 4-States | -6485.85 | 9.31 | 13075.60 | 0.00 | 0 | 12995.75 | 0.00 |
| B5 | SDM | -3241.01 | 11.20 | 6561.67 | 6.60 | 2 | 6502.09 | 6.60 |
| | HTSD | -3238.51 | 11.11 | 6556.68 | 1.61 | 2 | 6497.10 | 1.61 |
| | HTUVSD | -3237.71 | 13.05 | 6571.00 | 0.00 | 0 | 6499.52 | 0.00 |
| | SAC | -3237.71 | 13.24 | 6571.00 | 0.00 | 0 | 6499.52 | 0.00 |
| | Streak | -3237.71 | 13.24 | 6571.00 | 0.00 | 0 | 6499.52 | 0.00 |
| | 4-States | -3237.71 | 12.85 | 6571.00 | 0.00 | 0 | 6499.52 | 0.00 |
| B6 | SDM | -4087.32 | 9.71 | 8256.52 | 2.77 | 2 | 8194.70 | 2.77 |
| | HTSD | -4087.65 | 8.98 | 8257.20 | 3.44 | 2 | 8195.37 | 3.44 |
| | HTUVSD | -4087.32 | 10.04 | 8272.90 | 2.77 | 0 | 8198.73 | 2.77 |
| | SAC | -4085.93 | 11.67 | 8270.13 | 0.00 | 0 | 8195.95 | 0.00 |
| | Streak | -4085.93 | 11.67 | 8270.13 | 0.00 | 0 | 8195.95 | 0.00 |
| | 4-States | -4085.93 | 11.67 | 8270.13 | 0.00 | 0 | 8195.95 | 0.00 |
| B7 | SDM | -2821.05 | 10.24 | 5724.70 | 202.33 | 2 | 5662.15 | 202.27 |
| | HTSD | -2830.44 | 7.59 | 5743.48 | 221.11 | 2 | 5680.93 | 221.05 |
| | HTUVSD | -2821.05 | 10.24 | 5741.23 | 202.33 | 0 | 5666.18 | 202.27 |
| | SAC | -2719.93 | 10.90 | 5538.98 | 0.09 | 0 | 5463.93 | 0.03 |
| | Streak | -2719.91 | 15.55 | 5538.96 | 0.06 | 0 | 5463.91 | 0.00 |
| | 4-States | -2724.35 | 12.95 | 5547.83 | 8.93 | 0 | 5472.77 | 8.87 |

| Experiment | Modell | \mathcal{LL} | $Optim$ | BIC | G^2 | df | AIC_c | ΔAIC_c |
|------------|----------|----------------|---------|----------|--------|------|----------|----------------|
| B8 | SDM | -1424.45 | 12.53 | 2922.67 | 45.84 | 2 | 2869.03 | 45.84 |
| | HTSD | -1427.52 | 11.32 | 2928.82 | 51.99 | 2 | 2875.18 | 51.99 |
| | HTUVSD | -1424.45 | 12.53 | 2937.43 | 45.84 | 0 | 2873.09 | 45.84 |
| | SAC | -1401.52 | 14.23 | 2891.58 | 0.00 | 0 | 2827.24 | 0.00 |
| | Streak | -1401.52 | 14.22 | 2891.58 | 0.00 | 0 | 2827.24 | 0.00 |
| | 4-States | -1404.62 | 12.45 | 2897.77 | 6.19 | 0 | 2833.44 | 6.19 |
| B9 | SDM | -784.83 | 70.48 | 1643.44 | 7.24 | 2 | 1589.80 | 7.24 |
| | HTSD | -787.74 | 42.54 | 1649.26 | 13.05 | 2 | 1595.62 | 13.05 |
| | HTUVSD | -784.84 | 75.36 | 1658.21 | 7.25 | 0 | 1593.87 | 7.25 |
| | SAC | -781.21 | 73.26 | 1650.96 | 0.00 | 0 | 1586.62 | 0.00 |
| | Streak | -781.21 | 63.86 | 1650.96 | 0.00 | 0 | 1586.62 | 0.00 |
| | 4-States | -781.21 | 29.03 | 1650.96 | 0.00 | 0 | 1586.62 | 0.00 |
| C1 | SDM | -855.67 | 96.02 | 1819.26 | 8.00 | 3 | 1741.71 | 8.00 |
| | HTSD | -858.03 | 103.34 | 1823.98 | 12.72 | 3 | 1746.43 | 12.72 |
| | HTUVSD | -855.67 | 112.73 | 1840.85 | 8.00 | 0 | 1747.87 | 8.00 |
| | SAC | -851.67 | 111.75 | 1832.84 | 0.00 | 0 | 1739.86 | 0.00 |
| | Streak | -851.67 | 89.68 | 1832.84 | 0.00 | 0 | 1739.86 | 0.00 |
| | 4-States | -851.67 | 56.80 | 1832.84 | 0.00 | 0 | 1739.86 | 0.00 |
| C2 | SDM | -3723.71 | 15.25 | 7569.64 | 23.28 | 3 | 7477.56 | 23.28 |
| | HTSD | -3735.05 | 12.03 | 7592.32 | 45.96 | 3 | 7500.24 | 45.96 |
| | HTUVSD | -3723.71 | 15.26 | 7594.08 | 23.28 | 0 | 7483.62 | 23.28 |
| | SAC | -3712.07 | 18.49 | 7570.80 | 0.00 | 0 | 7460.34 | 0.00 |
| | Streak | -3712.07 | 18.49 | 7570.80 | 0.00 | 0 | 7460.34 | 0.00 |
| | 4-States | -3712.07 | 18.49 | 7570.80 | 0.00 | 0 | 7460.34 | 0.00 |
| C3 | SDM | -15258.13 | 19.84 | 30658.57 | 193.91 | 3 | 30546.29 | 193.91 |
| | HTSD | -15271.60 | 18.87 | 30685.52 | 220.86 | 3 | 30573.23 | 220.86 |
| | HTUVSD | -15258.13 | 19.85 | 30687.04 | 193.91 | 0 | 30552.30 | 193.91 |
| | SAC | -15161.17 | 23.91 | 30493.12 | 0.00 | 0 | 30358.39 | 0.00 |
| | Streak | -15161.17 | 23.91 | 30493.12 | 0.00 | 0 | 30358.39 | 0.00 |
| | 4-States | -15161.17 | 23.90 | 30493.12 | 0.00 | 0 | 30358.39 | 0.00 |
| D1 | SDM | -3190.38 | 28.03 | 6497.95 | 14.50 | 4 | 6408.86 | 9.67 |
| | HTSD | -3192.32 | 25.81 | 6501.84 | 18.38 | 4 | 6412.74 | 13.55 |
| | HTUVSD | -3190.46 | 39.36 | 6514.85 | 14.65 | 2 | 6413.04 | 9.82 |
| | SAC | -3185.54 | 45.85 | 6505.03 | 4.83 | 2 | 6403.22 | 0.00 |
| | Streak | -3217.86 | 25.46 | 6569.66 | 69.46 | 2 | 6467.85 | 64.64 |
| | 4-States | -3204.79 | 23.84 | 6543.52 | 43.33 | 2 | 6441.72 | 38.5 |
| D2 | SDM | -1376.53 | 20.02 | 2855.77 | 1.18 | 4 | 2781.33 | 0.46 |
| | HTSD | -1376.83 | 81.22 | 2856.38 | 1.79 | 4 | 2781.94 | 1.07 |
| | HTUVSD | -1376.62 | 61.63 | 2870.64 | 1.38 | 2 | 2785.61 | 0.65 |
| | SAC | -1376.30 | 54.46 | 2869.98 | 0.72 | 2 | 2784.95 | 0.00 |
| | Streak | -1376.97 | 45.65 | 2871.33 | 2.07 | 2 | 2786.30 | 1.34 |
| | 4-States | -1376.73 | 27.24 | 2870.86 | 1.59 | 2 | 2785.82 | 0.87 |

| Experiment | Modell | \mathcal{LL} | $Optim$ | BIC | G^2 | df | AIC_c | ΔAIC_c |
|------------|----------|----------------|---------|---------|--------|------|---------|----------------|
| D3 | SDM | -2081.21 | 20.41 | 4245.91 | 13.56 | 4 | 4184.54 | 3.56 |
| | HTSD | -2082.65 | 18.68 | 4248.80 | 16.45 | 4 | 4187.44 | 6.45 |
| | HTUVSD | -2082.05 | 21.57 | 4255.19 | 15.25 | 3 | 4188.26 | 5.25 |
| | SAC | -2080.71 | 24.58 | 4252.52 | 12.58 | 3 | 4185.59 | 2.58 |
| | Streak | -2085.96 | 24.58 | 4263.01 | 23.07 | 3 | 4196.08 | 13.07 |
| | 4-States | -2079.43 | 24.71 | 4249.94 | 10.00 | 3 | 4183.01 | 0.00 |
| D4 | SDM | -1812.45 | 23.47 | 3708.50 | 4.97 | 4 | 3647.03 | 0.00 |
| | HTSD | -1822.66 | 23.59 | 3728.92 | 25.39 | 4 | 3667.44 | 20.42 |
| | HTUVSD | -1813.67 | 25.19 | 3718.55 | 7.42 | 3 | 3651.50 | 2.45 |
| | SAC | -1813.58 | 27.71 | 3718.38 | 7.24 | 3 | 3651.32 | 2.27 |
| | Streak | -1832.87 | 27.71 | 3756.95 | 45.81 | 3 | 3689.89 | 40.84 |
| | 4-States | -1818.32 | 27.57 | 3727.85 | 16.72 | 3 | 3660.80 | 11.75 |
| D5 | SDM | -1280.33 | 16.49 | 2663.42 | 6.24 | 4 | 2588.94 | 4.23 |
| | HTSD | -1283.53 | 14.50 | 2669.81 | 12.63 | 4 | 2595.33 | 10.62 |
| | HTUVSD | -1280.38 | 17.32 | 2678.19 | 6.34 | 2 | 2593.11 | 4.32 |
| | SAC | -1278.22 | 21.48 | 2673.87 | 2.02 | 2 | 2588.80 | 0.01 |
| | Streak | -1279.16 | 21.58 | 2675.75 | 3.90 | 2 | 2590.68 | 1.89 |
| | 4-States | -1278.22 | 13.76 | 2673.87 | 2.01 | 2 | 2588.79 | 0.00 |
| E1 | SDM | -1519.85 | 15.50 | 3142.40 | 4.78 | 4 | 3067.97 | 2.26 |
| | HTSD | -1519.92 | 52.17 | 3142.55 | 4.94 | 4 | 3068.12 | 2.41 |
| | HTUVSD | -1519.79 | 37.97 | 3156.96 | 4.67 | 2 | 3071.93 | 2.14 |
| | SAC | -1518.83 | 62.94 | 3155.03 | 2.74 | 2 | 3070.01 | 0.22 |
| | Streak | -1520.66 | 44.43 | 3158.71 | 6.41 | 2 | 3073.68 | 3.89 |
| | 4-States | -1518.72 | 23.22 | 3154.82 | 2.53 | 2 | 3069.79 | 0.00 |
| E2 | SDM | -1033.09 | 13.39 | 2166.35 | 0.86 | 4 | 2094.52 | 0.26 |
| | HTSD | -1034.22 | 12.92 | 2168.60 | 3.12 | 4 | 2096.77 | 2.52 |
| | HTUVSD | -1033.15 | 14.32 | 2180.77 | 0.97 | 2 | 2098.73 | 0.37 |
| | SAC | -1032.96 | 15.34 | 2180.40 | 0.60 | 2 | 2098.36 | 0.00 |
| | Streak | -1045.05 | 13.13 | 2204.58 | 24.78 | 2 | 2122.54 | 24.19 |
| | 4-States | -1033.37 | 14.95 | 2181.21 | 1.41 | 2 | 2099.17 | 0.81 |
| E3 | SDM | -2488.03 | 17.46 | 5088.47 | 6.01 | 4 | 5004.19 | 0.00 |
| | HTSD | -2489.14 | 15.58 | 5090.71 | 8.24 | 4 | 5006.43 | 2.23 |
| | HTUVSD | -2488.49 | 17.90 | 5105.46 | 6.93 | 2 | 5009.16 | 0.93 |
| | SAC | -2488.36 | 19.64 | 5105.20 | 6.67 | 2 | 5008.89 | 0.66 |
| | Streak | -2504.43 | 17.67 | 5137.35 | 38.82 | 2 | 5041.04 | 32.81 |
| | 4-States | -2488.11 | 19.08 | 5104.69 | 6.17 | 2 | 5008.39 | 0.16 |
| E4 | SDM | -2453.94 | 14.52 | 5020.81 | 4.20 | 4 | 4936.02 | 0.00 |
| | HTSD | -2456.75 | 12.20 | 5026.43 | 9.82 | 4 | 4941.64 | 5.62 |
| | HTUVSD | -2455.46 | 14.44 | 5039.97 | 7.23 | 2 | 4943.09 | 3.03 |
| | SAC | -2454.52 | 16.54 | 5038.09 | 5.35 | 2 | 4941.21 | 1.15 |
| | Streak | -2481.71 | 14.46 | 5092.48 | 59.74 | 2 | 4995.59 | 55.54 |
| | 4-States | -2513.11 | 16.60 | 5155.28 | 122.54 | 2 | 5058.39 | 118.34 |

| Experiment | Modell | \mathcal{LL} | $Optim$ | BIC | G^2 | df | AIC_c | ΔAIC_c |
|------------|----------|----------------|---------|----------|--------|------|----------|----------------|
| E5 | SDM | -1012.29 | 9.18 | 2122.66 | 1.62 | 4 | 2052.96 | 0.77 |
| | HTSD | -1012.65 | 16.85 | 2123.40 | 2.35 | 4 | 2053.69 | 1.51 |
| | HTUVSD | -1012.50 | 13.96 | 2137.10 | 2.04 | 2 | 2057.49 | 1.2 |
| | SAC | -1011.90 | 13.41 | 2135.90 | 0.85 | 2 | 2056.30 | 0.00 |
| | Streak | -1013.77 | 17.21 | 2139.64 | 4.58 | 2 | 2060.03 | 3.74 |
| | 4-States | -1012.02 | 10.96 | 2136.14 | 1.08 | 2 | 2056.53 | 0.23 |
| E6 | SDM | -3277.17 | 12.95 | 6670.34 | 8.59 | 4 | 6582.44 | 8.09 |
| | HTSD | -3281.65 | 11.33 | 6679.31 | 17.56 | 4 | 6591.41 | 17.06 |
| | HTUVSD | -3277.29 | 13.74 | 6687.15 | 8.83 | 2 | 6586.72 | 8.34 |
| | SAC | -3273.56 | 14.87 | 6679.69 | 1.37 | 2 | 6579.25 | 0.87 |
| | Streak | -3274.42 | 13.37 | 6681.41 | 3.10 | 2 | 6580.98 | 2.60 |
| | 4-States | -3273.12 | 9.23 | 6678.81 | 0.50 | 2 | 6578.38 | 0.00 |
| E7 | SDM | -2578.27 | 16.72 | 5267.57 | 15.25 | 4 | 5184.70 | 0.00 |
| | HTSD | -2596.94 | 16.38 | 5304.90 | 52.58 | 4 | 5222.03 | 37.33 |
| | HTUVSD | -2584.86 | 18.63 | 5296.61 | 28.42 | 2 | 5201.92 | 13.18 |
| | SAC | -2579.24 | 19.81 | 5285.36 | 17.17 | 2 | 5190.67 | 1.92 |
| | Streak | -2582.40 | 18.33 | 5291.68 | 23.50 | 2 | 5197.00 | 8.25 |
| | 4-States | -2594.59 | 17.72 | 5316.05 | 47.87 | 2 | 5221.37 | 32.62 |
| E8 | SDM | -6021.45 | 18.31 | 12214.69 | 12.76 | 4 | 12083.05 | 12.76 |
| | HTSD | -6031.09 | 16.56 | 12233.98 | 32.05 | 4 | 12102.34 | 32.05 |
| | HTUVSD | -6021.48 | 19.36 | 12249.11 | 12.83 | 0 | 12091.18 | 12.83 |
| | SAC | -6015.07 | 20.27 | 12236.28 | 0.00 | 0 | 12078.35 | 0.00 |
| | Streak | -6015.07 | 17.96 | 12236.28 | 0.00 | 0 | 12078.35 | 0.00 |
| | 4-States | -6058.52 | 19.33 | 12323.19 | 86.90 | 0 | 12165.26 | 86.90 |
| E9 | SDM | -2295.60 | 13.08 | 4701.06 | 8.59 | 4 | 4619.36 | 7.68 |
| | HTSD | -2298.75 | 12.67 | 4707.36 | 14.89 | 4 | 4625.66 | 13.98 |
| | HTUVSD | -2293.94 | 13.66 | 4713.44 | 5.27 | 2 | 4620.09 | 4.36 |
| | SAC | -2291.76 | 15.06 | 4709.08 | 0.91 | 2 | 4615.73 | 0.00 |
| | Streak | -2295.49 | 12.19 | 4716.55 | 8.38 | 2 | 4623.20 | 7.48 |
| | 4-States | -2292.85 | 14.68 | 4711.26 | 3.09 | 2 | 4617.91 | 2.19 |
| E10 | SDM | -4823.16 | 13.27 | 9815.04 | 7.88 | 4 | 9686.51 | 7.88 |
| | HTSD | -4830.38 | 12.16 | 9829.48 | 22.32 | 4 | 9700.95 | 22.32 |
| | HTUVSD | -4822.73 | 13.23 | 9847.91 | 7.01 | 0 | 9693.72 | 7.01 |
| | SAC | -4819.22 | 15.07 | 9840.90 | 0.00 | 0 | 9686.71 | 0.00 |
| | Streak | -4819.22 | 13.16 | 9840.90 | 0.00 | 0 | 9686.71 | 0.00 |
| | 4-States | -4819.22 | 14.86 | 9840.90 | 0.00 | 0 | 9686.71 | 0.00 |
| E11 | SDM | -19216.35 | 10.99 | 38633.68 | 384.96 | 4 | 38472.73 | 384.96 |
| | HTSD | -19257.75 | 12.92 | 38716.48 | 467.76 | 4 | 38555.53 | 467.76 |
| | HTUVSD | -19216.35 | 17.66 | 38673.88 | 384.96 | 0 | 38480.74 | 384.96 |
| | SAC | -19023.86 | 14.12 | 38288.92 | 0.00 | 0 | 38095.78 | 0.00 |
| | Streak | -19023.86 | 11.20 | 38288.92 | 0.00 | 0 | 38095.78 | 0.00 |
| | 4-States | -19023.86 | 12.73 | 38288.92 | 0.00 | 0 | 38095.78 | 0.00 |

| Experiment | Modell | \mathcal{LL} | $Optim$ | BIC | G^2 | df | AIC_c | ΔAIC_c |
|------------|----------|----------------|---------|----------|-------|------|----------|----------------|
| E12 | SDM | -3370.68 | 10.68 | 6907.26 | 9.07 | 4 | 6781.57 | 9.07 |
| | HTSD | -3374.09 | 8.54 | 6914.07 | 15.89 | 4 | 6788.38 | 15.89 |
| | HTUVSD | -3370.71 | 10.62 | 6940.50 | 9.13 | 0 | 6789.72 | 9.13 |
| | SAC | -3366.14 | 11.42 | 6931.37 | 0.00 | 0 | 6780.59 | 0.00 |
| | Streak | -3366.14 | 9.05 | 6931.37 | 0.00 | 0 | 6780.59 | 0.00 |
| | 4-States | -3366.14 | 11.39 | 6931.37 | 0.00 | 0 | 6780.59 | 0.00 |
| F1 | SDM | -3958.44 | 30.56 | 8169.29 | 1.73 | 6 | 7977.30 | 1.73 |
| | HTSD | -3959.36 | 59.94 | 8171.12 | 3.56 | 6 | 7979.13 | 3.56 |
| | HTUVSD | -3958.48 | 63.61 | 8219.86 | 1.81 | 0 | 7989.57 | 1.81 |
| | SAC | -3957.58 | 50.90 | 8218.05 | 0.00 | 0 | 7987.75 | 0.00 |
| | Streak | -3957.58 | 48.30 | 8218.05 | 0.00 | 0 | 7987.75 | 0.00 |
| | 4-States | -3957.58 | 44.96 | 8218.05 | 0.00 | 0 | 7987.75 | 0.00 |
| F2 | SDM | -2257.19 | 33.28 | 4758.89 | 19.98 | 6 | 4574.93 | 19.98 |
| | HTSD | -2260.09 | 155.29 | 4764.69 | 25.78 | 6 | 4580.73 | 25.78 |
| | HTUVSD | -2257.36 | 92.84 | 4808.12 | 20.30 | 0 | 4587.49 | 20.30 |
| | SAC | -2247.20 | 72.63 | 4787.81 | 0.00 | 0 | 4567.19 | 0.00 |
| | Streak | -2247.23 | 94.59 | 4787.87 | 0.06 | 0 | 4567.24 | 0.06 |
| | 4-States | -2247.20 | 48.72 | 4787.81 | 0.00 | 0 | 4567.19 | 0.00 |
| G1 | SDM | -1597.91 | 10.83 | 3268.54 | 0.60 | 2 | 3215.97 | 0.60 |
| | HTSD | -1597.68 | 10.86 | 3268.09 | 0.14 | 2 | 3215.52 | 0.14 |
| | HTUVSD | -1597.61 | 12.21 | 3282.49 | 0.00 | 0 | 3219.44 | 0.00 |
| | SAC | -1597.61 | 12.88 | 3282.49 | 0.00 | 0 | 3219.44 | 0.00 |
| | Streak | -1597.61 | 12.88 | 3282.49 | 0.00 | 0 | 3219.44 | 0.00 |
| | 4-States | -1597.61 | 12.53 | 3282.49 | 0.00 | 0 | 3219.44 | 0.00 |
| G2 | SDM | -2458.97 | 9.25 | 4996.54 | 17.23 | 2 | 4938.02 | 17.23 |
| | HTSD | -2462.56 | 8.68 | 5003.72 | 24.42 | 2 | 4945.20 | 24.42 |
| | HTUVSD | -2458.97 | 9.33 | 5012.26 | 17.23 | 0 | 4942.06 | 17.23 |
| | SAC | -2450.35 | 10.94 | 4995.03 | 0.00 | 0 | 4924.82 | 0.00 |
| | Streak | -2450.35 | 10.94 | 4995.03 | 0.00 | 0 | 4924.82 | 0.00 |
| | 4-States | -2450.35 | 10.94 | 4995.03 | 0.00 | 0 | 4924.82 | 0.00 |
| H1 | SDM | -1755.33 | 15.82 | 3622.49 | 11.14 | 3 | 3540.95 | 11.14 |
| | HTSD | -1758.52 | 15.45 | 3628.86 | 17.51 | 3 | 3547.32 | 17.51 |
| | HTUVSD | -1755.34 | 16.00 | 3644.86 | 11.14 | 0 | 3547.07 | 11.14 |
| | SAC | -1749.76 | 18.67 | 3633.71 | 0.00 | 0 | 3535.93 | 0.00 |
| | Streak | -1749.76 | 18.66 | 3633.71 | 0.00 | 0 | 3535.93 | 0.00 |
| | 4-States | -1751.05 | 16.56 | 3636.28 | 2.57 | 0 | 3538.50 | 2.57 |
| H2 | SDM | -18013.60 | 10.51 | 36171.44 | 45.02 | 3 | 36057.23 | 45.02 |
| | HTSD | -18021.31 | 9.64 | 36186.86 | 60.44 | 3 | 36072.65 | 60.44 |
| | HTUVSD | -18013.60 | 10.52 | 36200.29 | 45.02 | 0 | 36063.25 | 45.02 |
| | SAC | -17991.09 | 13.13 | 36155.27 | 0.00 | 0 | 36018.23 | 0.00 |
| | Streak | -17991.09 | 13.13 | 36155.27 | 0.00 | 0 | 36018.23 | 0.00 |
| | 4-States | -17991.09 | 13.13 | 36155.27 | 0.00 | 0 | 36018.23 | 0.00 |

| Experiment | Modell | \mathcal{LL} | $Optim$ | BIC | G^2 | df | AIC_c | ΔAIC_c |
|------------|----------|----------------|---------|----------|-------|------|---------|----------------|
| H3 | SDM | -1520.19 | 13.10 | 3152.19 | 4.18 | 3 | 3070.65 | 4.18 |
| | HTSD | -1519.92 | 20.32 | 3151.66 | 3.65 | 3 | 3070.12 | 3.65 |
| | HTUVSD | -1519.32 | 16.03 | 3172.82 | 2.45 | 0 | 3075.04 | 2.45 |
| | SAC | -1518.09 | 22.39 | 3170.37 | 0.00 | 0 | 3072.59 | 0.00 |
| | Streak | -1518.09 | 23.37 | 3170.37 | 0.00 | 0 | 3072.59 | 0.00 |
| | 4-States | -1518.10 | 16.49 | 3170.38 | 0.00 | 0 | 3072.59 | 0.00 |
| I1 | SDM | -2595.79 | 16.13 | 5345.13 | 4.57 | 4 | 5231.97 | 4.57 |
| | HTSD | -2596.57 | 15.37 | 5346.69 | 6.13 | 4 | 5233.53 | 6.13 |
| | HTUVSD | -2595.75 | 16.41 | 5375.76 | 4.49 | 0 | 5240.05 | 4.49 |
| | SAC | -2593.50 | 18.18 | 5371.27 | 0.00 | 0 | 5235.56 | 0.00 |
| | Streak | -2593.50 | 16.22 | 5371.27 | 0.00 | 0 | 5235.56 | 0.00 |
| | 4-States | -2593.50 | 17.81 | 5371.27 | 0.00 | 0 | 5235.56 | 0.00 |
| I2 | SDM | -4898.64 | 10.85 | 9920.00 | 3.65 | 4 | 9825.35 | 0.00 |
| | HTSD | -4899.36 | 8.88 | 9921.44 | 5.09 | 4 | 9826.79 | 1.44 |
| | HTUVSD | -4899.09 | 11.83 | 9938.43 | 4.54 | 2 | 9830.27 | 0.90 |
| | SAC | -4899.70 | 13.17 | 9939.64 | 5.76 | 2 | 9831.48 | 2.11 |
| | Streak | -4943.38 | 11.29 | 10027.00 | 93.12 | 2 | 9918.84 | 89.47 |
| | 4-States | -4906.02 | 11.94 | 9952.29 | 18.41 | 2 | 9844.13 | 14.76 |
| J1 | SDM | -3550.75 | 28.07 | 7345.97 | 12.07 | 6 | 7162.04 | 12.07 |
| | HTSD | -3552.19 | 44.81 | 7348.86 | 14.96 | 6 | 7164.93 | 14.96 |
| | HTUVSD | -3549.43 | 47.60 | 7392.22 | 9.43 | 0 | 7171.63 | 9.43 |
| | SAC | -3544.71 | 49.49 | 7382.79 | 0.00 | 0 | 7162.20 | 0.00 |
| | Streak | -3544.71 | 49.42 | 7382.79 | 0.00 | 0 | 7162.20 | 0.00 |
| | 4-States | -3546.82 | 36.44 | 7387.01 | 4.22 | 0 | 7166.42 | 4.22 |

Literaturverzeichnis

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csak (Hg.), *Second international symposium on information theory*. Budapest, Hungary: Academia Kiado.
- Allen, D. M. (1974). The relationship between variable selection and data augmentation and a method of prediction. *Technometrics*, 16(1), 125–127.
- Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, 6(4), 451–474.
- Anderson, J. R., Bothell, D., Lebiere, C. & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language*, 38(4), 341–380.
- Anderson, J. R. & Lebiere, C. (1998). *The atomic components of thought*. Mahwah: Erlbaum.
- Ashby, F. G. (1992). *Multidimensional models of perception and cognition*. Hillsdale: Lawrence Erlbaum Associates.
- Ashby, F. G. & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, 93(2), 154–179.
- Atkinson, R. C. & Shiffrin, R. M. (1968). Human memory: a proposed system and its control processes. In K. W. Spence & J. T. Spence (Hg.), *The psychology of learning and motivation: Advances in research and theory* (S. 89–195). New York: Academic Press.
- Baddeley, A. D. (1986). *Working memory*. New York: Oxford University Press.
- Baddeley, A. D. (2002). The concept of episodic memory. In A. D. Baddeley, J. P. Aggleton & M. A. Conway (Hg.), *Episodic memory: New directions in research* (S. 1–10). New York: Oxford University Press.
- Baddeley, A. D. & Hitch, G. (1974). Working memory. In G. A. Bowers (Hg.), *Recent advances in learning and motivation* (Bd. 8, S. 47–89). New York: Academic Press.
- Baddeley, A. D. & Warrington, E. K. (1970). Amnesia and the distinction between long- and short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 9(2), 176–189.
- Bamber, D. & Van Santen, J. P. H. (1985). How many parameters can a model have and still be testable? *Journal of Mathematical Psychology*, 29(4), 443–473.
- Banks, W. P. (1970). Signal detection theory and human memory. *Psychological Bulletin*, 74(2), 81–99.

- Bastin, C. & Van der Linden, M. (2003). The contribution of recollection and familiarity to recognition memory: A study of the effects of test format and aging. *Neuropsychology*, *17*(1), 14–24.
- Bastin, C., Van der Linden, M., Michel, A.-P. & Friedman, W. J. (2004). The effects of aging on location-based and distance-based processes in memory for time. *Acta Psychologica*, *116*(2), 145–171.
- Bayen, U. J., Murnane, K. & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 197–215.
- Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, *10*, 12–21.
- Burnham, K. P. & Anderson, D. R. (2003). *Model selection and multimodel inference: a practical information theoretic approach*. New York: Springer.
- Calabrese, P., Markowitsch, H. J., Durwen, H. F., Widlitzek, H., Haupts, M., Holinka, B. et al. (1996). Right temporofrontal cortex as critical locus for ecphory of old episodic memories. *Journal of Neurology, Neurosurgery, & Psychiatry*, *61*(3), 304–310.
- Cary, M. & Reder, L. M. (2003). A dual-process account of the list-length and strength-based mirror effects in recognition. *Journal of Memory and Language*, *49*(2), 231–248.
- Cave, C. B. & Squire, L. R. (1992). Intact and long-lasting repetition priming in amnesia. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(3), 509–520.
- Cermak, L. S. & O'Connor, M. (1983). The anterograde and retrograde retrieval ability of a patient with amnesia due to encephalitis. *Neuropsychologia*, *21*(3), 213–234.
- Cohen, N. J. & Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science*, *210*(4466), 207–210.
- Collins, A. M. & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 240–247.
- Comblain, C., D'Argembeau, A., Van der Linden, M. & Aldenhoff, L. (2004). The effect of ageing on the recollection of emotional and neutral pictures. *Memory*, *12*(6), 673–684.
- Conover, W. J. (1999). *Practical nonparametric statistics*. New York: John Wiley & Sons.
- Cowan, N., Nugent, L. D., Elliott, E. M. & Geer, T. (2000). Memory-search and rehearsal processes and the word length effect in immediate recall: A synthesis in reply to Service. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *53A*(3), 666–670.
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302.
- Curran, H. V. & Hildebrandt, M. (1999). Dissociative effects of alcohol on recollective

- experience. *Consciousness & Cognition: An International Journal*, 8(4), 497–509.
- D'Argembeau, A. & Van der Linden, M. (2004). Identity but not expression memory for unfamiliar faces is affected by ageing. *Memory*, 12(5), 644–654.
- D'Argembeau, A., Van der Linden, M., Comblain, C. & Etienne, A.-M. (2003). The effects of happy and angry expressions on identity and expression memory for unfamiliar faces. *Cognition & Emotion*, 17(4), 609–622.
- D'Argembeau, A., Van der Linden, M., Etienne, A.-M. & Comblain, C. (2003). Identity and expression memory for happy and angry faces in social anxiety. *Acta Psychologica*, 114(1), 1–15.
- Davison, A. C. & Hinkley, D. V. (1997). *Bootstrap methods and their application*. New York: Cambridge University Press.
- Della Sala, S., Laiacona, M., Spinnler, H. & Trivelli, C. (1993). Autobiographical recollection and frontal damage. *Neuropsychologia*, 31(8), 823–839.
- Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*, 39(1), 1–38.
- Dewhurst, S. A. (2001). Category repetition and false recognition: Effects of instance frequency and category size. *Journal of Memory and Language*, 44(1), 153–167.
- Dewhurst, S. A. & Anderson, S. J. (1999). Effects of exact and category repetition in true and false recognition memory. *Memory & Cognition*, 27(4), 665–673.
- Dewhurst, S. A. & Farrand, P. (2004). Investigating the phenomenological characteristics of false recognition for categorised words. *European Journal of Cognitive Psychology*, 16(3), 403–416.
- Dewhurst, S. A. & Parry, L. A. (2000). Emotionality, distinctiveness and recollective experience. *European Journal of Cognitive Psychology*, 12, 541–551.
- Diana, R. A., Peterson, M. J. & Reder, L. M. (2004). The role of spurious feature familiarity in recognition memory. *Psychonomic Bulletin & Review*, 11(1), 150–156.
- Diana, R. A., Reder, L. M., Arndt, J. & Park, H. (2006). Models of recognition: A review of arguments in favor of a dual-process account. *Psychonomic Bulletin & Review*, 13(1), 1–21.
- Dodson, C. S., Prinzmetal, W. & Shimamura, A. P. (1998). Using Excel to estimate parameters from observed data: An example from source memory data. *Behavior Research Methods, Instruments, & Computers*, 30(3), 517–526.
- Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General*, 121(3), 275–277.
- Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition*, 24(4), 523–533.
- Dunn, J. C. (2004). Remember-know: A matter of confidence. *Psychological Review*, 111(2), 524–542.
- Ebbinghaus, H. (1885). *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie*. Darmstadt: Wissenschaftliche Buchgesellschaft.

- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7(1), 1–26.
- Efron, B. (1983). Estimating the error rate of a prediction rule: Improvement on cross-validation. *Journal of the American Statistical Association*, 78(382), 316–331.
- Efron, B. & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.
- Eldridge, L. L., Sarfatti, S. & Knowlton, B. J. (2002). The effect of testing procedure on remember-know judgments. *Psychonomic Bulletin & Review*, 9(1), 139–145.
- Erdfelder, E. (2000). *Multinomiale Modelle in der kognitiven Psychologie (Habilitationsschrift)*. Bonn: Philosophische Fakultät der Rheinischen Friedrich-Wilhelms-Universität in Bonn.
- Erdfelder, E. & Bayen, U. J. (2002). The Four-States model of memory retrieval experiences. In *43rd annual meeting of the psychonomic society* (Kansas City, Missouri Aufl.).
- Erdfelder, E. & Buchner, A. (1995). *Process dissociation measurement models: good versus better* (Bericht). Rheinische Friedrich-Wilhelms-Universität Bonn: Psychologisches Institut.
- Erdfelder, E., Cüpper, L. & Auer, T.-S. (2006). Multinomiale Verarbeitungsbaummodelle. In P. A. Joachim F.; Frensch (Hg.), *Handbuch der Allgemeinen Psychologie – Kognition* (S. 760–769). Göttingen: Hogrefe.
- Erdfelder, E., Cüpper, L., Auer, T. S. & Undorf, M. (Im Druck). The Four-States model of memory retrieval experiences. *Zeitschrift für Psychologie – Journal of Psychology*.
- Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition*, 10, 680–698.
- Gardiner, J. M. (2000). On the objectivity of subjective experiences and auto-noetic and noetic consciousness. In E. Tulving (Hg.), *Memory, consciousness, and the brain: The Tallinn conference* (S. 159–172). Philadelphia: Psychology Press.
- Gardiner, J. M. (2002). Episodic memory and auto-noetic consciousness: a first-person approach. In A. D. Baddeley, J. P. Aggleton & M. A. Conway (Hg.), *Episodic memory: New directions in research* (S. 11–30). New York: Oxford University Press.
- Gardiner, J. M. & Gregg, V. H. (1997). Recognition memory with little or no remembering: Implications for a detection model. *Psychonomic Bulletin & Review*, 4(4), 474–479.
- Gardiner, J. M. & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition*, 18(1), 23–30.
- Gardiner, J. M. & Java, R. I. (1991). Forgetting in recognition memory with and without recollective experience. *Memory & Cognition*, 19(6), 617–623.
- Gardiner, J. M. & Java, R. I. (1993). Recognition memory and awareness – an experimental approach. *European Journal of Cognitive Psychology*, 5(3), 337–346.
- Gardiner, J. M., Java, R. I. & Richardson-Klavehn, A. (1996). How level of proces-

- sing really influences awareness in recognition memory. *Canadian Journal of Experimental Psychology*, 50(1), 114–122.
- Gardiner, J. M., Kaminska, Z., Dixon, M. & Java, R. I. (1996). Repetition of previously novel melodies sometimes increases both remember and know responses in recognition memory. *Psychonomic Bulletin & Review*, 3(3), 366–371.
- Gardiner, J. M., Konstantinou, I., Karayianni, I. & Gregg, V. H. (2005). Memory awareness following speeded compared with unspeeded picture recognition. *Experimental Psychology*, 52(2), 140–149.
- Gardiner, J. M. & Parkin, A. J. (1990). Attention and recollective experience in recognition memory. *Memory & Cognition*, 18(6), 579–583.
- Gardiner, J. M., Ramponi, C. & Richardson-Klavehn, A. (1999). Response deadline and subjective awareness in recognition memory. *Consciousness & Cognition: An International Journal*, 8(4), 484–496.
- Gardiner, J. M., Ramponi, C. & Richardson-Klavehn, A. (2002). Recognition memory and decision processes: A meta-analysis of remember, know, and guess responses. *Memory*, 10(2), 83–98.
- Gardiner, J. M. & Richardson-Klavehn, A. (2000). Remembering and knowing. In E. Tulving & F. I. M. Craik (Hg.), *The Oxford handbook of memory* (S. 229–244). Oxford: Oxford University Press.
- Gardiner, J. M., Richardson-Klavehn, A. & Ramponi, C. (1997). On reporting recollective experiences and direct access to memory systems. *Psychological Science*, 8(5), 391–394.
- Gillund, G. & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, 91(1), 1–67.
- Glanzer, M., Adams, J. K., Iverson, G. J. & Kim, K. (1993). The regularities of recognition memory. *Psychological Review*, 100(3), 546–567.
- Glanzer, M., Kim, K., Hilford, A. & Adams, J. K. (1999). Slope of the receiver-operating characteristic in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(2), 500–513.
- 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., Squire, L. R. & Mandler, G. (1984). The information that amnesic patients do not forget. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(1), 164–178.
- Green, D. M. & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Oxford: John Wiley & Sons.
- Gregg, V. H. & Gardiner, J. M. (1994). Recognition memory and awareness: A large effect of study-test modalities on “know” responses following a highly perceptual orienting task. *European Journal of Cognitive Psychology*, 6(2), 131–147.
- Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin*, 75(6), 424–429.
- Griffiths, D. P., Dickinson, A. & Clayton, N. S. (1999). Declarative and episodic

- memory: What can animals remember about their past? *Trends in Cognitive Sciences*, 3(2), 74–80.
- Grünwald, P. (2000). Model selection based on minimum description length. *Journal of Mathematical Psychology*, 44(1), 133–152.
- Habib, R. (2001). On the relation between conceptual priming, neural priming, and novelty assessment. *Scandinavian journal of psychology*, 42, 187–195.
- Habib, R., Nyberg, L. & Tulving, E. (2003). Hemispheric asymmetries of memory: The HERA model revisited. *Trends in Cognitive Sciences*, 7(6), 241–245.
- Hager, W. (1992). *Jenseits von Experiment und Quasi-Experiment. Zur Struktur psychologischer Versuche und zur Ableitung von Vorhersagen*. Göttingen: Hogrefe.
- Hebb, D. E. (1949). *The organization of behavior: a neuropsychological theory*. New York: John Wiley & Sons.
- Herrmann, T. (1979). *Psychologie als Problem. Herausforderungen der psychologischen Wissenschaft*. Stuttgart: Klett-Cotta.
- Hirshman, E. & Master, S. (1997). Modeling the conscious correlates of recognition memory: Reflections on the remember-know paradigm. *Memory & Cognition*, 25(3), 345–351.
- Hu, X. (1999). Multinomial processing tree models: An implementation. *Behavior Research Methods, Instruments & Computers*, 31(4), 689–695.
- Huron, C. & Danion, J.-M. (2002). Impairment of constructive memory in schizophrenia. *International Clinical Psychopharmacology*, 17(3), 127–133.
- Huron, C., Danion, J.-M., Rizzo, L., Killofer, V. & Damiens, A. (2003). Subjective qualities of memories associated with the picture superiority effect in schizophrenia. *Journal of Abnormal Psychology*, 112(1), 152–158.
- Huron, C., Giersch, A. & Danion, J. M. (2002). Lorazepam, sedation, and conscious recollection: A dose-response study with healthy volunteers. *International Clinical Psychopharmacology*, 17(1), 19–26.
- Huron, C., Servais, C. & Danion, J.-M. (2001). Lorazepam and diazepam impair true, but not false, recognition in healthy volunteers. *Psychopharmacology*, 155(2), 204–209.
- Hurvich, C. M. & Tsai, C.-L. (1989). Regression and time series selection in small samples. *Biometrika*, 76(2), 297–307.
- Jacobs, A. M. & Grainger, J. (1994). Models of visual word recognition: Sampling the state of the art. *Journal of Experimental Psychology: Human Perception and Performance*, 20(6), 1311–1334.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30(5), 513–541.
- Jacoby, L. L. & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, 110(3), 306–340.
- Jacoby, L. L. & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General*, 118(2), 126–135.

- Jacoby, L. L. & Witherspoon, D. (1982). Remembering without awareness. *Canadian Journal of Psychology*, 36(2), 300–324.
- James, W. (1890). *The principles of psychology*. New York: Henry Holt and Co.
- Johnson, M. K., Hashtroudi, S. & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, 114(1), 3–28.
- Karayianni, I. & Gardiner, J. M. (2003). Transferring voice effects in recognition memory from remembering to knowing. *Memory & Cognition*, 31(7), 1052–1059.
- Kass, R. & Raftery, A. (1995). Bayes factors and model uncertainty. *Journal of the American Statistical Association*, 90, 773–795.
- Klauer, K. C. & Wegener, I. (1998). Unraveling social categorization in the Who said what? paradigm. *Journal of Personality and Social Psychology*, 75(5), 1155–1178.
- Knowlton, B. J., Ramus, S. J. & Squire, L. R. (1992). Intact artificial grammar learning in amnesia: Dissociation of classification learning and explicit memory for specific instances. *Psychological Science*, 3(3), 172–179.
- Konstantinou, I. & Gardiner, J. M. (2005). Conscious control and memory awareness when recognising famous faces. *Memory*, 13(5), 449–457.
- Kopelman, M. D., Stanhope, N. & Kingsley, D. (1999). Retrograde amnesia in patients with diencephalic, temporal lobe or frontal lesions. *Neuropsychologia*, 37(8), 939–958.
- Landauer, T. K. & Freedman, J. L. (1968). Information retrieval from long-term memory: Category size and recognition time. *Journal of Verbal Learning and Verbal Behavior*, 7(2), 291–295.
- Levine, B., Black, S. E., Cabeza, R., Sinden, M., McIntosh, A. R., Toth, J. P. et al. (1998). Episodic memory and the self in a case of isolated retrograde amnesia. *Brain: A Journal of Neurology*, 121(10), 1951–1973.
- Macho, S. (2002). Cognitive modeling with spreadsheets. *Behavior Research Methods, Instruments, & Computers*, 34(1), 19–36.
- Macho, S. (2004). Modeling associative recognition: A comparison of two-high-threshold, two-high-threshold signal detection, and mixture distribution models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(1), 83–97.
- Macmillan, N. A. & Creelman, C. D. (1996). Triangles in ROC space: History and theory of “nonparametric” measures of sensitivity and response bias. *Psychonomic Bulletin & Review*, 3(2), 164–170.
- MacRae, C. N., Schloerscheidt, A. M., Bodenhausen, G. V. & Milne, A. B. (2002). Creating memory illusions: Expectancy-based processing and the generation of false memories. *Memory*, 10(1), 63–80.
- Malamut, B. L., Saunders, R. C. & Mishkin, M. (1984). Monkeys with combined amygdalo-hippocampal lesions succeed in object discrimination learning despite 24-hour intertrial intervals. *Behavioral Neuroscience*, 98(5), 759–769.

- Manly, B. F. (1997). *Randomization and Monte Carlo methods in biology*. London: Chapman and Hall.
- Mantyla, T. (1997). Recollections of faces: Remembering differences and knowing similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(5), 1203–1216.
- Markowitsch, H. J., Calabrese, P., Liess, J., Haupts, M., Durwen, H. & Gehlen, W. (1993). Retrograde amnesia after traumatic injury of the fronto-temporal cortex. *Journal of Neurology, Neurosurgery & Psychiatry*, 56(9), 988–992.
- Maylor, E. A. & Mo, A. (1999). Effects of study-test modality on false recognition. *British Journal of Psychology*, 90(4), 477–493.
- Meiser, T. & Bröder, A. (2002). Memory for multidimensional source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(1), 116–137.
- Mitchell, K. J. & Johnson, M. K. (2000). Source monitoring: Attributing mental experiences. In E. Tulving & F. I. M. Craik (Hg.), *The Oxford handbook of memory* (S. 179–195). Oxford: Oxford University Press.
- Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology*, 44(1), 190–204.
- Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47(1), 90–100.
- Myung, I. J. & Pitt, M. A. (2002). Mathematical modeling. In H. Pashler & J. Wixted (Hg.), *Stevens' handbook of experimental psychology: Methodology in experimental psychology* (Bd. 4, S. 429–460). Hoboken: John Wiley & Sons.
- Navarro, D. J., Pitt, M. A. & Myung, I. J. (2004). Assessing the distinguishability of models and the informativeness of data. *Cognitive Psychology*, 49(1), 47–84.
- Nega, C. (2005). Perceptual effects and recollective experience in face recognition. *Experimental Psychology*, 52(3), 224–231.
- Nelson, K. (1993). The psychological and social origins of autobiographical memory. *Psychological Science*, 4(1), 7–14.
- Nocedal, J. & Wright, S. (2005). *Numerical optimization*. New York: Springer.
- O'Connor, M., Verfaellie, M. & Cermak, L. S. (1995). Clinical differentiation of amnesic subtypes. In A. D. Baddeley, B. A. Wilson & F. N. Watts (Hg.), *Handbook of memory disorders* (S. 53–80). New York: John Wiley & Sons.
- Parkin, A. J., Reid, T. K. & Russo, R. (1990). On the differential nature of implicit and explicit memory. *Memory & Cognition*, 18(5), 507–514.
- Parkin, A. J. & Walter, B. M. (1992). Recollective experience, normal aging, and frontal dysfunction. *Psychology and Aging*, 7(2), 290–298.
- Perner, P. & Ruffman, T. (1995). Episodic memory and autonoetic consciousness: Developmental evidence and a theory of childhood amnesia. *Journal of Experimental Child Psychology*, 59(3), 516–548.
- Peterson, L. & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology: General*, 58(3), 193–198.

- Pitt, M. A., Kim, W. & Myung, I. J. (2003). Flexibility versus generalizability in model selection. *Psychonomic Bulletin & Review*, 10(1), 29–43.
- Pitt, M. A., Kim, W., Navarro, D. J. & Myung, J. I. (2006). Global model analysis by parameter space partitioning. *Psychological Review*, 113(1), 57–83.
- Pitt, M. A., Myung, I. J. & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, 109(3), 472–491.
- Pollack, I. & Norman, D. A. (1964). A non-parametric analysis of recognition experiments. *Psychonomic Science*, 1(5), 125–126.
- Propper, R. E. & Christman, S. D. (2004). Mixed- versus strong right-handedness is associated with biases towards “remember” versus “know” judgements in recognition memory: Role of interhemispheric interaction. *Memory*, 12(6), 707–714.
- Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition*, 21(1), 89–102.
- Rajaram, S. (1996). Perceptual effects on remembering: Recollective processes in picture recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(2), 365–377.
- Rajaram, S. (1998). The effects of conceptual salience and perceptual distinctiveness on conscious recollection. *Psychonomic Bulletin & Review*, 5(1), 71–78.
- Rajaram, S. (1999). Assessing the nature of retrieval experience: Advances and challenges. In B. H. Challis & B. M. Velichkovsky (Hg.), *Stratification in cognition and consciousness* (S. 255–275). Amsterdam: J. Benjamins.
- Rajaram, S. & Geraci, L. (2000). Conceptual fluency selectively influences knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(4), 1070–1074.
- Rajaram, S., Hamilton, M. & Bolton, A. (2002). Distinguishing states of awareness from confidence during retrieval: Evidence from amnesia. *Cognitive, Affective & Behavioral Neuroscience*, 2(3), 227–235.
- Ramponi, C., Barnard, P. J. & Nimmo-Smith, I. (2004). Recollection deficits in dysphoric mood: An effect of schematic models and executive mode? *Memory*, 12(5), 655–670.
- Ratcliff, R., Sheu, C. & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological review*, 99(3), 518–535.
- Read, T. R. C. & Cressie, N. R. C. (1988). *Goodness-of-fit statistics for discrete multivariate data*. New York: Springer.
- Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6(6), 855–863.
- Reder, L. M., Angstadt, P., Cary, M., Erickson, M. A. & Ayers, M. S. (2002). A reexamination of stimulus-frequency effects in recognition: Two mirrors for low- and high-frequency pseudowords. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(1), 138–152.
- Reder, L. M., Nhoyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P. & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgments in a continuous recogni-

- tion paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(2), 294–320.
- Richardson-Klavehn, A., Gardiner, J. M. & Java, R. I. (1996). Memory: Task dissociations, process dissociations and dissociations of consciousness. In G. D. Underwood (Hg.), *Implicit cognition* (S. 85–158). Oxford: Oxford University Press.
- Rissanen, J. (1996). Fisher information and stochastic complexity. *IEEE Transactions on Information Theory*, 42(1), 40–47.
- Roberts, S. & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, 107(2), 358–367.
- Roediger, H. L., III, Marsh, E. J. & Lee, S. C. (2002). Kinds of memory. In D. Medin & H. Pashler (Hg.), *Stevens' handbook of experimental psychology: Memory and cognitive processes* (S. 1–42). Hoboken: John Wiley & Sons.
- Rotello, C. M., Macmillan, N. A. & Reeder, J. A. (2004). Sum-difference theory of remembering and knowing: a two-dimensional signal-detection model. *Psychological Review*, 111(3), 588–616.
- Rothkegel, R. (1999). AppleTree: A multinomial processing tree modeling program for macintosh computers. *Behavior Research Methods, Instruments & Computers*, 31(4), 696–700.
- Schacter, D. L. (1990). Perceptual representation systems and implicit memory: Toward a resolution of the multiple memory systems debate. *Annals of the New York Academy of Sciences*, 608, 543–571.
- Schwartz, G. (1978). Estimating the dimension of a model. *Annals of statistics*, 6, 461–464.
- Scoville, W. B. & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery & Psychiatry*, 20, 11–21.
- Service, E. (1998). The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 51A(2), 283–304.
- Shallice, T. & Warrington, E. K. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *Quarterly Journal of Experimental Psychology*, 22(2), 261–273.
- Shao, J. (1996). Bootstrap model selection. *Journal of the American Statistical Association*, 91, 655–665.
- Shao, J. & Tu, D. (1995). *The jackknife and bootstrap*. New York: Springer.
- Sharbaugh, C., Viet, S. M., Fraser, A. & McMaster, S. B. (2003). Comparable measures of cognitive function in human infants and laboratory animals to identify environmental health risks to children. *Environmental Health Perspectives*, 111(13), 1630–1639.
- Shimamura, A. P. & Squire, L. R. (1991). The relationship between fact and source memory: Findings from amnesic patients and normal subjects. *Psychobiology*, 19(1), 1–10.
- Sloman, S. A., Hayman, C. G., Ohta, N., Law, J. & al. et. (1988). Forgetting in primed

- fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(2), 223–239.
- Snodgrass, J. G. & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, 117(1), 34–50.
- Sonntag, P., Gokalsing, E., Olivier, C., Robert, P., Burglen, F., Kauffmann-Muller, F. et al. (2003). Impaired strategic regulation of contents of conscious awareness in schizophrenia. *Consciousness and Cognition: An International Journal*, 12(2), 190–200.
- Squire, L. R. (1992). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. *Journal of Cognitive Neuroscience*, 4(3), 232–243.
- Squire, L. R., Knowlton, B. & Musen, G. (1993). The structure and organization of memory. *Annual Review of Psychology*, 44, 453–495.
- Stahl, C. & Klauer, K. C. (Im Druck). HMMTree: A computer program for latent-class hierarchical multinomial processing tree models. *Behavior Research Methods, Instruments, & Computers*.
- Stegmüller, W. (1973). *Logische Analyse der Struktur ausgereifter physikalischer Theorien: "Non-statement view" von Theorien (Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie, Band II, Studienausgabe, Teil D)*. Berlin: Springer.
- Stone, M. (1974). Cross-validation choice and assessment of statistical predictions. *Journal of the Royal Statistical Society B*, 39, 44–47.
- Strack, F. & Förster, J. (1995). Reporting recollective experiences: Direct access to memory systems? *Psychological Science*, 6(6), 352–358.
- Toth, J. P. (1996). Conceptual automaticity in recognition memory: Levels-of-processing effects on familiarity. *Canadian Journal of Experimental Psychology*, 50(1), 123–138.
- Tukey, J. W. (1958). Bias and confidence in not quite large samples. *Annals of mathematical statistics*, 29, 614.
- Tulving, E. (1983). *Elements of episodic memory*. Oxford: Oxford University Press.
- Tulving, E. (1985a). How many memory systems are there? *American Psychologist*, 40(4), 385–398.
- Tulving, E. (1985b). Memory and consciousness. *Canadian Psychology*, 26(1), 1–12.
- Tulving, E. (1991). Concepts in human memory. In L. R. Squire, N. M. Weinberger, G. Lynch & J. L. McGaugh (Hg.), *Memory: Organization and locus of change* (S. 3–32). New York: Oxford University Press.
- Tulving, E. (1993). What is episodic memory? *Current Directions in Psychological Science*, 2(3), 67–70.
- Tulving, E. (1995). Organization of memory: Quo vadis? In M. S. Gazzaniga (Hg.), *The cognitive neurosciences* (S. 839–847). Cambridge, MA: MIT Press.
- Tulving, E. (2002a). Episodic memory and common sense: How far apart? In A. D.

- Baddeley, J. P. Aggleton & M. A. Conway (Hg.), *Episodic memory: New directions in research* (S. 269–287). New York: Oxford University Press.
- Tulving, E. (2002b). Episodic memory: From mind to brain. *Annual Review of Psychology*, 53(1), 1–25.
- Tulving, E. (2005). Episodic memory and autoeogenesis: Uniquely human? In H. S. Terrace & J. Metcalfe (Hg.), *The missing link in cognition: Origins of self-reflective consciousness* (S. 3–56). Oxford: Oxford University Press.
- Tulving, E. & Donaldson, W. (1972). *Organization of memory*. New York: Academic Press.
- Tulving, E., Hayman, C. A. & Macdonald, C. A. (1991). Long-lasting perceptual priming and semantic learning in amnesia: A case experiment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(4), 595–617.
- Tulving, E. & Markowitsch, H. J. (1998). Episodic and declarative memory: Role of the hippocampus. *Hippocampus*, 8(3), 198–204.
- Tulving, E., Markowitsch, H. J., Kapur, S., Habib, R. & Houle, S. (1994). Novelty encoding networks in the human brain: positron emission tomography data. *Neuroreport*, 5(18), 2525–2528.
- Tulving, E. & Schacter, D. L. (1990). Priming and human memory systems. *Science*, 247(4940), 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(4), 336–342.
- Viskontas, I. V., McAndrews, M. P. & Moscovitch, M. (2000). Remote episodic memory deficits in patients with unilateral temporal lobe epilepsy and excisions. *Journal of Neuroscience*, 20(15), 5853–5857.
- Warrington, E. K. & Weiskrantz, L. (1982). Amnesia: A disconnection syndrome? *Neuropsychologia*, 20(3), 233–248.
- Wehr, T. & Wippich, W. (2004). Typography and color: Effects of salience and fluency on conscious recollective experience. *Psychological Research/Psychologische Forschung*, 69(1–2), 138–146.
- Westermann, R. (1987). *Strukturalistische Theorienkonzeption und empirische Forschung in der Psychologie. Eine Fallstudie*. Berlin: Springer.
- Westmeyer, H. (1989). *Psychological theories from a structuralist point of view*. New York: Springer.
- Westmeyer, H. (1992). *The structuralist program in psychology: Foundations and applications*. Toronto: Hogrefe & Huber Publishers.
- Wheeler, M. A., Stuss, D. T. & Tulving, E. (1997). Toward a theory of episodic memory: The frontal lobes and autoeogenic consciousness. *Psychological Bulletin*, 121(3), 331–354.
- Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, 114(1), 152–176.
- Wixted, J. T. & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review*, 11(4), 616–641.

- Xu, M. & Bellezza, F. S. (2001). A comparison of the multimemory and detection theories of know and remember recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(5), 1197–1210.
- Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(6), 1341–1354.
- Yonelinas, A. P. (1999). Recognition memory ROCs and the dual-process signal-detection model: Comment on Glanzer, Kim, Hilford, and Adams (1999). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(2), 514–521.
- Yonelinas, A. P. (2001). Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology: General*, 130(3), 361–379.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46(3), 441–517.
- Yonelinas, A. P., Dobbins, I., Szymanski, M. D., Dhaliwal, H. S. & King, L. (1996). Signal-detection, threshold, and dual-process models of recognition memory: ROCs and conscious recollection. *Consciousness and Cognition: An International Journal*, 5(4), 418–441.
- Yonelinas, A. P. & Jacoby, L. L. (1994). Dissociations of processes in recognition memory: Effects of interference and of response speed. *Canadian Journal of Experimental Psychology*, 48(4), 516–534.
- Yonelinas, A. P. & Jacoby, L. L. (1995). The relation between remembering and knowing as bases for recognition – effects of size congruency. *Journal of Memory and Language*, 34(5), 622–643.
- Yonelinas, A. P., Kroll, N. E. A., Dobbins, I., Lazzara, M. & Knight, R. T. (1998). Recollection and familiarity deficits in amnesia: Convergence of remember-know, process dissociation, and receiver operating characteristic data. *Neuropsychology*, 12(3), 323–339.