

Diagnose: Confirmation bias

Wie die anfängliche Überzeugtheit von einer klinisch-psychologischen Prüfhypothese, die Berufserfahrung und das Fachwissen die Validität psychologischer Diagnosen beeinflussen.

Dissertation zur Erlangung des Doktorgrades
der Philosophischen Fakultät
der Christian-Albrechts-Universität
zu Kiel

Vorgelegt von
Britta Dreger

Kiel

15.03.2012

Erstgutachter: Prof. Dr. Günter Köhnken

Zweitgutachter: Prof. Dr. Thomas Bliesener

Tag der mündlichen Prüfung: 22.06.2012

Durch den zweiten Prodekan, Prof. Dr. Martin Krieger zum Druck genehmigt: 11.07.2012

Für Tino

Dank

Ich danke meinem Doktorvater Prof. Dr. Günter Köhnken sehr für seine Begleitung und die wertvollen Anmerkungen zu meiner Arbeit. Die Möglichkeit, meine Leidenschaft für die Psychologische Diagnostik mit Elementen aus der Rechtspsychologie verknüpfen zu können, bot mir außerordentliche Gestaltungsmöglichkeiten und Freiräume, die ich sehr zu schätzen weiß. Prof. Dr. Thomas Bliesener danke ich sehr dafür, dass er sich als Zweitgutachter zur Verfügung gestellt hat.

Katja vom Schemm hat mein wissenschaftliches Arbeiten von Anfang an begleitet und war mir stets eine konstruktiv-kritische und motivierende Kollegin. Auch aus der Ferne hatte sie während des Endspurts einen ordnenden Einfluss auf meine Gedanken und einen beruhigenden auf meine Emotionen. Ich danke Dir!

Allen Probanden sei recht herzlich für ihre Teilnahme an der zeitaufwändigen Studie gedankt. Ich rechne dies allen Studierenden und Berufstätigen sehr hoch an!

Dank gebührt insbesondere auch den Ratern aus den Reihen der studentischen Hilfskräfte, Forschungspraktikanten und Kolleginnen, die die vielen qualitativen Daten im Rahmen der Ratings aufbereitet haben. Carolin Sommer, Franz Möckel, Gesa Heinrich, Katrin Liethmann, Kerstin Limbrecht, Kerstin Neumann, Lena Abels, Milan Goetze, Sophie Kühne, Susann Roschelle, Swantje Maeß und Ursula Ziegenbruch, Euch sei herzlich für den unermüdlichen, zuverlässigen und zeitnahen Einsatz gedankt! Merle Hopert sei für die Diskussionen über das Studienmaterial und die Unterstützung bei der Rekrutierung sehr herzlich gedankt.

Bei Johannes Andres konnte ich meine statistischen Fragen unterbringen und mit der Hilfe von Eva Burchardt konnte ich einen gesunden Anspruch an mich entwickeln. Ihr schnelles Korrekturlesen war zudem „Gold“ wert. Silvia Gubi-Kelm danke ich insbesondere für das Mitleiden unter Gleichen!

Auch meinen Eltern und Schwiegereltern danke ich dafür, dass ihre Gedanken bei mir waren, als sie so oft auf mich verzichten mussten. Darüber hinaus geht eine Hochachtung an meinen Vater, der sich als Jurist der Aufgabe angenommen hat, meine Arbeit vorab zu lesen!

Ein ganz persönlicher Dank gebührt dem großartigen Team der Campus Krippe, das es mir ermöglicht hat, weitestgehend ohne schlechtes Gewissen Mutter und Promovierende zu sein.

Meinem Sohn Linus danke ich dafür, dass er so fröhlich, so mutig und so einzigartig ist. Du hast mich während der Endphase der Arbeit so unglaublich motiviert!

Aber am meisten danke ich meinem Mann Tino. Danke, dass Du mich durch alle Täler der Arbeit begleitet hast und stets Vertrauen hattest, dass ich den Spagat schaffe. Danke für die vielen Stunden Programmierarbeit und Lösung aller meiner technischen und grafischen Probleme! Danke, dass Du die unzähligen Wochenenden die Verantwortung als Kurzzeit-Alleinerziehender angenommen hast, um unserem Trotzkopf die Zeit zu vertreiben.

Tino, diese Arbeit gibt es, weil es Dich gibt!

Vorwort

„Die Aufgabe von Psychologen ist es, das Wissen über den Menschen zu vermehren und ihre Kenntnisse und Fähigkeiten zum Wohle des einzelnen und der Gesellschaft einzusetzen.“ (BDP, 1999).

Die vorliegende Arbeit hat es zum Ziel, mögliche Urteilsfehler und Verzerrungsmechanismen zu untersuchen, die im Zusammenhang mit dem Stellen der Diagnose einer psychischen Störung befürchtet werden müssen. Der Präambel der ethischen Richtlinien der DGPs und des BDP entsprechend, stellt dies eine angestrebte Vermehrung des Wissens dar, die sich konkret auf potentielle Schwachstellen des menschlichen Denken und Handelns im professionellen Kontext bezieht. Denn nach fester Überzeugung der Verfasserin ist nur dann ein verantwortungsvolles Handeln zum Wohl des einzelnen und der Gesellschaft möglich, wenn eine Fachdisziplin sich mit der eigenen Fehlbarkeit kritisch und wissenschaftlich auseinandersetzt, um aus den gewonnenen Erkenntnissen etwaige Maßnahmen zur Vermeidung von Fehlern abzuleiten. Hierzu soll mit der vorliegenden Arbeit ein kleiner Beitrag geleistet werden.

Aus Gründen der stilistischen Vereinfachung und der besseren Lesbarkeit wird in der vorliegenden Arbeit vorwiegend die maskuline Form der Bestimmungswörter verwendet. Personenbezogene Ausdrücke sind geschlechtsneutral zu verstehen und schließen immer männliche und weibliche Personen ein.

Inhaltsverzeichnis

1	Einleitung.....	1
2	Theoretischer und empirischer Hintergrund.....	7
2.1	Urteilsfehler und Urteilsverzerrungen als menschliches Versagen?.....	7
2.1.1	Kognitive Täuschungen und kognitive Verzerrungen	7
2.1.2	Der confirmation bias	9
2.1.3	Der Mensch als Hypothesentester	19
2.2	Ablauf einer Hypothesenprüfung	21
2.2.1	Modellvorstellungen zum Ablauf des (sozialen) Hypothesentestens	21
2.2.1.1	Modell des sozialen Hypothesentestens nach Trope und Liberman (1996)	22
2.2.1.2	Ablauf einer Hypothesenprüfung nach Hager und Weißmann (1991).....	24
2.2.2	Ablauf des diagnostischen Prozesses	26
2.2.3	Integration	28
2.3	Das Modell des konfirmatorischen Hypothesentestens	30
2.3.1	Überblick.....	30
2.3.2	Kognitive Verzerrungsprozesse.....	35
2.3.2.1	Überschätzung der a priori Wahrscheinlichkeit der Prüfhypothese	35
2.3.2.2	Selektive Speicherung bzw. selektiver Abruf hypothesenkonsistenter Evidenz.....	37
2.3.2.3	Hypothesenkonsistente Interpretation uneindeutiger Evidenz.....	40
2.3.2.4	Affirmation bias in der Inferenz aus der vorliegenden Evidenz.....	41
2.3.3	Konfirmatorische Teststrategie.....	43
2.3.3.1	Diagnostische vs. pseudodiagnostische Teststrategie	44
2.3.3.2	Positive vs. negative Teststrategie.....	47
2.3.3.3	Kombination von positiver und pseudodiagnostischer Teststrategie	56
2.3.4	Motivationale Prozesse.....	57

2.4	Ergänzungsmöglichkeiten.....	59
2.4.1	Faktoren innerhalb der Person des Hypothesentesters	59
2.4.1.1	Persönlichkeitsmerkmale	60
2.4.1.2	Intern-repräsentierter Anfangsverdacht	61
2.4.1.3	Der Experte als Hypothesentester	62
2.4.2	Art der zu prüfenden Hypothese.....	64
2.4.2.1	Valenz der Prüfhypothese	64
2.4.2.2	Devianzhypothesen: Übertragbarkeit auf andere Anwendungskontexte ..	66
2.5	Confirmation bias, Berufserfahrung und Fachwissen.....	69
2.6	Maßnahmen zur Vermeidung des confirmation bias	75
3	Ableitung der Fragestellung	85
3.1	Einflussvariablen	85
3.1.1	Überzeugtheit von der Prüfhypothese	85
3.1.2	Berufserfahrung und Fachwissen	87
3.2	Abhängige Variablen	87
3.2.1	AVn nach dem Modell des konfirmatorischen Hypothesentestens	87
3.2.2	AVn bzgl. der Bestätigungstendenzen nach Hager & Weißmann (1991).....	91
3.2.3	AV Verhaltenskonsequenz in Anlehnung an Haverkamp (1993)	92
3.3	Störvariablen.....	93
3.4	Inhaltliche Hypothesen	94
4	Operationalisierung	97
4.1	Einflussvariablen	97
4.1.1	Überzeugtheit von der Prüfhypothese	97
4.1.2	Berufserfahrung.....	98
4.1.3	Fachwissen	100
4.1.3.1	Fallmaterialien in der Bedingung Fachwissen relevant (Essstörung)	101
4.1.3.1.1	Wahl des Essstörungskontextes	101

4.1.3.1.2	Konstruktion und Eignungsprüfung der Fallbeschreibung	104
4.1.3.1.3	Konstruktion der Zusatzinformationen	105
4.1.3.2	Fallmaterialien in der Bedingung Fachwissen irrelevant (Brandstiftung)	108
4.1.3.2.1	Konstruktion und Eignungsprüfung der Fallbeschreibung	109
4.1.3.2.2	Konstruktion der Zusatzitems	111
4.2	Abhängige Variablen	111
4.2.1	AV1: Schlussfolgerung	111
4.2.2	AV 2: Subjektive Sicherheit	112
4.2.3	AV 3: Güte der Evidenzregeln	112
4.2.4	AV 4: Bestätigungstendenz B	115
4.2.5	AV 5: Hypothesenkonsistente Interpretation uneindeutiger Evidenz und Bestätigungstendenz I	119
4.2.6	AV 6: Verhaltenskonsequenzen	124
4.3	Störvariablen	124
4.4	Begründung der Wahl einer Online-Studie	125
4.5	Studienablauf	126
5	Ergebnisdarstellung	131
5.1	Vorbemerkungen	131
5.1.1	Eingesetzte Auswertungsverfahren	131
5.1.2	Wahl des Signifikanzniveaus	132
5.1.3	Bericht von Effektgrößen und p-Werten	133
5.1.4	Beurteilungsmaßstab für die Übereinstimmung von Raterurteilen	134
5.2	Voranalysen	135
5.2.1	Rücklaufquote bzw. Teilnahmebereitschaft	135
5.2.2	Ausschluss von Probanden	136
5.2.3	Analyse der Beurteilerüberstimmungen bzgl. der AVn	138
5.2.3.1	AV 3 Güte der Evidenzregeln	138

5.2.3.2	AV 6 Verhaltenskonsequenzen	140
5.2.4	Beschreibung der Stichprobe	140
5.2.4.1	Beschreibung der Stichprobe: Überzeugtheit von der Prüfhypothese	141
5.2.4.2	Beschreibung der Stichprobe: Berufserfahrung	141
5.2.4.2.1	Beschreibung der Erfahrenen.....	141
5.2.4.2.2	Vergleich mit der Population deutscher Psychotherapeuten	142
5.2.4.2.3	Beschreibung der Unerfahrenen	143
5.2.4.2.4	Vergleich mit der Population Psychologiestudierender in Deutschland.....	143
5.2.5	Beschreibung der Stichprobe: Fachwissen.....	144
5.3	Hypothesengeleitete Darstellung der Ergebnisse	145
5.3.1	PH-A _i : Verzerrtes Urteil	145
5.3.1.1	PH-A ₁ : Schlussfolgerung (AV 1)	145
5.3.1.2	PH-A ₂ : Subjektive Sicherheit (AV 2).....	148
5.3.2	PH-B _i : Asymmetrische Evidenzregeln	150
5.3.2.1	PH-B ₁ : Überwiegen der positiven Teststrategie (AV 3a)	150
5.3.2.2	PH-B ₂ : Überwiegen nicht diagnostischer Merkmale (AV 3b)	151
5.3.2.3	PH-B ₃ : Positive Evidenzregeln mit nicht diagnostischen Merkmalen	152
5.3.2.4	PH-B ₄ und PH-B ₅ : Berücksichtigung alternativer Hypothesen (AV 3c) .	152
5.3.2.4.1	PH-B ₄ : Anzahl berücksichtigter Hypothesen.....	152
5.3.2.4.2	PH-B ₅ : Überwiegen der Evidenzregeln zur Prüfhypothese	154
5.3.3	PH-C: Hypothesenkonsistente Interpretation uneindeutiger Evidenz	155
5.3.4	PH-D: Bestätigungstendenz I (AV 5b)	157
5.3.5	PH-E _i : Bestätigungstendenz B (AV 4).....	158
5.3.5.1	PH-E ₁ : Wichtigkeit	158
5.3.5.2	PH-E ₂ : Einfluss auf die Bewährung der Prüfhypothese	160
5.3.6	PH-F _i : Verhaltenskonsequenzen (AV 6)	162
5.3.6.1	PH-F ₁ : Themen der ersten Therapiesitzung	163
5.3.6.2	PH-F ₂ : Hausaufgaben.....	163

6 Diskussion.....	165
6.1 Interpretation der Ergebnisse der vorliegenden Studie.....	165
6.1.1 Ergebnisinterpretation: Bewertung der internen Validität	165
6.1.2 Ergebnisinterpretation: Überzeugtheit von der Prüfhypothese	166
6.1.2.1 PH-A _{ia} : Verzerrtes Urteil	166
6.1.2.2 PH-B _{ia} : Asymmetrische Evidenzregeln	168
6.1.2.3 PH-C: Hypothesenkonsistente Interpretation uneindeutiger Evidenz	171
6.1.2.4 PH-D: Bestätigungstendenz I.....	172
6.1.2.5 PH-E _{ia} : Bestätigungstendenz B	172
6.1.2.6 PH-F: Verhaltenskonsequenzen.....	172
6.1.3 Ergebnisinterpretation: Berufserfahrung	173
6.1.4 Ergebnisinterpretation: Fachwissen.....	177
6.1.5 Explorativ: Interaktion von Berufserfahrung und Fachwissen.....	179
6.2 Bewertung der Generalisierbarkeit der Ergebnisse	179
6.2.1 Teilnahmebereitschaft.....	179
6.2.2 Repräsentativität hinsichtlich der Variablen Berufserfahrung.....	180
6.2.3 Umsetzung als Online-Studie	182
6.3 Kritische Würdigung	182
6.4 Bedeutung der Ergebnisse für Praxis und Forschung.....	184
6.4.1 Implikationen für die klinische Praxis.....	184
6.4.2 Implikationen für die Forschung	185
Literaturverzeichnis	189
Abbildungsverzeichnis.....	201
Tabellenverzeichnis.....	205
Anhang A - Fallmaterialien Essstörung	209
Anhang B - Fallmaterialien Brandstiftung.....	239
Anhang C - Studie aus Probandensicht.....	273
Anhang D - Ratingmaterial für das OPUS.....	293
Lebenslauf.....	313

1 Einleitung

Die *Diagnose* ist für den klinisch tätigen Psychologen von außerordentlicher Bedeutung, dient sie doch ganz im Sinne ihrer sprachlichen Wurzeln im Griechischen dem *gründlichen Kennenlernen* bzw. *Entscheiden* und bildet somit die Basis jeglicher Intervention.

Durch den Abgleich von erhobenen Befunden mit umschriebenen Krankheitsbildern klassifiziert der Psychologe die Symptome des Patienten. Allerdings würde der fachlich kompetente Psychologe nicht sofort das Vorliegen einer Bulimie annehmen, wenn ihm bekannt wird, dass ein Patient mit großer Regelmäßigkeit Essanfälle hat, sondern erst noch Informationen zu weiteren Symptomen und alternativen Störungen einholen. In diesem Prozess kann man das *gründliche Kennenlernen* wiedererkennen.

Die Zuordnung der Diagnose ist jedoch kein Selbstzweck. Ihr Wert liegt vor allem darin, dass aufgrund der gründlichen Beschreibung des Patienten nunmehr relevante *Entscheidungen* für die Behandlung abgeleitet werden können. Die Wahl der richtigen Therapieelemente ist essentiell, weil man nicht davon ausgehen kann, dass die Konsequenz einer Falschbehandlung lediglich das Ausbleiben von Therapieerfolg ist, sondern sogar Schädigungen denkbar sind. Würde man bspw. jemandem ein Training der sozialen Kompetenzen zukommen lassen, bei dem das Vorliegen einer Antisozialen Persönlichkeitsstörung übersehen wurde, so vermehrt man die sozialen Fertigkeiten dieser Person, was in einer Zunahme ihrer manipulativen Verhaltensweisen bzw. deren Erfolg resultieren und letztlich Dritten schaden könnte.

Die Richtigkeit bzw. die Validität der Diagnose ist demnach eine notwendige - wenngleich noch keine hinreichende - Bedingung für Therapieerfolg, und muss oberste Priorität für den Diagnostiker haben.

Die Validität psychischer Diagnosen ist in der Geschichte der Psychologie bzw. Psychiatrie jedoch wiederholt in Frage gestellt worden. So spricht Saß (1999) von einer Krise der psychiatrischen Diagnostik Ende der 1930er Jahre in den USA, die sich bspw. darin offenbart habe, dass sich die Häufigkeit der Diagnose Schizophrenie zwischen verschiedenen Kliniken und Regionen extrem unterschieden habe. Da nicht anzunehmen ist, dass diese gefundenen Häufigkeiten Entsprechungen in der zugrundeliegenden Population haben, ist davon auszugehen, dass der Vorgang der Zuschreibung einer Person zu derselben Störungsklasse unterschiedlich gehandhabt worden sein muss. In diesem Fall kann nicht von der Validität der Diagnose gesprochen werden. Aufsehenerregend war in vergleichbarer Weise die Publikation von Rosenhan (1973), in der davon berichtet wird, dass gesunde Personen, die aufgrund der Vortäuschung von auditiven

Halluzinationen in verschiedene psychiatrischen Kliniken aufgenommen wurden, trotz ihres in Folge normalen Verhaltens nicht als gesund erkannt worden seien.

Über die potentiellen Fehlerquellen wurden viele Annahmen gemacht. So wurde bspw. vermutet, in der Person des Diagnostikers Verzerrungsquellen zu finden, sei es aufgrund seines Geschlechts oder der ethnischen Zugehörigkeit (z.B. Pottick, Kirk, Hsieh & Tian, 2007) oder eines Persönlichkeitsmerkmals wie der kognitiven Komplexität (z.B. Garb & Lutz, 2001; Spengler & Strohmmer, 1994; Walker & Spengler, 1995). Auch verschiedene Ansätze der Diagnostik selbst, bspw. kategoriale vs. behaviorale Diagnostik (vgl. Herbert, Nelson & Herbert, 1988) oder klinische vs. statistische Urteilsbildung (Grove, Zald, Lebow, Snitz, Nelson, 2000) sowie die verhaltenstherapeutische vs. analytische Therapieschule (vgl. Langer & Abelson, 1974) wurden ins Spiel gebracht.

Auch der menschliche Denkkaparat an sich steht in dem Verdacht, die Validität der Diagnose zu beeinträchtigen, denn häufig lassen sich Hinweise darauf finden, dass Menschen bspw. Schwierigkeiten mit dem Bayesianischen Denken haben (für eine Überblicksarbeit z.B. Cosmides & Tooby, 1996). Dies zeigt sich i.d.R. daran, dass Menschen bei der Abschätzung, wie wahrscheinlich es ist, dass eine Person *P* tatsächlich unter Krankheit *K* mit einer Prävalenzrate von 1 zu 1 000 leidet, wenn ein Test *T* mit einer Rate von 5% fälschlicherweise ein positives Ergebnis anzeigt, mehrheitlich die inkorrekte Antwort 95% geben. In ihrer Arbeit, die u.a. Ärzte als Probanden einschloss, berichten Casscells, Schoenberg und Graboys (1978), dass die korrekte Antwort (Chance von 2%) nur von 18% der Probanden gegeben wurde, wohingegen am häufigsten, nämlich von 34% der Probanden, die typische, falsche Antwort (s.o.) genannt wurde. Die Autoren folgern daraus, dass die Fähigkeit, Ergebnisse von Laboruntersuchungen zu verstehen, suboptimal ausgeprägt sei, was u.U. eine eingeschränkte Validität der resultierenden Diagnose nach sich ziehen könnte.

Viele weitere Aspekte des menschlichen Denkens bzw. der Informationsverarbeitung sind im Rahmen der (sozialen) Kognitionspsychologie als Verzerrungsquellen betrachtet worden. Abschnitt 2.1.2 dieser Arbeit gibt einen Überblick darüber. Der so genannte *confirmation bias* ist eine von ihnen. Im Sinne von Turk und Salovey (1985) ist darunter die Neigung zu einer Informationsverarbeitung zu verstehen, die solche Informationen bevorzugt, die einer (Vor-) Einstellung entsprechen. Die Bedeutsamkeit dieser Neigung für die Psychologie wurde von Nickerson (1998, p. 175) wie folgt beschrieben: „*If one were to attempt to identify a single problematic aspect of human reasoning that deserves attention above all others, the confirmation bias would have to be among the candidates for consideration*”

Nickerson beim Wort nehmend, soll in der vorliegenden Arbeit daher das Auftreten des *confirmation bias* im Zusammenhang mit dem Stellen einer klinisch-psychologischen Diagnose analysiert werden.

Eine Grundvoraussetzung für die Übertragbarkeit der Erkenntnisse der sozial-psychologischen Grundlagenforschung zum *confirmation bias* auf das Stellen einer klinisch-psychologischen Diagnose ist die Erkenntnis, dass dieser Vorgang einen Informationsverarbeitungsprozess im Rahmen einer interpersonellen Interaktion darstellt (vgl. Strong, 1968). Die Autorin führt daher in Abschnitt 2.2 die Argumentation, dass der diagnostische Prozess, d.h. alle Maßnahmen, die der Gewinnung von relevanten Informationen zum Treffen professioneller Entscheidungen dienen (Amelang & Zielinski, 2002), mit dem Ablauf des Testens einer sozialen Hypothese (Trope & Liberman, 1996) gleichzusetzen ist. Erkenntnisse darüber, wie Menschen Hypothesen der Art *Wählt mein neuer Nachbar die Grünen?* testen und welche Fehlerquellen dabei relevant sind, können demnach helfen, Verzerrungsmechanismen auch im klinischen Bereich zu beleuchten.

Deshalb wird in Abschnitt 2.3 eine ausführliche Darstellung des Modells des konfirmatorischen Hypothesentestens nach Schulz-Hardt und Köhnken (2000) vorgenommen. Die Entwicklung dieses Modells - basierend auf Befunden der Sozialpsychologie - ist im Kern motiviert durch die dramatischen Geschehnisse im Zusammenhang mit Aufsehen erregenden Gerichtsprozessen wegen vermeintlichen sexuellen Missbrauchs von Kindern in den USA, England und Europa in den 1980er und 1990er Jahren. Schulz-Hardt und Köhnken (2000) veranschaulichen den *confirmation bias* anhand des Montessori-Prozesses, im Zuge dessen dem Erzieher eines Kindergartens 55 fortgesetzte Handlungen des sexuellen Kindesmissbrauch und der körperlichen Misshandlung von Kindergartenkindern vorgeworfen wurden. Seinen Ursprung hatte der Prozess in der Aussage des damals vierjährigen M., der auf die Frage einer Freundin der Familie, was er bekomme, wenn er Fieber habe, antwortete: „*Der R. [der Erzieher] steckt mir den Finger in den Po.*“ Obwohl M. nicht auf weitere Fragen zu der Aussage einging, wurde die Kindergartenleitung informiert und Elternabende abgehalten, die die anderen Eltern über die Anschuldigung informierten. Bei der Kinderschutzorganisation, bei der die Freundin der Familie arbeitete, war man zu diesem Zeitpunkt davon überzeugt: „*Wenn ein Kind betroffen ist, dann müssen auch mehrere betroffen sein.*“ Die meisten Eltern befragten ihre Kinder dann solange und unter Einsatz von Druck bzw. Belohnung, bis Aussagen über den erwarteten Missbrauch gemacht wurden. Auch im Laufe der Ermittlungsarbeit machten die Kinder immer detailreichere, aber teilweise auch immer abstrusere Angaben.

Dies betrachten die Autoren als Selbstbestätigung im Sinne des *confirmation bias*, da das Verhalten der Eltern darauf schließen lässt, dass sie von dem Verdacht des Missbrauchs bereits überzeugt waren, von den Kindern lediglich noch die erwartete Bekräftigung einholten und dabei suggestiv vorgingen. Letzteres entwertete die Aussagen der Kinder jedoch, da so eine wissenschaftlich fundierte aussagepsychologische Begutachtung der Glaubhaftigkeit unmöglich wurde, weswegen der Erzieher letztendlich freizusprechen war.

Nickersons (1998) Hinweis auf die Notwendigkeit, gerade den *confirmation bias* besonders zu beachten, wird an diesem Beispiel auf drastische Art und Weise veranschaulicht, da der Prozess ausschließlich Verlierer zurückgelassen und eine Wahrheitsfindung unmöglich gemacht hat.

Wenngleich man unterstellen kann, dass die Tragweite von invaliden klinisch-psychologischen Diagnosen - verursacht durch einen *confirmation bias* - geringer ist als diejenige im Montessori-Prozess, so ist eine Nicht- oder Falschbehandlung doch zumindest für den Betroffenen und sein direktes soziales Umfeld eine persönliche Tragödie. Es wäre daher fahrlässig einfach anzunehmen, dass der Diagnostiker aufgrund seines professionellen Hintergrunds per se vor dem *confirmation bias* gefeit ist. Aus Sicht der Verfasserin ist dieser Nachweis empirisch zu erbringen.

Die einschlägigen Arbeiten werden daher in Abschnitt 2.3 referiert und kritisch bewertet. Daraus wird ableitbar sein, dass ein Mangel an Arbeiten herrscht, die die individuelle Überzeugtheit des Hypothesentesters von einer an ihn herangetragenen Prüfhypothese berücksichtigen. In den experimentellen Arbeiten wurden i.d.R. seitens des Versuchsleiters zu prüfende Hypothesen vorgegeben und stillschweigend unterstellt, dass die Probanden in diesem Sinne agiert haben. Die Nachteile dieses Paradigmas werden diskutiert.

Bislang wurde die Vergleichbarkeit der Situationen beim Hypothesentesten im Alltag und in der Diagnostik betont. Selbstverständlich sollen aber auch die Unterschiede zwischen Laien (z.B. dem Hausbewohner, der sich fragt, ob der neue Mieter die Partei die Grünen wählt) und Experten (z.B. dem Psychologen, der das Vorliegen einer Bulimie überprüfen muss) nicht übergangen werden. Im Unterschied zum Testen der sozialen Hypothesen im Alltag, bei der sich die Hypothesentester in der Mehrheit nicht bedeutsam hinsichtlich ihrer Erfahrungen und ihrem Wissen über die hypostasierten Zusammenhänge unterscheiden werden, gibt es zwischen Diagnostikern, die mit der Abklärung einer klinisch-psychologischen Störung beauftragt werden, durchaus relevante Unterschiede im Hinblick auf die Berufserfahrung und das spezifische Fachwissen. Es ist unmittelbar einsichtig, dass der Psychotherapeut, der bereits seit 25 Jahren praktiziert,

einen anderen Erfahrungshintergrund für die Diagnose mitbringt, als derjenige, der gerade die theoretischen Seminare zur Diagnostik im Psychologiestudium absolviert hat. Die Befunde hinsichtlich der Frage, ob die Berufserfahrung daher ein Schutz- oder gar ein Risikofaktor ist, werden in Abschnitt 2.5 diskutiert.

Zusammenfassend kann festgehalten werden, dass es das Ziel der vorliegenden Arbeit ist, das Auftreten des Phänomens des *confirmation bias* im Kontext des Stellens einer klinisch-psychologischen Diagnose durch Berufserfahrene zu analysieren, um dabei die Übertragbarkeit des Modells des konfirmatorischen Hypothesentestens (Schulz-Hardt & Köhnken, 2000) zu überprüfen. Darüber hinaus soll die Wahrscheinlichkeit eines selbstbestätigenden Urteils in Abhängigkeit von der subjektiven Überzeugtheit in die Prüfhypothese, der Dauer der Berufserfahrung und dem fallspezifischen Wissen untersucht werden.

2 Theoretischer und empirischer Hintergrund

Im Folgenden werden die theoretischen Annahmen und empirischen Befunde vorgestellt und diskutiert, die für die vorliegende Arbeit relevant sind. Einleitend wird klargestellt, dass der Mensch trotz der in der Empirie umfassend bestätigten Urteilsfehler – wie zum Beispiel des *confirmation bias* – in der Lage ist, als Hypothesentester zu fungieren (2.1). Anschließend werden drei Modelle des Ablaufs des Hypothesentestungsprozesses vorgestellt (2.2), bevor auf die Konzeptionierung des konfirmatorischen Hypothesentestens nach Schulz-Hardt und Köhnken (2000) eingegangen werden kann (2.3), zu dem nachfolgend Ergänzungsmöglichkeiten diskutiert werden (2.4). Auf den konkreten Anwendungsfall der vorliegenden Arbeit bezogen ist sodann der Abschnitt 2.5, der den Zusammenhang von *confirmation bias*, Berufserfahrung und Fachwissen thematisiert. Im letzten Abschnitt wird die Vermeidung der Selbstbestätigung (2.6) thematisiert.

2.1 Urteilsfehler und Urteilsverzerrungen als menschliches Versagen?

Das erste Ziel ist es, einen Überblick über die möglichen Urteilsfehler zu geben, die in der Fachliteratur diskutiert werden. Auf den *confirmation bias* als zentralem Urteilsfehler der vorliegenden Arbeit wird dabei das Hauptaugenmerk gerichtet.

2.1.1 Kognitive Täuschungen und kognitive Verzerrungen

Die alltäglichen Erfahrungen eines jeden Einzelnen bestätigen immer wieder, dass die bereits aus der Antike bekannte Weisheit *Irren ist menschlich!* auch heute noch Gültigkeit besitzt. Zentral für die vorliegende Arbeit sind genau solche Fehler, die eine fehlerhafte Wahrnehmung bzw. Erkenntnis der Realität beinhalten. In der Forschung zur sozialen Kognition bzw. sozialen Informationsverarbeitung werden diese als *cognitive illusions* (vgl. Pohl, 2004) oder *cognitive biases* bezeichnet (Caverni, Fabre & Gonzalez, 1990). Ihnen allen ist nach Pohl (2004) gemeinsam, dass sie uns vor Augen führen, dass die Menschen in ihrem Denken von dem objektiven Standard abweichen.

Pohl (2004) ist zu entnehmen, dass der Begriff *cognitive illusion* – im Deutschen besser *kognitive Täuschung* – an denjenigen der *optischen Täuschung* angelehnt ist. Der Autor beschreibt vier Bestimmungsstücke. Erstens führt eine kognitive Täuschung zu einer Wahrnehmung, einem Urteil oder einer Erinnerung, die zuverlässig von der Realität

abweicht, wobei diese Abweichung zweitens systematisch, d.h. in einer vorhersagbaren Richtung erfolgt. Außerdem treten die Illusionen drittens unfreiwillig und damit meist unbewusst auf und sind daher viertens nur schwerlich – wenn überhaupt – zu vermeiden.

Caverni et al. (1990) definieren *cognitive biases* – im Deutschen am ehesten als *kognitive Verzerrungen zu übersetzen* – in ganz ähnlicher Weise, indem u. a. betont wird, dass gelegentliche und zufällige Fehler nicht als Verzerrungen zu interpretieren sind. Außerdem machen sie klar, dass dasjenige, was nach heutiger Erkenntnis als *bias* gilt, diesen negativen Status schnell verlieren kann, wenn sich nämlich die Theorie, d.h. der derzeitige Rahmen für die Definition von Norm und Normabweichung, durch neue Erkenntnisse ändert.

Die untersuchten Phänomene beinhalten durchaus sehr unterschiedliche Aspekte. Einerseits sind kognitive Täuschungen im Bereich des Denkens, d.h. beim Anwenden einer Regel zu beobachten, also bspw. dann, wenn Personen eine Wahrscheinlichkeit schätzen, eine Hypothese testen oder eine versteckte Regel finden müssen. Genannt seien an dieser Stelle der *confirmation bias* (Oswald & Grosjean, 2004) und die *illusory correlation* (Fiedler, 2004; Fiedler, Hemmeter & Hofmann, 1984).

Täuschungen beim *Urteilen* sind nach Pohl (2004) hingegen diejenigen, die auftreten, wenn Menschen unter Unsicherheit eine subjektive Einschätzung eines präsentierten Stimulusobjekts abgeben sollen. Klassischerweise kommen in diesem Zusammenhang Urteilsheuristiken zum Einsatz, die nach Strack und Deutsch (2002) als einfache Faustregeln auf leicht zu erhaltende Informationen angewendet werden und unter geringem Verarbeitungsaufwand ein hinreichend genaues Urteil ermöglichen. Wie die Umschreibung *hinreichend genau* es aber bereits andeutet, können sie auch eine Urteilsverzerrung nach sich ziehen. Bekannt sind hier v.a. die Verfügbarkeitsheuristik (Tversky & Kahnemann, 1974), die Repräsentativitätsheuristik (Kahnemann & Tversky, 1972) und der Ankereffekt (vgl. Englich, Mussweiler & Strack, 2006; Gigerenzer & Engel, 2006; Tversky & Kahnemann, 1974).

Auch in Bezug auf das Erinnern sind Täuschungen möglich, d.h. immer dann, wenn zuvor enkodiertes Material zu einem späteren Zeitpunkt erinnert und abgerufen werden muss. Untersucht werden in diesem Zusammenhang Effekte eines *labellings* (Pohl, 2004), der *missinformation effect* (Loftus & Palmer, 1974) und der *hindsight bias* (Pohl, 2004).

In der vorliegenden Arbeit steht mit dem *confirmation bias* eine kognitive Täuschung im Mittelpunkt, die dem Bereich des Denkens zuzuordnen ist, da sein Auftreten mit dem Testen von Hypothesen verknüpft ist. Es schließt sich daher nun eine eingehendere Betrachtung des Phänomens an. Zuvor sei allerdings angemerkt, dass mit der Fokussierung

auf genau diese kognitive Täuschung die Bedeutsamkeit der anderen Verzerrungen nicht herabgesetzt werden soll, denn es ist unmittelbar einsichtig, dass auch beim Testen einer Hypothese Verzerrungen des Erinnerns wirksam werden können, da Speicherung und Abruf von Informationen hierfür ebenfalls relevant sind.

2.1.2 *Der confirmation bias*

Der Begriff *confirmation bias* geht auf Wason (1960) zurück und wird von Turk und Salovey (1985) als Neigung zu einer Informationsverarbeitung verstanden, die solche Informationen bevorzugt, die einer oder mehreren (Vor-)Einstellungen entsprechen. Im deutschen Sprachraum wird der *confirmation bias* als *Bestätigungstendenz* (Bördlein, 2000; Effler, 1986; Gadenne & Oswald, 1986) oder *Bestätigungsfehler* (Gadenne, 1982; Fiedler, 1983) übersetzt. Bevor eine theoretische Annäherung vorgenommen werden kann, werden zunächst einschlägige empirischen Arbeiten zum *confirmation bias* präsentiert. Da die Anfänge der Forschung zur Bestätigungstendenz im Bereich der (sozialen) Kognitionspsychologie gemacht wurden, wird mit diesem Themengebiet begonnen.

Der confirmation bias in der (sozialen) Kognitionsforschung. Auf empirischer Ebene konnte einerseits im Rahmen kognitionspsychologischer Untersuchungen zur Konzeptbildung und andererseits im Bereich des Testens von sozialen Hypothesen über die Persönlichkeit von unbekanntem Personen gezeigt werden, dass Personen dazu neigen, die zu testende Hypothese auch zu bestätigen.

Der kognitionspsychologische Forschungsstrang zum *confirmation bias* bezieht sich vornehmlich auf die Frage, inwieweit Personen – spezifiziert als intelligente junge Erwachsene – bei der Lösung einer konzeptuellen Aufgabe entweder ausschließlich bestätigende Informationen oder doch sowohl bestätigende als auch widerlegende Informationen suchen, wie es im wissenschaftlichen Sinne der Hypothesenprüfung durch Widerlegung in Anlehnung an Popper (1959) als Goldstandard angesehen wird. Da die Probanden von Wason (1960) in seinem *rule discovery task paradigm* mehrheitlich versuchten, ihre selbst generierte Hypothese über die Regel zur Bildung eines Zahlentripels zu verifizieren anstelle sie zu falsifizieren (s. auch Abschnitt 2.3.3.2 zur Erläuterung des *rule discovery tasks*), sprach Wason von einem *confirmation bias*. Später zeigten Wason und Johnson-Laird (1972) ein äquivalentes Vorgehen der Versuchspersonen bei der Testung der logischen Verknüpfung *Wenn P, dann Q*.

Die Sozialpsychologie interessierte sich im Zusammenhang mit dem Testen sozialer Hypothesen über die Persönlichkeit von unbekanntem Personen für den *confirmation bias*. Snyder und Swann stellten sich 1978 die Frage, mit welcher Strategie Personen eine Hypothese über die Persönlichkeit einer Zielperson testen. Das klassische Untersuchungsparadigma stellte ihre Probanden entweder vor die Aufgabe, herauszufinden, ob eine Zielperson extravertiert ist oder ob sie introvertiert ist. Die Probanden suchten durch die Auswahl von Fragen aus einer vorgefertigten Liste in der Folge vornehmlich solche Informationen, die geeignet waren, die Hypothese zu bestätigen.

Der confirmation bias in der rechtspsychologischen Forschung. Da sozialpsychologische Konzeptionen in vielen psychologischen Anwendungsfeldern eine Rolle spielen, wurde die Forschung nach und nach auf diese übertragen. Eine besondere Beachtung fand der *confirmation bias* dabei in der rechtspsychologischen Forschung, was nach Ansicht der Verfasserin darin begründet sein dürfte, dass im juristischen Kontext in der Regel über Normverletzungen und deren Ahndung zu entscheiden ist. Urteilsfehler wie der *confirmation bias* können in diesem Zusammenhang fatale Konsequenzen für den zu Unrecht Angeklagten oder Verurteilten haben – im Zweifelsfall sogar letale.

So konnten Lord, Ross und Lepper (1979) zeigen, dass die Beschäftigung mit Argumenten für *und* gegen einen Sachverhalt (hier die Todesstrafe), die die betreffende Person entweder befürwortet oder ablehnt, nicht etwa zu einer Relativierung der vorherigen Meinung führt, sondern vielmehr zu ihrer Verfestigung. Konfrontiert mit zwei sich widersprechenden Arbeiten bewerteten die Probanden jeweils die mit ihrer Meinung konforme Studie als methodisch gelungener und betonten die methodischen Schwachpunkte der widersprechenden Studie.

Bei Ammann (1999) und Baer (2002) schließlich befanden mehr Probanden einen Sportlehrer (in der Studie von Ammann) bzw. einen Zauberer (bei Baer) des sexuellen Missbrauchs für schuldig, wenn ihnen nach der Vorgabe der Videoaufnahme der Turnstunde bzw. der Tonbandaufnahme des Berichts über die Zaubervorstellung durch ein Kindergartenkind der Hinweis präsentiert wurde, gegen den Akteur werde strafrechtlich ermittelt. Der Verdacht hatte sich selbst bestätigt, da in der Kontrollgruppe mit identischem Material nur selten ein vergleichbarer Vorwurf erhoben wurde.

Die Arbeit von Semin und DePoot (1997) ergab, dass die Erwartung einer Person über die Glaubwürdigkeit einer Zielperson die Formulierung von Interviewfragen dahingehend beeinflusste, ob die Zielperson als aktiv Handelnde dargestellt wurde oder nicht. Dies wurde von den Autoren daran festgemacht, ob eher solche Fragen ausgewählt werden, in denen die Zielperson das Subjekt im Rahmen des Satzbaus einnahm (*Did you dance*

with Peter?) oder die Objektposition im Sinne einer passiven Rolle (*Did Peter dance with you?*). Personen, die in der Rolle des polizeilichen Ermittlers von der Glaubwürdigkeit des Opfers einer mutmaßlichen Vergewaltigung überzeugt waren, wählten Fragen aus, die den mutmaßlichen Täter als Akteur darstellen, wohingegen Personen, die von einer geringen Glaubwürdigkeit ausgingen, eher Fragen wählten, die nahe legten, dass das vermeintliche Opfer die aktivere Rolle übernommen hat. Hier zeigte sich also nicht nur in einem anderen Kontext eine Bestätigungstendenz, sondern vielmehr ist aus den Befunden ableitbar, dass die Auswirkungen einer Meinung oder Erwartung nicht auf die Einstellung der betreffenden Person begrenzt sind, sondern auch in die Interaktion (hier die Befragungssituation) mit relevanten Personen hineinreichen, wobei einschränkend anzumerken ist, dass es sich bei der Arbeit um eine PC-gestützte Studie handelte und keine reale Interviewsituation umgesetzt wurde.

Kassin, Goldstein und Savitsky (2003) ließen ihre Probanden ebenfalls in die Rolle von polizeilichen Ermittlern schlüpfen und zeigten den Einfluss einer Schuld- bzw. einer Unschuldsvermutung erstens auf die Auswahl von Fragen, zweitens den Einsatz vorgelegter Verhörtechniken aus dem Arsenal der *Reid-Technik* (vgl. Inbau, Reid, Buckley & Jayne, 2001) und drittens das Ausmaß des ausgeübten Drucks auf eine Zielperson, die eines (Schein-) Diebstahls verdächtigt wird. Im Sinne des *confirmation bias* zeigte sich auf den drei genannten Ebenen ein negativer Effekt der induzierten Schuldannahme. Hill, Memon und McGeorge (2008) erweiterten diese Befunde, indem sie ihre interviewenden Probanden in einer analogen Situation (statt Scheindiebstahl allerdings ein Schummeln bei einer Leistungsmessung) die Fragen sogar selbst generieren ließen. Auch hier zeigte sich eine erhöhte Zahl an schuldzuweisenden Fragen bei Personen, denen eine höhere Schuldwahrscheinlichkeit induziert wurde im Gegensatz zu Personen mit einer geringeren. Während die Autoren in der zweiten Studie keinen Zusammenhang zwischen einer schuldzuweisenden Befragung (im Gegensatz zu einer neutralen) und der Häufigkeit von Geständnissen finden konnten, zeigte sich in der dritten Untersuchung, dass Personen, die auf schuldzuweisende Fragen antworten, von Außenstehenden als schuldiger, nervöser und defensiver beurteilt wurden, was als selbsterfüllende Prophezeiung der anfänglichen Befragung unter schuldannehmenden Bedingungen gewertet werden muss. Der Effekt einer schuldzuweisenden Befragung als Konsequenz eines *confirmation bias* ist daher weniger im Erzeugen falscher Gedächtnisse, sondern in der späteren Wahrnehmung des Angeklagten durch die Urteiler im Gerichtsprozess zu sehen.

Während die zuvor beschriebenen Studien ihr Hauptaugenmerk auf das Interview legten, leistet die Studie von O'Brien (2009) einen Beitrag zur Generalisierung, da ihre Probanden auch bei der Bearbeitung von Erinnerungsaufgaben, der Interpretation

mehrdeutiger und inkonsistenter Informationen und der Wahl zu verfolgender Spuren vermehrt dann einen *confirmation bias* im Sinne einer abschließenden Feststellung der Schuld eines Verdächtigen aufwiesen, wenn sie im Unterschied zu anderen nach ca. der Hälfte der Urteilsprozesse angewiesen wurden, den derzeitigen Hauptverdächtigen zu benennen, was als Explizierung einer Hypothese verstanden werden kann.

Rassin, Eerland und Kuijpers (2010) klärten weitere Rahmenbedingungen des *confirmation bias* bei der polizeilichen Ermittlung auf. So zeigte sich im Falle einer Verbalisierung der Schuld-/ Unschuldannahme eines Verdächtigen nach dem Lesen einer ersten Fallbeschreibung, dass von den schuldanehmenden Probanden potentiell belastende Ermittlungsrichtungen präferiert wurden (57%), während die Unschuld vermutenden Teilnehmer diese Präferenz nicht zeigten (45%). In weiterführenden Studien, die auf eine Verbalisierung der Annahme verzichteten, sank der Anteil allerdings auf 37% (Studie 2). Interessanter Weise hing diese Präferenz jedoch auch von der Schwere des Delikts ab, d.h. im Falle schwererer Verletzungen des Opfers bei einer Körperverletzung wurden mehr belastende Ermittlungsrichtungen ausgewählt als bei geringeren Verletzungen. Aus rationalen Gründen sollte die Schwere des Delikts allerdings höchstens das Ausmaß der Kraftanstrengung im Zuge der Ermittlungsarbeiten erhöhen, nicht jedoch den Inhalt bzw. die Richtung. Ebenfalls problematisch erscheint der Befund, dass die Präferenz für die belastenden Ermittlungen mit der späteren Schuldzuschreibung assoziiert war und zwar unabhängig davon, ob die zusätzlich vorgelegten Ergebnisse der Ermittlungspfade das erwartete Ergebnis lieferten oder nicht (Studie 3).

Ask und Granhag (2005, Studie 2) konnten wiederum zeigen, dass eine weniger verzerrte Vorgehensweise resultiert, wenn eine plausible Alternative, d.h. ein weiterer möglicher Täter, von Beginn an verfügbar ist.

Abschließend zum Forschungskomplex des *confirmation bias* in rechtspsychologischen Arbeiten sei jedoch darauf hingewiesen, dass die zuvor beschriebenen Studien hauptsächlich mit Studierenden durchgeführt wurden, die sich (i.d.R.) in die Rolle eines ermittelnden Polizeibeamten hineinversetzen mussten. Es besteht daher Bedarf an Replikationsstudien mit berufserfahrenen Probanden bzw. Experten. Im Bereich der klinisch-psychologischen Diagnostik ist dieser Schritt bereits größtenteils unternommen worden. Da speziell dieser Anwendungskontext im Fokus der vorliegenden Arbeit liegt, soll eine erste Auswahl an einschlägigen Studien hier bereits ausführlicher vorgestellt werden.

Der confirmation bias in der klinisch-psychologischen Diagnostik. In einer Arbeit von Dailey (1952) zeigte sich, dass Personen, die schon nach dem Lesen der ersten Hälfte eines Lebenslaufs einer Zielperson angewiesen wurden, Einschätzungen/Urteile

abzugeben und damit zu einem frühen Zeitpunkt Schlussfolgerungen zu ziehen, am Ende schlechter vorhersagen konnten, was die Zielperson in einem Persönlichkeitsfragebogen ankreuzen würde, als Personen, die vor der ersten Schlussfolgerung die vollständige Biografie lesen durften. Für die Vorhersage stand den beiden Probandengruppen allerdings dieselbe Menge an Informationen zur Verfügung, sie unterschieden sich lediglich darin, ob sie zwischendurch bereits angehalten waren, Schlussfolgerungen zu ziehen. Es zeigte sich daher, dass ein Beobachter von seinen eigenen Urteilen über eine zu beobachtende Person in dem Sinne beeinflusst ist, dass eine frühzeitige Schlussfolgerung, seine Fähigkeit behindert, von nachfolgenden Daten zu profitieren und diese korrekt in seinen Eindruck zu integrieren. Was bereits für sich genommen beunruhigend ist, bekommt eine weitere Brisanz durch den Befund des *impact of the first three minutes* (p. 968) von Sandifer, Hordern und Green (1970). Die Autoren zeigten einer Gruppe von Psychiatern, die durchschnittlich zehn Jahre Erfahrung in ihrem Berufsfeld besaßen, Aufnahmen diagnostischer Interviews mit Patienten verschiedener Einrichtungen und baten sie, zu festgelegten Zeitpunkten alle Gedanken zum jeweiligen Patienten auf ein Tonband zu sprechen sowie auf einer Liste mit Symptomen diejenigen zu markieren, die sie beim Patienten beobachtet hatten. Anhand der Angaben, die die Probanden nach den ersten drei Minuten der gezeigten Videoaufnahme machten, konnten unabhängige Rater bereits ihr späteres Urteil vorhersagen. Zudem hatten die Teilnehmer zu diesem Zeitpunkt bereits die Hälfte aller insgesamt als beobachtet markierten Symptome notiert.

Während bei den beiden bislang erwähnten Studien der Effekt einer selbst vorgenommenen Schlussfolgerung bzw. eigener Gedanken im Mittelpunkt stand, interessierten Temerlin (1968) interpersonelle Einflüsse auf das Stellen einer psychiatrischen Diagnose. Konkret interessierte er sich dafür, ob die Vorgabe bzw. Suggestion einer bestimmten Störung durch einen Konföderierten mit hohem Ansehen die spätere Diagnosestellung beeinflusst. Ohne über den genauen Studienhintergrund informiert gewesen zu sein, wurde dafür ein professioneller Schauspieler angeleitet, einen psychisch gesunden Mann und dessen Lebensgeschichte darzustellen (die Kriterien für einen psychisch gesunden Mann stellt Temerlin ausführlich dar). Die so entstandene Tonbandaufnahme wurde nun verschiedenen Personengruppen in unterschiedlichen Kontexten zwecks UV-Manipulation vorgespielt. Der Experimentalgruppe, bestehend aus 45 Studierenden und jeweils 25 klinischen Psychologen und Psychiatern, wurde unmittelbar vor Darbietung dieses Tonbands mitgeteilt, der Patient auf dem Band sei „*a very interesting man because he looks neurotic, but actually is quite psychotic*“ (Temerlin, 1968, p. 350). Drei hinsichtlich der professionellen Identität gematchten Kontrollgruppen wurde das Tonband entweder ohne eine vorherige Suggestion oder mit umgekehrter Suggestion der Gesundheit

oder in einem abgeänderten Kontext eines Einstellungsinterviews präsentiert. Eine vierte Kontrollgruppe bestand aus Laienrichtern, denen mitgeteilt wurde, an einer Erprobung eines neuen *sanity hearings* teilzunehmen. Als Nachweis der Validität wertet der Autor das Ergebnis, dass niemand der 77 Personen aus den Kontrollgruppen ein Urteil im Hinblick auf einen psychotischen Zustand abgab und immerhin 80% zu dem korrekten Urteil der Gesundheit kamen. Anders hingegen in der Experimentalgruppe: nur 8% attestierten ihm psychische Gesundheit, während die anderen in Abhängigkeit ihrer professionellen Identität zur Psychose (z.B. 60% der Psychiater) oder Neurose (78% der Studierenden) tendierten. Der Autor schlussfolgerte daraus, dass die Suggestion im Sinne eines interpersonellen Einflusses Auswirkungen auf die Diagnosestellung nehmen kann und betrachtet die Abhängigkeit von der professionellen Identität in erster Linie als Folge der medizinischen Sozialisation der Psychiater, die einerseits in einer stark hierarchisierten Struktur arbeiten und andererseits durch einen enormen Druck zur Vermeidung falsch negativer Diagnosen geprägt sind. Statt einer Berufsschelte ist aus Sicht der Verfasserin daher vornehmlich zu schlussfolgern, dass Meinungen anerkannter Fachleute das eigene Urteil beeinflussen können und da der berufliche Alltag des Klinischen Psychologen dadurch geprägt sein wird, dass ihm Patienten von Arztbriefen oder Überweisungsschriften samt (Verdachts-) Diagnosen begleitet vorgestellt werden, ist ein Verzerrungspotential festzustellen.

Wenn sich interpersonelle Einflüsse auf den Prozess der Diagnosestellung feststellen lassen, liegt die Vermutung nahe, dass auch kontextuelle Aspekte eine verzerrende Wirkung mit sich bringen können. Genau diese Annahme untersuchte Rosenhan (1973). In diesem Zusammenhang verschafften sich acht Personen, die so genannten Pseudopatients, unentdeckt Zugang zu insgesamt zwölf psychiatrischen Kliniken. Alle gaben bei der Aufnahme vor, Stimmen zu hören. Direkt nach dem Vollzug der Aufnahme hörten die Pseudopatients mit dem Vorspielen der Symptome auf. Neben der Vorgabe der falschen Symptome sowie falschem Namen, Wohnort und Beruf wurden keine Veränderungen an der Lebensgeschichte und den Lebensumständen vorgenommen.

Rosenhan (1973) gibt an, dass die Pseudopatients nie entdeckt wurden. Vielmehr seien alle (mit einer Ausnahme) mit der Diagnose einer Schizophrenie in Remission anstatt der Diagnose „Gesundheit“ nach – im Mittel – 19 Tagen entlassen worden. Außerdem sei eigentlich auffälliges Verhalten – konkret das offensichtliche Anfertigen ausführlicher Notizen – nicht hinterfragt worden, da man bei Schizophrenen ungewöhnliches Verhalten als durchaus typisch erachtet habe. Auch habe es niemanden verwundert, wie die unauffällige Lebensgeschichte der Pseudopatients zu der Schizophrenie passen könne. Rosenhan (1973, p. 253) schlussfolgert daher, dass *„having once been labeled schizophrenic, there is nothing the pseudopatient can do to overcome the tag.“* Außerdem sieht er seine Annahme

bestätigt, dass innerhalb einer psychiatrischen Klinik nicht zwischen Gesunden und Kranken unterschieden werden könne (vgl. auch Rosenhan, 1975). Reliabilität und damit auch Validität der Diagnosestellung sei daher in Frage zu stellen. Die Veröffentlichung dieser Arbeit traf die Psychiatrie nachvollziehbarer Weise mitten ins Herz und ließ viele Leser geschockt oder geradezu empört zurück. Davis (1976) fand es daher nicht überraschend, dass die sich an die Publikation anschließende Diskussion mehr Hitze als Licht mit sich brachte.

Es folgten einerseits Replikationsansätze mit dem Ziel zu zeigen, dass Pseudopatienten durchaus erkannt werden können. So z.B. zu lesen bei Shipko (1982), der von der Vorstellung eines weiteren Pseudopatienten berichtet, der von den Klinikmitarbeitern durchaus als schwingungsfähig eingeschätzt wurde und dessen Berichte über aggressive Fantasien als normal eingestuft wurden. Es wurde daher kein psychotischer Prozess festgestellt, obwohl auch dieser Pseudopatient von auditorischen Halluzinationen berichtete. Es schienen also nicht alle Psychiater Schwierigkeiten mit dem Unterschied zwischen Krankheit und Gesundheit zu haben.

Aber auch methodische Kritik wurde geübt (vgl. z.B. Weiner, 1975). So argumentiert u. a. Spitzer (1975), dass die Psychiater durchaus entdeckt haben, dass die anfangs krank wirkenden Pseudopatienten nicht länger Anzeichen der Krankheit zeigten, indem sie sie mit der Diagnose Schizophrenie in Remission entlassen haben. Nach seiner Ansicht wurde mit der Abwesenheit der Krankheitszeichen die Gesundheit hinreichend entdeckt. Eine symptomvortäuschende gesunde Person bei der Aufnahme zu erkennen oder festzustellen, dass eine im Krankenhaus normal agierende Person die Krankheit anfangs vorgetäuscht hat, ist seiner Ansicht nach kein Vorwurf, den man dem Personal machen könne und tangiere auch nicht Rosenhans Forschungsfrage. In vergleichbarer Weise schlägt Millon (1975) vor, Rosenhans Kontexthypothese zu überprüfen, indem einerseits gesunde und andererseits andernorts als krank identifizierte Personen ihre wahren Geschichten berichten. Würden Gesunde aufgenommen und Kranke abgewiesen, wäre Rosenhans These bestätigt. Als weitere Möglichkeit der Überprüfung schlägt er vor, Psychiater in eine unbekannte Klinik zu schicken, die zur Hälfte mit gesunden und kranken (Pseudo-) Patienten belegt ist.

Andererseits löste das Experiment eine durchaus berechtigte Systemkritik an der Behandlung psychisch Kranker zu dieser Zeit aus (vgl. z.B. Rabichow & Pharis, 1974), auf die u. a. Crown (1975) mit der Forderung nach Wandlung der psychiatrischen Krankenhäuser hin zu einer therapeutischen Gemeinschaft mit einer kontinuierlichen Beobachtung der Patienten reagierte. Außerdem wurden Veränderungen an den Klassifikationssystemen angestoßen. Positive Effekte des Experiments sind also auf jeden Fall festzuhalten.

Neben der Bedeutung für die Betrachtung des *confirmation bias* ist an dem Umgang mit dem Rosenhan-Experiment in der Literatur festzuhalten, dass die Kontroverse über weite Strecken sehr emotional geführt wurde, und dass sie wie eine nicht heilende Wunde immer wieder aufbricht. Konkret zeigte sich 2005 ein Rezidiv durch hitzige Diskussionen im *Journal of Nervous and Mental Disease* im Zusammenhang mit der von Spitzer, Lilienfeld und Miller (2005) kritisierten Darstellung von Slater (2004), die mit einer *response to a nonresponse to criticisms of a nonstudy* endet (Lilienfeld, Spitzer & Miller, 2005). Die Autoren verleihen in dieser Antwort der Hoffnung Ausdruck, dass man ihnen verzeihen möge, die Arbeit von Slater (2004) als Experiment (im Sinne einer Replikation von Rosenhan schildert sie einen Selbstversuch bei verschiedenen Psychiatern) betrachtet und als solche einer wissenschaftlichen Kritik unterzogen zu haben. Denn die Autorin habe sich aller Vorwürfe mit dem Argument entledigt, sie habe gar keine Studie durchgeführt, was als *nonresponse* aufgefasst wird. Ihre *nonstudy* sei vielmehr ein Ergebnis ihres kreativen Schreibens als Journalistin.

Abseits von *pseudopatients* und *nonstudys* hat jedoch die Arbeitsgruppe um Strohmmer (vgl. Strohmmer & Chiodo, 1984; Strohmmer & Newman, 1983; Strohmmer, Shivy, Chiodo, 1990) in 1980er und 1990er Jahren einen großen Beitrag zum Verständnis des *confirmation bias* in der klinisch-psychologischen Diagnostik geleistet. In der Regel wurde bei diesen Studien aufbauend auf dem Paradigma von Snyder und Swann (1978) das diagnostische Vorgehen von psychologischen Beratern bzw. Therapeuten beim Testen von Hypothesen über die Persönlichkeit oder pathologische Auffälligkeiten untersucht.

Diese Studien spielen für die vorliegende Arbeit eine so zentrale Rolle, dass sie, auch um Redundanzen zu vermeiden, im Zusammenhang mit dem Modell des konfirmatorischen Hypothesentestens (s. Abschnitt 2.3) ausführlich beschrieben werden. An dieser Stelle sei im Sinne einer ersten Bilanz festgehalten, dass die Unterschiedlichkeit der oben dargestellten Arbeiten, die für sich alle beanspruchen, den *confirmation bias* zu untersuchen, immens ist und von Hager und Weißmann (1991) durch die Subsumierung des *primacy effects* (Asch, 1946), des *Versuchsleitereffekts* (Rosenthal, 1976), der *selbsterfüllenden Prophezeiung* (Merton, 1948) und vielen mehr weiter auf die Spitze getrieben wird. Zu Recht werfen Hager und Weißmann (1991) in diesem Zusammenhang die Frage nach der Sinnhaftigkeit einer solch breiten Definition des *confirmation bias* auf, der so viele unterschiedliche Phänomene aus den verschiedenen Teildisziplinen der Psychologie zu vereinigen scheint. An dieser Stelle ist daher eine Auseinandersetzung mit dem theoretischen Konzept des *confirmation bias* angezeigt.

Eine theoretische Annäherung. Die oben zitierten empirischen Arbeiten enthalten größtenteils auch Annahmen über die Ursachen und Bedingungen des *confirmation bias*. Dabei deutet die inhaltliche und methodische Unterschiedlichkeit der Studien bereits an, dass die Konzeption der Ursachen ähnlich heterogen ausfallen dürfte wie die Operationalisierungen. Auf theoretischer Ebene haben in chronologischer Reihenfolge Snyder und Swann (1978), Klayman und Ha (1987), Hager und Weißmann (1991), Trope und Liberman (1996) sowie Oswald und Grosjean (2004) wertvolle Beiträge zum Verständnis des *confirmation bias* geleistet (für eine ältere Überblicksarbeit s. auch Nickerson, 1998). Die unterschiedlichen Konzeptionen sollen in der Folge vorgestellt werden, bevor abschließend Stellung bezogen wird, welches Verständnis des *confirmation bias* dieser Arbeit zugrunde liegt.

Die frühen Arbeiten zum *confirmation bias* von Wason (1960), Wason und Johnson-Laird (1972) sowie die Arbeiten in Anlehnung an das Paradigma von Snyder und Swann (1978) fokussieren auf den Prozess der Informationssuche, indem die Probanden entweder geeignete Prüfinstanzen für zu testende Regeln suchen sollen oder Fragen auswählen bzw. generieren, um Hinweise auf ein Persönlichkeitsmerkmal zu finden. Konkret wird Letzteres von Snyder und Swann (1978) als *confirmatory search strategy* beschrieben und als Beleg für den *confirmation bias* gewertet. Sowohl Klayman und Ha (1987) als auch Trope und Liberman (1996) argumentieren jedoch, dass die beschriebenen Phänomene eher im Sinne einer *positiven Teststrategie* zu verstehen sind, die nicht notwendigerweise einem *confirmation bias* entspricht, weil aufgrund logischer Überlegungen gezeigt werden kann, dass unter bestimmten Bedingungen auch bei der Suche nach bestätigender Evidenz eine Falsifikation der Hypothese möglich ist. Da für eine umfassende Darstellung der Diskussion weitere Konzepte vorzustellen sind, sei auf Abschnitt 2.3.3.2 verwiesen, in welchem eine klare Abgrenzung vorgenommen werden kann.

Für die jüngeren Arbeiten kann festgehalten werden, dass die Gleichsetzung von positiver Teststrategie und *confirmation bias* weitgehend aufgegeben wurde und heute eine umfassendere Betrachtung weiterer Schritte im Testungsprozess relevant ist. So definieren Oswald und Grosjean (2004, p. 79) wie folgt: „*Confirmation bias means, that information is searched for, interpreted, and remembered in such a way that systematically impedes the possibility that the hypothesis could be rejected – that is, it fosters the immunity of the hypothesis.*“ Es wird dementsprechend davon ausgegangen, dass es Mechanismen gibt, die *systematisch* dazu führen, dass die Hypothese stärker bestätigt wird als die Evidenz es objektiv erlaubt. In diesem Sinne ist der *confirmation bias* als Selbstbestätigungstendenz zu verstehen und konfirmatorische Mechanismen

als Prozesse, die systematisch die Bestätigung der zu testenden Hypothesen begünstigen und im selben Moment die Wahrscheinlichkeit der Zurückweisung verringern. Welche Mechanismen dies sein können, wird in der Folge darzulegen sein (s. Abschnitt 2.3).

In Anlehnung an Hager und Weißmann (1991) wird der *confirmation bias* in dieser Arbeit als *Bestätigungstendenz* verstanden, nicht als *Bestätigungsfehler*, denn Letzteres impliziert, dass zwangsläufig Fehler resultieren, was nicht der Fall ist, da die Fehlerhaftigkeit vom wahren Zustand abhängt. Mit dem Begriff der *Tendenz* soll vielmehr ausgedrückt werden, dass es sich um ein Verhalten handelt, das nicht notwendigerweise eine verzerrende Wirkung hat, sondern nur eine bestimmte Wahrscheinlichkeit dafür mit sich bringt.

Desweiteren ist das Verständnis des *confirmation bias* in dieser Arbeit einerseits angelehnt an die Definition von Oswald und Grosjean (2004) und beinhaltet dementsprechend Mechanismen, die systematisch dazu führen, dass die Hypothese stärker bestätigt wird als die Evidenz es objektiv erlaubt. Auch die Definition der *Bestätigungstendenz* nach Hager und Weißmann (1991) findet sich im Verständnis der vorliegenden Arbeit wieder. Die Autoren definieren wie folgt:

„Wenn eine Person durch die Auswahl und/oder die Verarbeitung der Informationen zur Prüfung einer Hypothese den Ausgang der Hypothesenprüfung gezielt oder absichtsvoll derart beeinflusst, daß die Wahrscheinlichkeit einer Hypothesenbestätigung ansteigt, dann soll dieses Vorgehen als „Bestätigungstendenz“ beim Hypothesenprüfen bezeichnet werden. In Abhängigkeit davon, an welcher Stelle auf den Prüfprozeß Einfluß genommen wird, kann zwischen einer Bestätigungstendenz durch die Auswahl (A), Bewertung (B) oder (Um-)Interpretation (I) von Prüfinstanzen unterschieden werden.“ (S. 20, Rechtschreibung wie im Original).

Die Autoren differenzieren daher in drei *Bestätigungstendenzen*. Als ein Verhalten im Sinne der *Bestätigungstendenz A* wird die bevorzugte Auswahl von hypothesenbestätigenden Prüfinstanzen verstanden. Die *Bestätigungstendenz B* beinhaltet die Aufwertung hypothesenkonsistenter Daten, die sich bspw. darin manifestieren kann, dass die Wichtigkeit einer konformen Information bei größerer Überzeugtheit von der Hypothese größer eingeschätzt wird, als bei einer geringeren Überzeugtheit. Eine vergleichbare Abwertung hypotheseninkonsistenter Daten wird als *Widerlegungsvermeidungstendenz* bezeichnet. Im Falle einer Uminterpretation von hypotheseninkonsistenten Daten in hypothesenkonsistente sprechen Hager und Weißmann (1991) von der *Bestätigungstendenz I*.

Der Vollständigkeit halber sei angemerkt, dass die Autoren auch das jeweils umgekehrte Verhalten umschreiben und korrespondierend als Widerlegungstendenzen benennen.

2.1.3 *Der Mensch als Hypothesentester*

Nicht nur die konkreten Ausführungen zu der empirischen Befundlage bzgl. des *confirmation bias* an dieser Stelle, sondern auch der allgemeinere Überblick über Urteilsverzerrungen unter 2.1.1 deuten auf den ersten Blick an, dass der Mensch womöglich prinzipiell nicht in der Lage ist, valide Schlussfolgerungen zu ziehen. Diese Mutmaßung soll im folgenden Abschnitt thematisiert werden, indem unterschiedliche Sichtweisen zur prinzipiellen Befähigung des Menschen zum Urteilen bzw. Hypothesentesten diskutiert werden. In einem ersten Schritt wird die Frage adressiert, ob Menschen nicht adäquat testen *können* oder ob sie es nicht *wollen*. Desweiteren wird die Diskussion skizziert, ob Fehler bzw. Fehlurteile Resultate eines schweren kognitiven Defizits sind oder eher die Folge einer eigentlich umweltangepassten Strategie, die nicht zwangsläufig fehlerbehaftet sein muss. Abschließend ist zu klären, welches Bild des Menschen als Hypothesentester dieser Arbeit zugrunde liegt.

Kognitives oder motivationales Defizit? Durch die spontane Generierung oder durch die Vorgabe einer zu prüfenden Annahme wird eine Person zum Hypothesentester. So zum Beispiel der Mieter in einem Mehrfamilienhaus, der sich nach einer Begegnung mit dem neuen Nachbarn im Hausflur spontan fragt, ob dieser wohl eine Affinität zur Partei *Die Grünen* hat. Oder nehmen wir den psychologischen Gutachter, der beauftragt wird zu überprüfen, ob bei der pubertierenden Patientin eine Bulimie vorliegt. Müssen die Ursachen für Fehlentscheidungen bei solchen Fragestellungen nun im Versagen des kognitiven Systems des Menschen gesucht werden oder in seiner Motivation oder spielt gar beides eine Rolle?

Einen entscheidenden Hinweis liefern Wason und Johnson-Laird (1972), die ihre Probanden die Hypothese *all chairs have four legs* testen ließen und hier eine Präferenz für die Suche nach bestätigenden statt widerlegenden Informationen fanden. Denn motivationale Gründe für die Bestätigung der Hypothese, alle Stühle hätten vier Beine, sind kaum vorstellbar, so dass die emotionale Unterstützung einer Hypothese keine notwendige Bedingung für ein invalides Vorgehen darstellt. Hier müssen demnach kognitive Ursachen für die Fehlurteile verantwortlich gemacht werden.

Bleibt die Frage zu klären, ob die emotional positive Bewertung einer Hypothese eine hinreichende Bedingung für Verzerrungen beim Testen darstellt. Oswald und Grosjean (2004) führen an, dass auch bei emotional unterstützten Hypothesen die potentiellen Fehler und deren Folgekosten in die Testüberlegungen einfließen, so dass nicht generell geschlossen werden könne, eine erwünschte Hypothese werde immer invalide getestet.

Zusammenfassend ist zu sagen, dass die Motivation weder eine notwendige noch eine hinreichende Bedingung für das Auftreten von Verzerrungsprozessen darstellt. Damit ist selbstverständlich nicht ausgeschlossen, dass auch motivationale Prozesse zu Fehlentscheidungen beitragen können. Welche Rolle die Motivation für die vorliegende Arbeit spielt, wird unter Punkt 2.3.4 eingehend erläutert, es sei jedoch vorweggenommen, dass nach dieser hinführenden Diskussion eine Fokussierung auf die kognitiven Verzerrungen erfolgt, denn im Rahmen der untersuchten klinisch-psychologischen Diagnostik ist ein professioneller Kontext zu unterstellen, der die Rolle der Motivation in den Hintergrund rücken lässt.

Schwerwiegendes Defizit oder nur punktuell fehlgeleitete, umweltangepasste Strategie? Wenn nun das kognitive System des Menschen für einen großen Teil der Fehleranfälligkeit verantwortlich zu machen ist, muss die Frage diskutiert werden, ob der Mensch ein solches gravierendes kognitives Defizit aufweist, das ihn als Hypothesentester vollkommen ungeeignet erscheinen lässt oder ob die eingesetzten Strategien trotz einer potentiellen Fehleranfälligkeit nicht doch rational und angepasst sind. Um mit den Worten von Oswald (in press) zu sprechen, ist hiermit die Gegenüberstellung der Sichtweisen des Menschen als *cognitive miser* oder *rationalen Informationsverarbeiters* gemeint. Während Halder-Sinn (1993) die Sichtweise des *cognitive miser* vertritt, indem sie annimmt, ein grundlegendes Defizit in der Verarbeitung hochkomplexer, vielfältiger Informationen sei die Ursache für psychologische Urteilsfehler, die auch bei Experten aufträten, verteidigt Oswald (in press) die kognitiven Fähigkeiten des Menschen. Letztere plädiert dafür, den Menschen nicht als fehleranfälligen Hypothesentester zu betrachten, der lediglich vereinfachende Heuristiken anwendet, weil er es nicht besser kann. Ihre Ausführungen zeigen, dass der Einsatz von Heuristiken nicht als Unvermögen gewertet werden muss, sondern durchaus rationale Züge aufweist. Als wesentliches Argument führt sie das Bayes'sche Theorem an, mit dem gezeigt werden kann, dass es durchaus rational ist, eine Hypothese trotz Vorhandensein von Widersprüchen aufrechtzuerhalten. Denn nach dem Bayes'schen Theorem verringert ein Widerspruch lediglich die Wahrscheinlichkeit des Zutreffens der zu testenden Hypothese, muss jedoch nicht zu

deren Verwerfen führen. Als weiteres Argument führt sie an, dass es ebenso rational ist, nicht jede zur Verfügung stehende Information zu beachten, da dies eine hoffnungslose kognitive Überforderung darstellen würde.

Danach ist der Mensch kognitiv durchaus in der Lage, die Rolle des Hypothesentesters zu erfüllen, so dass eine theoretische und empirische Beschäftigung mit dem Prozess des Hypothesentestens und den möglichen Fehlern angezeigt und fruchtbar erscheint. Die Sichtweise des Menschen, die in dieser Arbeit vertreten wird, lehnt sich dabei eng an diejenige von Oswald (in press) an und kann wie folgt zusammengefasst werden: Erstens zeichnet es den Menschen aus, Annahmen überprüfen zu können, zweitens macht der Mensch weder alles richtig noch macht er alles falsch und drittens ist der Mensch weder Opfer irrationaler Gefühle noch eines defizitären Denkapparates.

Anliegen dieser Arbeit ist es deshalb, zu zeigen, wie es dazu kommen kann, dass intellektuell unbeeinträchtigte, emotional angepasste und unvoreingenommene Durchschnittserwachsene zu falschen Schlussfolgerungen nach dem Hypothesentesten gelangen können. Dieser Anspruch entspricht demjenigen von Schulz-Hardt und Köhnken (2000), auf deren Ausführungen auch das unter Punkt 2.3 dargestellte Modell des konfirmatorischen Hypothesentestens beruht.

2.2 Ablauf einer Hypothesenprüfung

Nachdem dargelegt wurde, dass der Mensch durchaus in der Lage ist, Hypothesen zu testen, ist es erforderlich, sich die Konzeptualisierung des Testungsprozesses näher anzuschauen. Zunächst werden die beiden Modelle des (sozialen) Hypothesentestens von Trope und Liberman (1996) sowie Hager und Weißmann (1991) vorgestellt, bevor ein einschlägiges Modell des (psycho-)diagnostischen Prozesses nach Jäger (1982) vergleichend erläutert wird.

2.2.1 Modellvorstellungen zum Ablauf des (sozialen) Hypothesentestens

Trope und Liberman (1996) sowie Hager und Weißmann (1991) stimmen überein, dass der Ablauf der Überprüfung einer sozialen Hypothese, d.h. einer Annahme über die soziale Umwelt, sich nicht wesentlich von dem Vorgehen bei der Prüfung wissenschaftlicher Annahmen unterscheidet. Beide Autorenpaare formulieren ihre Abläufe daher auf

einem hohen Abstraktionsniveau, so dass ein großer Geltungsbereich abgedeckt wird, und schlagen einen möglichen Ablauf einer Hypothesenprüfung in einem mehrstufig konzipierten Prozess vor.

2.2.1.1 *Modell des sozialen Hypothesentestens nach Trope und Liberman (1996)*

Trope und Liberman (1996) beschreiben die folgenden Komponenten und deren entsprechende Verknüpfungen: Der Prozess des Hypothesentestens besteht prinzipiell aus dem Durchlaufen von fünf unterschiedlichen Schritten. Hierzu gehören die Generierung einer Hypothese, die Ableitung von Evidenzregeln, die Informationssuche, die Identifizierung der Evidenz und die Schlussfolgerung.

Generierung der Hypothese. Als Erstes erfolgt die Generierung einer Hypothese (*hypothesis generation* im Original), die sich auf jede erdenkliche Annahme über Merkmale, Ursachen oder Zusammenhänge der sozialen Umwelt beziehen kann. Es ist einerseits denkbar, dass die Hypothese spontan aus dem sozialen Zusammenhang produziert wird. Beispielsweise könnte sich der Mieter im Mehrfamilienhaus nach der Beobachtung, dass der neue Nachbar immer mit dem Fahrrad fährt, fragen *Wählt mein Nachbar die Grünen?* Andererseits kann die Annahme auch von außen an die Person herangetragen werden, wie zum Beispiel an den Psychologen, der in dem Überweisungsschreiben eines Hausarztes die Verdachtsdiagnose *Bulimie* mitgeteilt bekommt und gebeten wird, diese Hypothese zu überprüfen. Darüber hinaus wären an dieser Stelle auch Alternativhypothesen zur Prüfhypothese zu generieren, die je nach Themengebiet entweder einen diffusen Zustand des nicht Zutreffens der Prüfhypothese (*Der neue Nachbar wählt irgendeine andere Partei oder gar keine*) oder eine konkrete Alternative (*Die Patientin ist anorektisch*) beinhalten kann. Der Vollständigkeit halber sei erwähnt, dass selbstverständlich auch Beispiele mit mehreren konkreten Alternativhypothesen denkbar sind.

Ableitung von Evidenzregeln. Nachdem die zu testende Hypothese feststeht, werden Regeln für die später zu suchende Evidenz aufgestellt. Diesen Prozess beschreiben die Autoren als *derivation*. An dieser Stelle wird folglich festgelegt, welche Informationen als Belege, als Hinweise oder als Widersprüche zu interpretieren sind, so dass hiermit der Wert der potentiellen Informationen bestimmt wird. Die Evidenzregeln können dabei aus dem gesunden Menschenverstand oder aus eigenem Fachwissen und aus fundierter

Literaturrecherche abgeleitet werden. Der Kontext der Hypothese wird dabei den Bezugsrahmen bedingen, d.h. der Mieter im Mehrfamilienhaus wird seinen gesunden Menschenverstand bzw. sein persönliches Bild von einem Wähler der Grünen aktivieren, während der Psychologe die Evidenzregeln unter anderem aus seinen bisherigen Erfahrungen, seinem Fachwissen und der einschlägigen Literatur ableiten sollte.

Die Evidenzregeln werden meist in Form von *Wenn-Dann-Regeln* aufgestellt und beinhalten die Regel „*Wenn* eine Annahme wahr ist, *dann* liegt wahrscheinlich eine bestimmte Evidenz vor.“ Der Mieter im Mehrfamilienhaus würde dementsprechend formulieren *Wenn mein Nachbar die Grünen wählt (Annahme ist wahr), dann sollte er den Müll sorgfältig trennen* (diese Tatsache sollte beobachtbar sein). Der Psychologe könnte beispielsweise folgende Regel aufstellen: *Wenn die Patientin an Bulimie leidet, dann müssen bei ihr im Zeitraum von mindestens drei Monaten mindestens zwei Heißhungeranfälle pro Woche aufgetreten sein bzw. auftreten.*

Informationssuche. Sobald die Evidenzregeln aufgestellt sind, kann die Informationssuche - *information gathering* - erfolgen, da vorher festgelegt wurde, welche Informationen relevant sind. Der Mieter könnte hier nun versuchen, sich zu erinnern, ob der Müll in den Gemeinschaftstonnen beim Einwerfen seiner letzten Tüte ordnungsgemäß getrennt war. Er könnte auch andere Nachbarn befragen, ob sie den neuen Nachbarn mit verschiedenen Mülltüten beobachtet haben oder gar selbst einen Tag lang die Mülltonnen beobachten. Dabei würde der Mieter die typischen Quellen für die Informationssuche anzapfen: das eigene Gedächtnis, Befragungen von anderen Beteiligten und eigene Beobachtungen. Dasselbe wird der Psychologen, der eine Essstörungsdiagnostik vorzunehmen hat, in seinen diagnostischen Prozess integrieren. Die Autoren merken zudem an, dass an dieser Stelle kein rein suchender Prozess stattfindet, sondern an dieser Stelle bereits eine Reihe von Entscheidungen getroffen werden müssen, z.B. wie viel Aufmerksamkeit welchen Informationen geschenkt wird oder wann hinreichend viele Informationen vorliegen.

Identifizierung der Evidenz. Dieser Schritt, den Trope und Libermann (1996) als *identification* bezeichnen, beinhaltet die Interpretation bzw. Kategorisierung der zuvor gefundenen Informationen. Mit Hilfe der im zweiten Schritt aufgestellten Regeln werden die Informationen als Hinweise oder Widersprüche eingestuft. Findet der Mieter heraus, dass der neue Nachbar mit vier unterschiedlichen Tüten zu den Mülltonnen gegangen ist, wird er dies als Beleg für die Wahlentscheidung einstufen. Der Wert jeder einzelnen Information der tatsächlich gefundenen Evidenz wird folglich in diesem Schritt festgelegt.

Schlussfolgerung. Am Ende des Testungsprozesses steht die Schlussfolgerung – *inference* – d.h. es wird eine integrative Bewertung der gefundenen und interpretierten Informationen vorgenommen. Es erfolgt eine Abschätzung der Wahrscheinlichkeit, dass die Hypothese vor dem Hintergrund der vorhandenen Evidenz wahr ist. Es erfolgt entweder die Annahme oder die Zurückweisung der Hypothese, wenn der Tester davon überzeugt ist, genug Informationen gefunden zu haben, oder er entscheidet sich dafür, weitere Informationen zu suchen. An dieser Stelle wird deutlich, dass der Ablauf des Testens nicht zwangsläufig linear erfolgen muss.

2.2.1.2 *Ablauf einer Hypothesenprüfung nach Hager und Weißmann (1991)*

Hager und Weißmann (1991) nehmen ebenso wenig wie Trope und Liberman (1996) einen Unterschied zwischen der Testung einer wissenschaftlichen und einer alltagspsychologischen Hypothese an und schlagen daher einen aus fünf Schritten bestehenden Prozess für jede Art von Hypothese vor, wobei sie zwischen Hypothese, Identifikation der potentiell hypothesenbestätigenden/-widerlegenden Prüfinstanzen, Bewertung der Informationen, Hypothesenprüfung und Hypothesenentscheidung sowie Bewertung der Hypothese differenzieren.

Hypothese. Die Hypothese bildet den Ausgangspunkt der Prüfung, durch die entschieden werden soll, ob der in der Hypothese angesprochene Sachverhalt in der Realität tatsächlich gegeben ist, d.h. ob die Hypothese gültig ist oder nicht. Die Autoren verstehen unter einer Hypothese „*jegliche Annahme, Vermutung oder Behauptung über Beziehungen oder Zusammenhänge zwischen Konstrukten, Variablen oder Sachverhalten*“ (1991, S.8). Sowohl die Annahme des Mieters im Mehrfamilienhaus, dass der neue Nachbar die Grünen wähle, als auch die Verdachtsdiagnose Bulimie des Psychologen sind daher auch nach Hager und Weißmann (1991) mögliche Hypothesen, mit denen der Prüfprozess beginnt. Insofern ist der hier genannte erste Schritt mit demjenigen von Trope und Liberman (1996) nach Ansicht der Verfasserin deckungsgleich.

Identifikation der potentiell hypothesenbestätigenden/-widerlegenden Prüfinstanzen. In diesem Schritt geht es um die Vorhersagen, welche Informationen (Prüfinstanzen) mit der Hypothese vereinbar, also hypothesenkonsistent, bzw. nicht vereinbar, d.h. hypotheseninkonsistent, sind. Im Anschluss daran werden die Prüfinstanzen „*erzeugt, hergestellt, aufgesucht oder sonstwie verfügbar gemacht*“, um

mit den Worten der Autoren zu sprechen (1991, S.10). Der Mieter im Mehrfamilienhaus würde dementsprechend als potentiell hypothesenbestätigende Prüfinstanz festlegen, dass eine korrekte Mülltrennung durch den neuen Nachbarn erfolgt und diese Information bspw. mit Hilfe von Gesprächen mit weiteren Mietern suchen. Der Psychologe würde an dieser Stelle zum Beispiel das Vorliegen einer Körperschemastörung, im Sinne einer verzerrten Wahrnehmung der eigenen Körperproportionen, als potentiell hypothesenwiderlegende Prüfinstanz identifizieren, die ihm der Test mit dem *subjective body dimensions apparatus* (Gila, Castro, Toro & Salamero, 1998) anzeigen könnte.

Nach Ansicht der Verfasserin fassen Hager und Weißmann (1991) in diesem Aspekt die von Trope und Liberman (1996) in Ableitung der Evidenzregel und Informationssuche getrennten Schritte zusammen. Festzuhalten ist jedoch die inhaltliche Übereinstimmung. Auch die in beiden Modellen nachfolgenden Schritte sind als weitgehend äquivalent zu werten, sowohl bei der *Identifizierung der Evidenz* als auch bei der Bewertung der Informationen hinsichtlich ihrer Eignung zur Hypothesenprüfung geht es um die Zuschreibung des Wertes der gefundenen Informationen.

Bewertung der Informationen. Im Rahmen der Bewertung wird festgelegt, in welchem Umfang eine Information etwas zur Prüfung der Hypothese beitragen kann oder soll. Dies berücksichtigt den Umstand, dass die erhobenen Daten nie absolut objektiv, zuverlässig und gültig sein können, sondern eine konkrete Einschätzung der Eignung vorgenommen werden muss. Wenn nun die 85jährige Frau Krause, die zum Lesen eine Brille benötigt, dem Mieter berichtet, dass der neue Nachbar den Müll am letzten Mittwoch getrennt habe, dann sind Objektivität und Zuverlässigkeit womöglich anders einzustufen als beim Psychologen, der mit Hilfe des standardisierten und validen Testverfahrens das Vorliegen einer Körperschemastörung ausschließen konnte. Während der Mieter also Zweifel an der Eignung der erhobenen Information haben sollte, würde der Psychologe wahrscheinlich die Eignung seiner Information bestätigen.

Hypothesenprüfung und Hypothesenentscheidung. An dieser Stelle soll mit Hilfe des Vergleichs des erhobenen Befundmusters mit dem erwarteten die Frage beantwortet werden, ob die Hypothese mit den neuen Informationen vereinbar ist oder nicht. Dies soll eine Aussage über den Status der Hypothese ermöglichen, die da lauten könnte „Die Hypothese konnte sich in dem betrachteten Versuch (nicht) bewähren.“ (Hager & Weißmann, 1991, S.11). Dieser Schritt korrespondiert mit der Schlussfolgerung, wie Trope und Liberman (1996) sie verstehen.

Bewertung der Hypothese. Der letzte Schritt trägt der Tatsache Rechnung, dass Hypothesen in der Regel in einen größeren Gesamtzusammenhang eingebettet sind und aufgrund anderer Erfahrungen oder Erkenntnisse schon eine gewisse subjektive Sicherheit darüber herrscht, dass die fragliche Hypothese zutreffend ist. Im Zuge der Bewertung der Hypothese kann es nun zu einer Steigerung oder Absenkung der subjektiven Sicherheit kommen, wenn nämlich die Konsequenzen der Hypothesenprüfung betrachtet werden, die sich nicht nur aus der vorangegangenen Entscheidung auf Stufe vier ergeben.

Nach Auffassung der Verfasserin unterscheiden sich die Abläufe v.a. in diesem letzten Schritt nach Hager und Weißmann (1991), der bei Trope und Liberman (1996) nicht explizit angesprochen wird. Da jedoch aus ihren Ausführungen nicht ersichtlich wird, dass ein derartiger Schritt abgelehnt wird, wertet die Verfasserin dieser Arbeit die fehlende Kongruenz als unwesentlich. Insofern stützt sich die vorliegende Arbeit im Hinblick auf das Verständnis des Ablaufs der Hypothesenprüfung sowohl auf die Konzeptualisierung von Hager und Weißmann (1991) als auch diejenige von Trope und Liberman (1996). Inwieweit eine Deckungsgleichheit auch mit dem Ablauf des (psycho-)diagnostischen Prozesses herrscht, wird im folgenden Abschnitt thematisiert.

2.2.2 *Ablauf des diagnostischen Prozesses*

Unter dem diagnostischen Prozess ist die Abfolge von Maßnahmen zu verstehen, die der Gewinnung diagnostisch relevanter Informationen für professionelle Entscheidungen dienen (vgl. Amelang & Zielinski, 2002).

Auftraggeber und Fragestellung. Zunächst tritt dabei ein Auftraggeber an den Diagnostiker mit einem in der Regel laienpsychologisch formulierten Anliegen heran und wünscht sich von diesem eine Beantwortung der Fragestellung, um in der Folge eine Entscheidung fundierter treffen zu können. Wenngleich die verschiedensten Anlässe für eine Diagnostik denkbar sind, die beispielhaft die Schulfähigkeit, die Geschlechtsidentität, die Fahreignung oder Berufseignung als Altenpflegerin einer Person betreffen können, so ist das Vorhandensein einer Fragestellung als Ausgangspunkt immer obligatorisch. Die Hypothesen, die im folgenden Schritt aufzustellen sind, sind daher notwendigerweise in dieses Anliegen eingebettet. Eine spontane Generierung, wie Trope und Liberman (1996) sie für den Ursprung der Hypothese zulassen, ist hier nicht vorgesehen. Man kann daher festhalten, dass der diagnostische Prozess an einem früheren Punkt mit stärkerer Bindung an äußere Umstände beginnt als das soziale bzw. wissenschaftliche Hypothesentesten.

Hypothesenformulierung. Der Diagnostiker formuliert anschließend Hypothesen, z.B. in Form von Erklärungen für ein Problemverhalten oder als Anforderungen, um für eine Maßnahme geeignet zu sein. Er wird hierfür sein Fachwissen über die Zusammenhänge zwischen den relevanten Variablen benutzen. Implizit sind hierfür auch die von Trope und Liberman (1996) angesprochenen Evidenzregeln bzw. Prüfinstanzen nach Hager und Weißmann (1991) relevant, da diese bspw. lauten würden, *Wenn bei einem Kind eine Intelligenzminderung vorliegt, dann sollte sich in den ersten Jahren der Grundschule ein Leistungsabfall zeigen* oder *Wenn die Person als Altenpflegerin geeignet ist, muss sie körperlich belastbar sein*. Als Hypothese würde dann entweder die Feststellung *Das Kind verfügt über eine unterdurchschnittliche Intelligenz* oder *Person A ist hinreichend körperlich belastbar* formuliert. Der (fachlich korrekt agierende) Diagnostiker wird daher viele Hypothesen aufstellen, um alle Alternativerklärungen für das Problemverhalten bzw. alle Anforderungen für die Eignung zu berücksichtigen.

Untersuchungsplanung und Durchführung der Untersuchung. An dieser Stelle geht es genau wie bei Trope und Liberman (1996) sowie Hager und Weißmann (1991) um die Informationssuche. Der Diagnostiker wird festlegen, welche Informationen er von wem und wie (ob z.B. in einem Gespräch oder mit einem Testverfahren) in welcher Reihenfolge erheben wird und nach dieser Planung die Untersuchung durchführen.

Datenauswertung. Nach Jäger (1982) hat die Datenauswertung das Ziel, die erhobenen Daten zu sichten, die Befunde zu ordnen und anschließend in psychologische Aussagen zu übersetzen. Es gilt demnach, auf Basis der erhobenen Daten zu entscheiden, wie intelligent das Kind bzw. wie körperlich belastbar Person A ist, z.B. in der Form *Das untersuchte Kind verfügt in Relation zu Kindern seiner Altersgruppe über eine unterdurchschnittliche allgemeine Intelligenz*. Es geht an dieser Stelle daher um eine Interpretation und Integration der Daten, wie sie in durchaus analoger Weise auch bei der Identifizierung der Evidenz (Trope & Liberman, 1996) bzw. der Bewertung der Informationen (Hager & Weißmann, 1991) vorgenommen wird.

Absicherung der Hypothesen. Im Einklang mit Hager und Weißmann (1991) in Bezug auf Hypothesenprüfung und Hypothesenbewertung geht es im Anschluss darum zu überprüfen, inwieweit die gefundenen Informationen (die soeben formulierten psychologischen Aussagen) sich mit den erwarteten (den vorab definierten Hypothesen) decken. Als Resultat gilt die Hypothese als abgesichert oder widerlegt.

Beantwortung der Fragestellung. Ist dies hinsichtlich aller Hypothesen geschehen, so kann die Ausgangsfragestellung beantwortet werden, d.h. es wird die Ursache des Leistungsabfalls identifiziert. Dieser Schritt ist notwendig, weil in der Regel viele Einzelhypothesen relevant sind und sticht daher etwas aus dem Ablauf von Trope und Liberman (1996) sowie Hager und Weißmann (1991) heraus. Da es sich aber um kein qualitativ unterschiedliches Merkmal handelt, sondern der Komplexität geschuldet ist, bleibt die Übereinstimmung der Abläufe nach Einschätzung der Verfasserin immer noch außerordentlich hoch.

2.2.3 Integration

Eine visuelle Veranschaulichung des hohen Grades an Übereinstimmung der drei skizzierten Modelle bietet Abbildung 1, die in je einem Drittel die oben skizzierten Modelle wiedergibt. Schritte, die entsprechend der obigen Erläuterungen als analog erachtet werden, sind einerseits auf derselben Höhe angeordnet und andererseits durch das Symbol des Pfeils aufeinander bezogen.

Der Veranschaulichung ist zu entnehmen, dass der *Ablauf einer Hypothesenprüfung* (vgl. Hager & Weißmann, 1991) als Alleinstellungsmerkmal die *Bewertung der Hypothese* aufweist. Der *diagnostische Prozess* (Jäger, 1982) unterscheidet sich von den anderen beiden Modellen nur durch seinen speziellen Rahmen: die *Fragestellung* und deren *Beantwortung*. Für die vorliegende Arbeit ist insbesondere die Beobachtung von Bedeutung, dass der diagnostische Prozess eine große konzeptuelle Nähe zum Hypothesentesten aufweist. Denn insofern scheint die natürlich in der Folge zu prüfende Annahme sinnvoll, dass auch im Zuge des diagnostischen Prozesses Fehlerquellen auftreten und zwar solche, die mit denjenigen beim sozialen Hypothesentesten in Verbindung stehen.

Diese Annahme steht zudem im Einklang mit dem Verständnis des Klinikers wie Leary und Miller (1986) es beschreiben, indem sie festhalten, dass “[...] *clinicians and counselors are not simply the passive recipient of whatever information their clients wish to divulge. Rather, they actively seek certain information, integrate that information to form ideas about the client and his or her dysfunction, and seek to test their hunches by gathering yet more information.*” (p. 135). Diese Annahme ist der schon in die 1960er Jahre zurückgreifenden Konzeptualisierung des Prozesses der Beratung bzw. Therapie als interpersonellem Einflussprozess zwischen Berater bzw. Therapeut und Klient geschuldet (Strong, 1968), die eine sozialpsychologische Untersuchung des Prozesses angezeigt erscheinen lässt.

Aufgrund der Vielzahl der vorzunehmenden Informationsverarbeitungsschritte im sozialen Kontext mit reichlich kognitiven Anforderungen ist es nun jedoch auch leicht vorstellbar, dass es eine Vielzahl von Fehlerquellen im Testungsprozess geben kann. Im Rahmen des *Modells des konfirmatorischen Hypothesentests* von Schulz-Hardt und Köhnken (2000) werden diese konzeptuell vereint und im nächsten Schritt dargestellt.

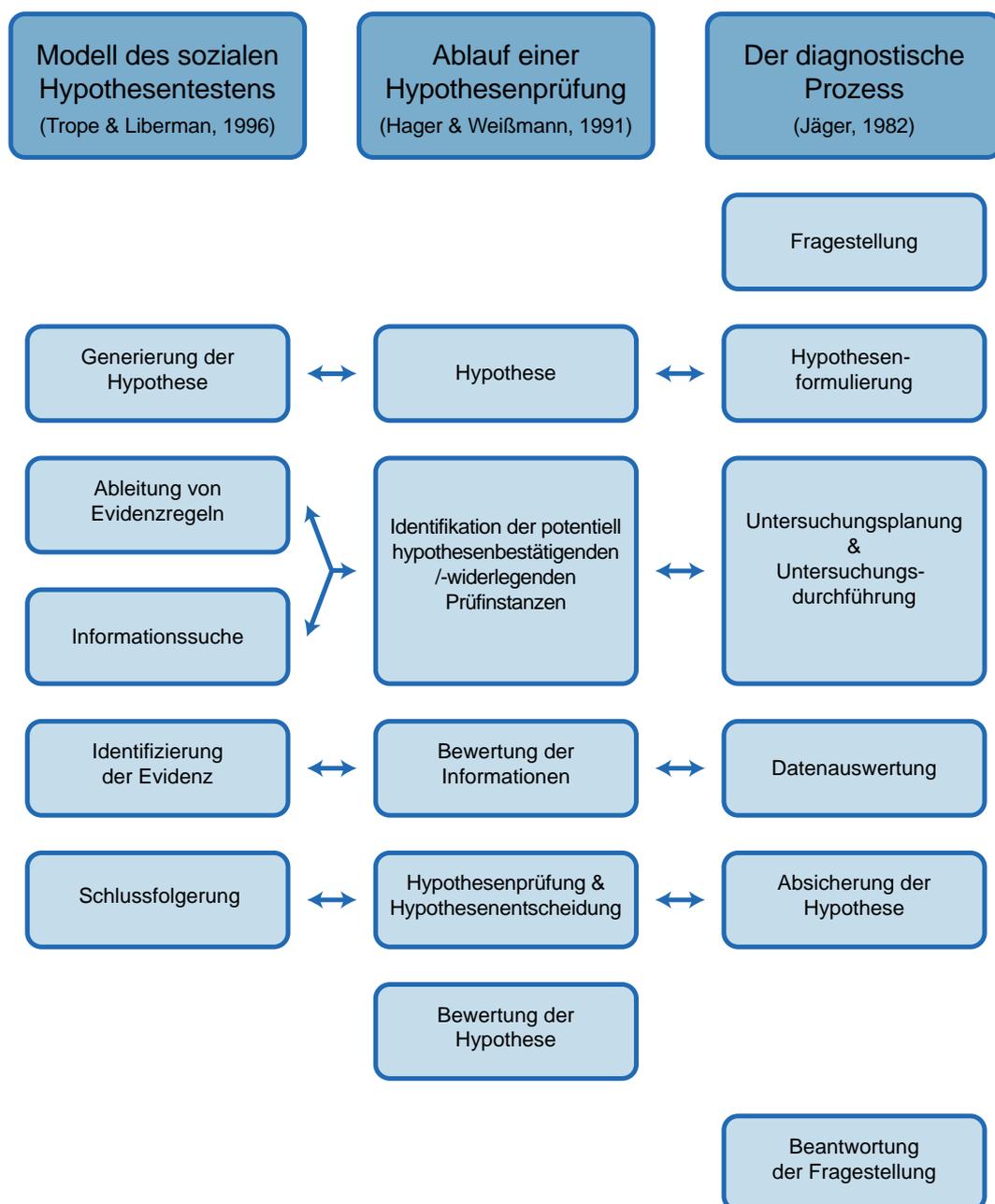


Abbildung 1. Integrative Darstellung dreier Konzeptionen des Ablaufs einer Hypothesentestung

2.3 Das Modell des konfirmatorischen Hypothesentestens

Das Modell des konfirmatorischen Hypothesentestens wurde 2000 von Schulz-Hardt und Köhnken vorgestellt und expliziert potentielle Verzerrungsmechanismen und ihre Wirkungsweise. Die Autoren nehmen drei Kategorien von beteiligten Mechanismen an: kognitive Prozesse, die Wahl der Teststrategie sowie motivationale Prozesse. Sie sprechen dann von konfirmatorischen bzw. Selbstbestätigungsmechanismen, wenn diese systematisch die Bestätigung der Hypothese begünstigen. Es soll nun zunächst überblicksartig das Modell vorgestellt werden, um einen Gesamteindruck zu ermöglichen.

2.3.1 Überblick

In Abbildung 2 sind alle Mechanismen überblicksartig dargestellt, bevor mit einer schrittweisen Erläuterung der einzelnen Mechanismen begonnen wird. Die theoretischen Überlegungen werden jeweils anhand von Beispielen veranschaulicht.

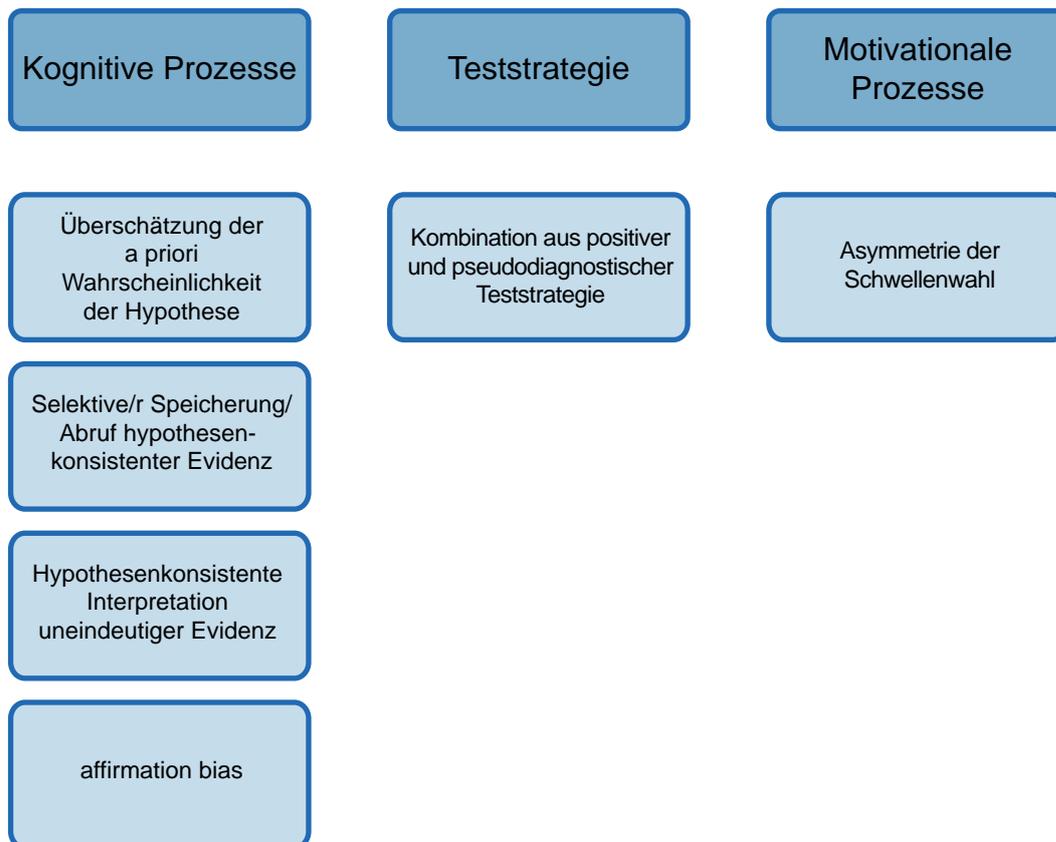


Abbildung 2. Veranschaulichung der konfirmatorischen Mechanismen (modifiziert nach Schulz-Hardt & Köhnken, 2000, S. 70)

Kognitive Prozesse. Die kognitiven Prozesse, die sich konfirmatorisch auswirken können, beziehen sich auf Verzerrungen, die während der Verarbeitung von Informationen eintreten können, wobei die Autoren davon ausgehen, dass diese Prozesse in der Regel unbewusst und ungewollt ablaufen und dementsprechend unabhängig davon auftreten, wie gewissenhaft, aufwändig und exakt die Informationen verarbeitet werden.

Unter dem Mechanismus der systematischen *Überschätzung der a priori Wahrscheinlichkeit der Prüfhypothese* wird das Phänomen verstanden, dass der Hypothesentester die Wahrscheinlichkeit, mit der die zu testende Hypothese ohne jedwede Berücksichtigung der vorliegenden Informationen wahr ist, höher einschätzt als diese tatsächlich ist. Dies ist problematisch, denn je wahrscheinlicher eine Hypothese ist, desto weniger Informationen wird der Tester suchen, bis er eine Entscheidung trifft, und bei einer Überschätzung würden weniger Informationen als gerechtfertigt bzw. sinnvoll eingeholt werden. Eine Überschätzung der Wahrscheinlichkeit der Hypothese würde bspw. resultieren, wenn der Mieter sich nicht mehr genau an das Wahlergebnis erinnert, aber aufgrund der aktuellen Nachrichten, die viele grüne Themen enthält den Zuspruch zur Partei mit 18% einschätzen würde, obwohl das Wahlergebnis tatsächlich bei 9% gelegen hat. Der Psychologe überschätzt womöglich auch die Basisrate, wenn er nicht die neueste epidemiologische Studie zur Bestimmung heranzieht, sondern aufgrund seiner Arbeit in einer auf Essstörungen spezialisierten Klinik mit einem hohen Anteil von 10% statt 1 bis 3% (Saß, Wittchen, Zaudig & Houben, 2003) ausgeht.

Ein verzerrender Einfluss wird auch der *selektiven Speicherung bzw. dem selektiven Abruf hypothesenkonsistenter Informationen* zugesprochen. Nach Schulz-Hardt und Köhnken (2000) handelt es sich hierbei um einen Filtermechanismus der Hypothese, der dazu führt, dass hypothesenkonsistente Informationen bei der Abspeicherung und dem Erinnern gegenüber hypotheseninkonsistenten Informationen einen Vorteil haben können. Obwohl der Mieter im Mehrfamilienhaus hinreichend motiviert und um eine objektive Suche in seinen Erinnerungen bemüht ist, fallen ihm (unbewusst!) nur solche Situationen ein, in denen der neue Nachbar mit dem Fahrrad weggefahren ist, eine Tüte aus dem Bio-Laden bei sich hatte oder der Brief eines Ökostromanbieters für den Nachbarn fälschlicherweise beim Mieter eingeworfen worden war. Dass der Nachbar neulich von der anstehenden Flugreise auf die Malediven erzählt hat oder nachts häufig das Licht draußen brennen lässt, würde ihm aufgrund der Filterfunktion nicht gleichermaßen leicht einfallen. Und für den Psychologen könnte dies bedeuten, dass ihm aus dem diagnostischen Gespräch mit der potentiellen Bulimikerin eher in Erinnerung bleibt, dass sie von nächtlichen Fressanfällen und intensivem Joggen erzählt als dass ihre Ausführungen erkennen lassen, dass sie keine Krankheitseinsicht

besitzt und sich sozial eher zurückzieht – beides Eigenschaften, die seltener mit einer Bulimie einhergehen.

Auch die *hypothesekonsistente Interpretation uneindeutiger Evidenz* stellt nach Ansicht der Autoren einen kognitiven Verzerrungsmechanismus dar, der genau dann vorliegt, wenn mehrdeutige, d.h. mit mehreren Alternativerklärungen in Einklang zu bringende Informationen bevorzugt im Sinne der zu testenden Hypothese interpretiert werden. Die Verzerrung besteht nun darin, dass die gefundene Information nicht als uneindeutig interpretiert wird, sondern tendenziell als hypothesekonsistent und damit bestätigend, weil sie ja immerhin mit der Hypothese vereinbar ist. Der Mieter im Mehrfamilienhaus hat es bspw. bei der Information *Der neue Nachbar fährt mit dem Fahrrad zur Arbeit* mit uneindeutiger Evidenz zu tun, denn dies ist nicht nur bei ökologischen Motiven denkbar, sondern auch bei gesundheitlichen (*Fahrrad fahren hält fit*), praktischen (*keine Parkplätze in der Innenstadt*) oder finanziellen Gründen eine sinnvolle Angelegenheit. Im Beispiel der Bulimie-Diagnostik stellt sich dieser Effekt wie folgt dar: Die Angabe *Die Patientin steht nach dem Essen sofort auf und geht in das Badezimmer* stellt eine objektiv uneindeutige Information dar, denn die Patientin könnte dort willentlich erbrechen oder sich einfach nur die Zähne putzen. Würde dies als Hinweis auf das Erbrechen der Patientin interpretiert, wäre diese fälschliche Interpretation als Beleg für das Vorliegen einer Bulimie eine entsprechende konfirmatorische Verzerrung.

Als vierten kognitiven Verzerrungsmechanismus nennen Schulz-Hardt und Köhnken (2000) den *affirmation bias*. Sie verstehen darunter die Tendenz, dass Menschen stärker auf das Auftreten als das Ausbleiben bestimmter Evidenz reagieren. Vom Schemm, Dreger und Köhnken (2008) führen weiter aus, dass diese Verzerrung sowohl auf einer Wahrnehmungs- als auch auf einer Bewertungsebene wirksam werden kann. Der wahrnehmungsmäßige *affirmation bias* bezieht sich darauf, dass Menschen zielsicherer die Existenz eines Ereignisses erfassen können als dessen Nonexistenz. Für den Mieter im Mehrfamilienhaus könnte der *affirmation bias* bedeuten, dass er vielleicht viel eher wahrnehmen wird, wenn der Nachbar am nächsten Tag seinen Fahrradhelm dabei hat, als wenn er am Briefkasten keinen Aufkleber mit dem Aufdruck *Atomkraft nein danke!* kleben hat. Den bewertungsmäßigen *affirmation bias* verstehen vom Schemm Dreger und Köhnken (2008) als unterschiedliche Einschätzung der Aussagekraft von anwesender vs. abwesender Evidenz. D.h. auch wenn es keinen wahrnehmungsmäßigen *affirmation bias* gab, da ein abwesendes Ereignis entdeckt wurde, so kann es trotzdem noch zu einer Verzerrung kommen, wenn nämlich den anwesenden Informationen ein höheres Gewicht beigemessen wird als den abwesenden. Die Autoren bedienen sich hier eines Beispiels aus der kriminaltechnischen Ermittlung und stellen die Frage, ob den gefundenen Blutspuren im Kofferraum (anwesende Information)

eines Verdächtigen wohl mehr Gewicht beigemessen wird als der Tatsache, dass registriert wurde, dass er keine Kampfspuren an den Armen aufweist (abwesende Information, die immerhin erkannt wurde).

Konfirmatorische Teststrategie. Unter dem Begriff der Teststrategie wird die grundsätzliche Methode verstanden, mit der man eine Hypothese überprüft, um zu einer Annahme oder Zurückweisung zu kommen. Dabei sind grundsätzlich verschiedene Strategien denkbar, z.B. können sich Teststrategien darin unterscheiden, ob eine Fokussierung auf die Prüfhypothese vorliegt oder auch die Alternativhypothesen beachtet werden, was als *pseudodiagnostisches* bzw. *diagnostisches Hypothesentesten* bezeichnet wird (vgl. Trope & Liberman, 1996). Außerdem kann sich die Teststrategie darin unterscheiden, ob bestätigende oder widerlegende Informationen gesucht werden, was auch als *positive* bzw. *negative Teststrategie* beschrieben wird (vgl. Schulz-Hardt & Köhnen, 2000).

Eine diagnostische Teststrategie zeichnet sich in diesem Zusammenhang durch die strikte Berücksichtigung von drei Merkmalen aus: die Diagnostizität und die Reliabilität der Evidenz sowie die Basisrate der Prüfhypothese. Eben diese Merkmale bleiben beim pseudodiagnostischen Vorgehen jedoch unberücksichtigt. Der Psychologe, betraut mit der Abklärung einer Bulimie-Erkrankung, berücksichtigt die Diagnostizität der Evidenz, wenn er bei der Bewertung der Information *die Patientin erbricht regelmäßig 1-2mal pro Tag* sowohl die Vereinbarkeit der Angabe mit den Diagnosekriterien für eine Bulimie abprüft, als auch berücksichtigt, dass ein regelmäßiges Erbrechen auch bei einer Anorexie auftreten kann. Die Information wird dahingehend bewertet, dass das Auftreten mit beiden Erklärungen vereinbar ist. Würde er lediglich feststellen, dass das Erbrechen mit der Prüfhypothese Bulimie vereinbar ist, würde er ein Merkmal der pseudodiagnostischen Strategie umsetzen. Die Reliabilität, also Zuverlässigkeit der Informationen bei einer Selbstauskunft und die Basisrate der Prüfhypothese, d.h. die Prävalenzrate für die Bulimie müssten ebenso beachtet werden.

Entscheidend für die Frage, ob eine konkrete Teststrategie eine positive oder negative Ausrichtung hat, ist die Art der gesuchten Evidenz. Wird Evidenz gesucht, die unter den von der Hypothese vorhergesagten Bedingungen wahrscheinlich ist, handelt es sich um eine positive Teststrategie. Eine negative Strategie verbirgt sich hingegen hinter der Suche nach Informationen bzw. Ereignissen, die nicht vorliegen dürften, wenn die Hypothese wahr ist. Im Beispiel des Mieters im Mehrfamilienhaus, der die Hypothese über die politische Ausrichtung des neuen Nachbarn testet, entspricht die Suche nach einer Information mithilfe der Frage *Trennt der Nachbar seinen Müll?* einer Frage im Sinne

der positiven Teststrategie. Denn das Merkmal *Müll trennen* ist eine Verhaltensweise die man Personen mit einer hohen ökologischen Verantwortlichkeit, wie den Anhängern der Grünen, unterstellt. Im Beispiel des Psychologen ist das Merkmal *Auftreten von Heißhungeranfällen mit Kontrollverlust* ein Ereignis, welches von der Hypothese Bulimie vorhergesagt wird. Der Mieter im Mehrfamilienhaus würde eine negative Teststrategie verfolgen, wenn er danach fragen würde, ob der neue Nachbar eine Fernreise mit dem Flugzeug geplant hat, weil man Letzteres bei einer Person mit hoher ökologischer Verantwortlichkeit nicht vermuten würde. Der Psychologe würde im Sinne einer negativen Strategie das Vorliegen von substantiellem Untergewicht abklären, da dieses gegen eine Bulimie sprechen würde.

Wie im Abschnitt 2.3.3.3 diskutiert wird, muss jedoch eine Kombination beider ungünstiger Strategien, also einer sowohl positiven als auch pseudodiagnostischen Strategie vorliegen, um ein konfirmatorisches Potential zu entfalten.

Motivationale Prozesse. Die dritte Kategorie der potentiellen Verzerrungsmechanismen stellen die motivationalen Prozesse dar, die sowohl von Trope und Liberman (1996) als auch Schulz-Hardt und Köhnken (2000) sehr eng gefasst werden und sich ausschließlich auf eine asymmetrische Wahl der Schwellen zur Annahme und Zurückweisung einer Hypothese beziehen.

Trope und Liberman (1996) nehmen an, dass sich der Tester zwei Schwellenwerte setzt: eine Schwelle für die Akzeptanz und eine für die Zurückweisung der Hypothese. Die Schwelle für die Akzeptanz beschreibt die minimale Überzeugung die notwendig ist, bevor die Hypothese akzeptiert wird. Die Höhe der Schwelle hängt dabei von der Höhe der antizipierten Kosten einer Fehlentscheidung ab. Je größer die vermuteten Kosten sind, desto höher wird der Schwellenwert ausfallen. Werden die Kosten einer falschen Annahme anders bewertet als die Kosten einer falschen Zurückweisung, fallen auch die Schwellenwerte unterschiedlich aus. Wählt der Tester nun eine niedrige Schwelle für die Annahme der Hypothese und eine höhere für deren Zurückweisung, weil er die Kosten für eine fälschliche Zurückweisung höher beurteilt als die Folgen einer fälschlichen Annahme der Hypothese, so reichen ihm zur Annahme der Hypothese weniger Informationen als zu deren Zurückweisung. Dies stellt eine weitere potentielle Begünstigung der zu testenden Hypothese und damit einen Verzerrungsmechanismus dar.

Nachdem zunächst ein grober Überblick über alle Verzerrungsmechanismen gegeben wurde, werden im Folgenden einschlägige empirische Arbeiten und theoretische Überlegungen zu den einzelnen Mechanismen referiert und bewertet. Hierfür sind

einerseits die Arbeiten aus der sozialen Kognitionsforschung relevant, aus denen die Autoren das Modell ableiten. Andererseits werden zusätzlich diejenigen empirischen Befunde aus der Untersuchung der klinisch-psychologischen Diagnostik angesprochen, die mit dem jeweiligen Mechanismus in Verbindung gebracht werden können. Dabei ist zu bedenken, dass letztere Forschungsarbeiten weitgehend unabhängig von dem Modell von Schulz-Hardt und Köhnken (2000) durchgeführt wurden und daher die Zuordnung zu den einzelnen Mechanismen eine subjektive Entscheidung der Verfasserin ist. Wenn kein eigenständiger Abschnitt zu einschlägigen Arbeiten aus der klinisch-psychologischen Diagnostik an den Erläuterung des Mechanismus angeschlossen ist, lagen nach den Recherchen der Verfasserin keine spezifischen Studien vor.

2.3.2 *Kognitive Verzerrungsprozesse*

Es erfolgt zunächst die differenzierte Abhandlung der vier Verzerrungsprozesse bei der Informationsverarbeitung.

2.3.2.1 *Überschätzung der a priori Wahrscheinlichkeit der Prüfhypothese*

Formal entspricht die a priori Wahrscheinlichkeit der Basisrate, mit der ein interessierendes Merkmal bei einem Objekt aus der zugehörigen Grundgesamtheit vorliegt.

Wenn der Mieter im Mehrfamilienhaus die a priori Wahrscheinlichkeit des Zutreffens seiner Hypothese einschätzen möchte, so ist u.a. das letzte Wahlergebnis der Grünen eine sinnvolle Annäherung an die Basisrate. Im Beispiel des Diplom-Psychologen, der mit der Abklärung des Verdachts auf das Vorliegen einer Bulimie-Erkrankung betraut wird, entspricht die a priori Wahrscheinlichkeit der Prävalenzrate der Bulimie.

An beiden Beispielen wird bereits deutlich, dass der Hypothesentester in den meisten, wenn nicht allen Fällen keine exakte Angabe der a priori Wahrscheinlichkeit wird machen können, sondern immer eine Schätzung vornehmen muss, die mehr oder weniger korrekt sein kann. Geringe Abweichungen aufgrund des Schätzvorgangs dürften jedoch keine gravierenden Konsequenzen haben, anders jedoch die systematische Überschätzung der a priori Wahrscheinlichkeit. Es bleibt zu klären, ob die Überschätzung zu einer Verzerrung im Sinne des *confirmation bias* führen kann, und darüber hinaus stellt sich die Frage, wie es zu einer solchen Überschätzung kommen kann. Einen Beitrag zu beiden Fragen liefert die Arbeit von Hodgins und Zuckerman (1993).

Sie konnten zeigen, dass eine erhöhte a priori Wahrscheinlichkeit unabhängig von dem konkreten Vorgehen und der erhaltenen Evidenz eine erhöhte Wahrscheinlichkeit zur Bestätigung der Hypothese bedingt. Probanden, die gebeten wurden die Basisraten von zu testenden Persönlichkeitsmerkmalen einer Zielperson als Prozentsatz der Studierenden zu schätzen, die an ihrer Universität über die entsprechenden Persönlichkeitsmerkmale (z.B. Extraversion, Pessimismus, Liberalität) verfügten, waren im späteren Verlauf der Bewertung von Informationen durch die Höhe ihrer geschätzten Basisrate beeinflusst: je höher die angenommene a priori Wahrscheinlichkeit war, desto höher wurden der Erklärungswert und die Sicherheit der Schlussfolgerungen angesehen, mit der aus den fiktiven Antworten der Zielperson auf die Hypothese zurück geschlossen werden könnte. Es resultierte folglich ein linearer Zusammenhang zwischen der Höhe der angenommenen Basisrate des Persönlichkeitsmerkmals und der Sicherheit bzgl. der Schlussfolgerungen. Die Wahrscheinlichkeit, die Hypothese zu bestätigen hing dementsprechend von der angenommenen Basisrate ab und zeigt somit die potentielle Gefahr der Überschätzung der a priori Wahrscheinlichkeit an.

Die Autoren erklären sich dieses Resultat dadurch, dass durch die Tatsache, dass eine Hypothese überhaupt geprüft wird, die Verfügbarkeit bzw. Zugänglichkeit erhöht wird, indem sie mit dem so genannten *hypothesis bias* argumentieren, den sie wie folgt umschreiben: „*the mere act of testing a hypothesis could make it more available or accessible and thus more subjective likely to occur*“ (p. 390). Daraus ist ableitbar, dass die Vorgabe von bzw. Beschäftigung mit einer Hypothese dazu führen kann, dass die angenommene a priori Wahrscheinlichkeit eine Überschätzung nach sich ziehen kann, was durch die erhöhte Verfügbarkeit erklärbar ist. In diesem Zusammenhang ist die Verfügbarkeitsheuristik (z.B. Strack & Deutsch, 2002) näher zu beleuchten, die vor allem dann angewendet wird, wenn Häufigkeiten oder Wahrscheinlichkeiten geschätzt werden müssen. Demzufolge kann diese Heuristik auch bei der Schätzung der a priori Wahrscheinlichkeit einer Hypothese zum Einsatz kommen. Nach Halder-Sinn (1993) bedingt die Leichtigkeit des Zugangs zu abgespeicherten Informationen die Schätzung der Häufigkeit dieser Informationen. Im Einklang mit diesem Verständnis veranschaulichen Strack und Deutsch (2002) die Strategie mit der Faustregel *Wenn mir ein Ereignis leicht einfällt, dann wird es wohl häufig auftreten*, oder *Wenn ich mir ein Ereignis leicht vorstellen kann, dann wird es wohl häufig vorkommen* anführen. Dementsprechend werden Ereignisse, die leichter erinnert/vorge stellt werden können für wahrscheinlicher gehalten. Man mag sich nun fragen, was an dieser Vorgehensweise problematisch ist, denn es ist plausibel, dass in vielen Fällen der Zusammenhang von Auftretenshäufigkeit und Leichtigkeit der Erinnerung begründet ist. Problematisch ist jedoch, dass einige seltene Ereignisse, d.h.

solche mit geringer Basisrate, durch eine besonders prägnante Erscheinung oder eine immense persönliche Bedeutung sehr gut erinnert oder vorgestellt werden können. Nach der Verfügbarkeitsheuristik würden diese ebenfalls leicht erinnerten/vorgestellten Phänomene auch eine höhere Basisrate erhalten als objektiv gerechtfertigt ist. Dementsprechend ist der Mieter im Mehrfamilienhaus durch die Berichterstattung dahingehend beeinflusst, dass Ereignisse wie in Fukushima oder bei Stuttgart 21 die Zustimmung zu grünen Themen leicht vorstellbar machen und dann ist die Schlussfolgerung durchaus plausibel, dass es wohl häufig vorkommen muss, dass die Grünen gewählt werden.

Ergänzend sei erwähnt, dass vom Schemm (2008a) auch die emotionale Tönung der Hypothese als mögliche Ursache für die Überschätzung anführt und dies u. a. am Beispiel des sexuellen Kindesmissbrauchs illustriert, der für die Eltern mit einem entsprechenden Verdacht hochgradig emotional belegt ist und so die Schätzung der Basisrate beeinflussen könnte.

2.3.2.2 *Selektive Speicherung bzw. selektiver Abruf hypothesenkonsistenter Evidenz*

Empirisch konnte der Gedächtnisvorteil hypothesenkonsistenter Informationen bei der Speicherung und beim Abruf unter anderem von Snyder und Cantor (1979) gezeigt werden. Ihre Probanden erhielten zum ersten Zeitpunkt ausführliche Informationen (extravertierte vs. introvertierte Merkmale) über den Verlauf einer Woche im Leben einer (fiktiven) Person Jane. Zwei Tage später vor die Aufgabe gestellt, die Berufseignung von Jane als Immobilienmaklerin (extravertierte Personen erwartet) oder Bibliothekarin (introvertierte Personen erwartet) zu bewerten, riefen die Probanden häufiger hypothesenkonsistente Informationen ab und bewerteten die Berufseignung ebenfalls in der Richtung der vorherrschenden Hypothese. Hierin zeigt sich der Mechanismus des selektiven Abrufs und dessen konfirmatorische Auswirkung.

Dieser Befund mag zunächst etwas verwundern. Denn weil inkonsistente Informationen unerwartet und überraschend sind, scheint es nahe liegend, dass ihnen mehr Aufmerksamkeit zukommt und sie intensiver verarbeitet werden. Die tiefere Verarbeitung könnte ihrerseits eine vielfältige Assoziation mit weiteren Angaben zur Folge haben, so dass auch ein Vorteil für hypotheseninkonsistente Angaben möglich ist (vgl. Oswald & Gadenne, 1986). So herrscht auch in der Gedächtnispsychologie der Befund vor, dass sowohl konsistente als auch inkonsistente Informationen besser erinnert werden als neutrale und ein klarer Vorteil von konsistenten Daten erklärungsbedürftig ist. Hilfreich in diesem Zusammenhang ist die Metaanalyse von Stangor und McMillan (1992) über den Einfluss von erwartungskonformen und erwartungswidrigen Informationen auf die

Erinnerungsleistung. Einerseits bestätigte diese Studie den generellen Vorteil von sowohl konsistenten als auch inkonsistenten Angaben gegenüber neutralen Informationen, aber andererseits konnten auch relative Erinnerungsvorteile für beide Informationstypen in Moderationsanalysen gefunden werden. Zusammenfassend können ihre Erkenntnisse dahingehend festgehalten werden, dass sich für hypothesenkonsistente Angaben u. a. insbesondere dann ein Erinnerungsvorteil ergibt, wenn die Verarbeitung der Informationen unter kognitiv stark fordernden Bedingungen erfolgt und das Stimulusmaterial verhältnismäßig mehrdeutig ist. Dass der Psychologe bei der Abklärung der Bulimie-Erkrankung kognitiv stark gefordert ist und die erhobenen Daten durchaus mehrdeutig sein können, weil ähnliche Symptome bei vielen Störungen auftreten können und eine ausgefeilte Differentialdiagnostik notwendig machen, ist unmittelbar einsichtig, so dass im Kontext der klinisch-psychologischen Diagnostik weitgehend von Rahmenbedingungen ausgegangen werden kann, die einen Erinnerungsvorteil der konsistenten Informationen bedingen.

Wie sich dieser Vorteil manifestieren kann, versuchen Wyer und Scull (1986) zu erklären. Sie bringen als mögliche Erklärung für den Gedächtnisvorteil von hypothesenkonsistenten Informationen die Annahme ins Spiel, dass Schemata eine Rolle für die Informationsverarbeitung spielen. Nach Stangor und McMillan (1992) können die hypothesenkonsistenten Informationen leichter in die Schemata-Repräsentationen integriert werden als die inkonsistenten, so dass sich eine stärkere Assoziation zwischen den konsistenten Angaben und dem Schema ergibt, welches von der Hypothese aktiviert wird. Dieser Vorteil bewirkt bei der anschließenden Bewertung der Informationen eine Verschiebung der Anteile konsistenter und inkonsistenter Informationen zugunsten der hypothesenkonsistenten Angaben. Die Wissensakkumulation von inkonsistenten Informationen könnte dementsprechend systematisch beeinträchtigt sein und die Waagschale der konsistenten Informationen bekäme mehr Gewicht, so dass die Wahrscheinlichkeit zur Annahme der Hypothese steigt, was die Voraussetzung ist, um als konfirmatorischer Prozess interpretiert zu werden.

Befunde aus der Forschung zur klinisch-psychologischen Diagnostik. Es finden sich sowohl Hinweise auf eine zu Gunsten der konfirmatorischen Informationen verzerrte Rekognitions- (Arkes & Harkness, 1980) als auch Abruffleistung (Strohmer et al., 1990), wenn Diagnostiker sich auf ihr Gedächtnis berufen müssen, um Informationen über eine zu begutachtende Person zu verwenden.

Arkes und Harkness (1980) haben aufgezeigt, dass Personen, die eine Diagnose stellen, später weniger gut erinnern können, welche Symptome, die zur Hypothese passen, genau

vorlagen. Zwar war die Erinnerung an tatsächlich vorhandene, hypothesenkonsistente Informationen genauer und neue irrelevante Items konnten besser zurückgewiesen werden (Experiment 2), allerdings tendierten die Probanden klar dazu, zuvor nicht präsentierte Informationen wiederzuerkennen, wenn diese mit der Diagnose im Einklang standen (Experiment 1). Besonders alarmierend scheint der Befund aus der dritten Studie, in der erstmals ein inkonsistentes Item in das Set der zu erinnernden Informationen aufgenommen wurde. Nun schnitt die Gruppe mit der Diagnose auf allen Maßen schlechter ab als die Gruppe ohne Diagnose. Die Autoren empfehlen daher eindringlich, nicht nur die gestellte Diagnose zu notieren, sondern auch die Symptome, auf denen sie beruht, um eine Überprüfung im Nachhinein möglich zu machen, da einem späteren Abruf aus dem Gedächtnis aufgrund der verzerrenden Prozesse nicht hinreichend vertraut werden sollte.

Nicht auf eine Wiedererkennung, sondern Abrufaufgabe fokussierend, ließen Strohmeyer et al. (1990, Experiment 1) ihren Probanden (im Mittel fünf Jahre Berufserfahrung) eine von drei Fallgeschichten per Post zukommen, die sich hinsichtlich der Relation von konfirmatorischen und diskonfirmatorischen Informationen bzgl. der Hypothese einer mangelnden Selbstkontrolle unterschieden (hoch konfirmatorisch, ausgeglichen, niedrig konfirmatorisch). Eine Woche später bekamen sie weitere Materialien zugeschickt und wurden aufgefordert, mindestens zehn Informationen zu erinnern, die ihnen für die Entscheidung relevant erscheinen. Sie produzierten daraufhin unabhängig von der Relation der Informationen mehr konfirmatorische als diskonfirmatorische.

Ihre Selbstkritik aufgreifend, erhöhten Strohmeyer und Shivy (1994, Experiment 1) im Einklang mit den Forderungen von Smith und Kida (1991; s. auch Edwards, 1983 sowie Hogarth, 1981) zunächst den Realitätsgehalt der Aufgabe, indem sie erfahrenen Diagnostikern (im Mittel 15 Jahre) einen hypothesenblinden, aber geschultem *mock client* an deren Arbeitsplatz vorstellten, der entsprechend des Fallmaterials aus früheren Studien (Strohmeyer et al., 1990) je zehn konfirmatorische und zehn diskonfirmatorische Informationseinheiten (in Bezug auf die Hypothese einer mangelnden Selbstkontrolle) übermittelte. Im Anschluss an diese Präsentation, die nicht unterbrochen werden durfte, wurden die Probanden gebeten, mindestens zehn solcher Faktoren anzugeben, die ihnen entscheidend für die Beurteilung vorkamen, ob dem Klienten Selbstkontrolle fehlt oder nicht. Trotz des ausgewogenen Verhältnisses von konfirmatorischen und diskonfirmatorischen Informationen wurden mehr konfirmatorische als diskonfirmatorische Informationen erinnert bzw. angegeben. In Experiment 2 (nun wieder mit schriftlichem Fallmaterial und Diagnostikern im Training) manipulierten die Autoren die Verantwortlichkeit für die Entscheidung, indem ihnen vor der Niederschrift der zehn Informationen mitgeteilt wurde, sie müssten ihre Antworten im Anschluss mit

einem anderen Diagnostiker diskutieren. Aber auch hier zeigte sich die Präferenz für konfirmatorische Informationen, so dass von einem robusten Gedächtniseffekt gesprochen werden kann.

2.3.2.3 *Hypothesenkonsistente Interpretation uneindeutiger Evidenz*

Einen Hinweis darauf, dass mehrdeutige Informationen bevorzugt im Sinne der zu testenden Hypothese interpretiert werden, liefert die Studie von Lord et al. (1979). Sie legten Befürwortern bzw. Gegnern der Todesstrafe zwei Studien über die Abschreckungswirkung dieser Strafe vor (s. auch Abschnitt 2.1.2). Uneindeutigkeit wurde in dieser Studie hergestellt, indem zwei sich widersprechende empirische Arbeiten vorgelegt wurden. Dass es nicht zu einer Anerkennung der Uneindeutigkeit und damit Reduktion der Sicherheit in die Voreinstellung zur Todesstrafe kam, sondern zu einer weiteren Bestätigung der eigenen Annahme, illustriert die bevorzugte hypothesenkonsistente Interpretation in erschreckender Konsequenz.

Wie aber kann man sich diesen Umstand erklären, der bei nüchterner Betrachtung so abwegig erscheint? Eine mögliche Erklärung liefern die Ergebnisse einer Studie von Evett, Devine, Hirt und Price (1994). Die Autoren untersuchten die Frage, ob die Aufgabe, eine spezifische Hypothese zu testen dazu führt, dass beim Tester Erwartungen erzeugt werden, welche Art von Evidenz er erhalten wird. Dafür wurde das klassische Paradigma des *trait hypotheses testing* (vgl. Snyder & Swann, 1978) mit einer Erweiterung umgesetzt: nach der Vorgabe einer Prüfhypothese (extravertiert vs. introvertiert vs. extravertiert oder introvertiert) und der Auswahl von einer aus zwölf Fragen sollten die Probanden angeben, welche Antwort der Zielperson sie erwarteten und dies begründen. Die Auswertung ergab, dass 90,4% entweder ein Ja oder ein Nein als Antwort erwarteten. Die weitergehende Analyse zeigte, dass 92,3% der Probanden auf eine hypothesenkonsistente Frage ein Ja erwarteten und 96,6% der Personen, die nach inkonsistenten Informationen fragte, ein Nein zu erhalten glaubten. Die überwältigende Mehrheit der Probanden erwartete also Informationen, die die Prüfhypothese bestätigen würden.

Diese flächendeckende Erzeugung von Erwartungen bzgl. der Aussagekraft der zu erhaltenden Informationen bietet nun Raum für die Annahme, dass die Wahrnehmung und Kategorisierung der folgenden Evidenz dahingehend beeinflusst wird, dass die Uneindeutigkeit der Evidenz nicht erkannt wird, sondern eine prüfhypothesenkonsistente Interpretation vorgenommen wird. In diesem Sinne sind auch die Ergebnisse von Darley und Gross (1983) interpretierbar, die ihren Probanden eine Videoaufzeichnung eines

Mädchens bei der Bearbeitung von schulischen Aufgabenstellungen zeigten, bei denen das Mädchen mal erfolgreich war und mal versagte, so dass insgesamt uneindeutiges Material vorlag, bei dem neutrale Beobachter als Kontrollpersonen eine altersangemessene Leistung des Kindes wahrnahmen. Beurteilten nun die wirklichen Probanden das Fähigkeitslevel des Mädchens auf Basis des identischen Videos, nachdem sie zuvor unterschiedliche Informationen über den sozioökonomischen Status des Kindes (hoch vs. niedrig) erhalten hatten, hing das Urteil mit dem induzierten Status zusammen: während Probanden aus der Gruppe mit nahegelegtem hohem sozioökonomischen Status das Fähigkeitsniveau des Mädchens über dem Klassenstufenniveau einschätzten, beurteilte die Gruppe mit nahegelegtem niedrigen sozioökonomischen Status das Mädchen als signifikant weniger leistungsstark. Ein weiteres Mal hat sich die Hypothese bzw. Erwartung als Filter erwiesen und durch die Konstruktion von Erwartungen die Uneindeutigkeit der Informationen in den Hintergrund treten lassen, so dass Raum für Verzerrungen frei wurde.

Befunde aus der Forschung zur klinisch-psychologischen Diagnostik. Strohmer et al. (1990, s. auch Abschnitt 2.3.2.2 für Experiment 1) beschäftigten sich in ihrem zweiten Experiment mit der Frage, ob es trotz einer Ausschaltung des Gedächtniseffektes zu einer bevorzugten Nutzung der konfirmatorischen gegenüber den disconfirmatorischen Informationen kommt, da sie in Anlehnung an Darley und Gross (1983) annehmen, dass “[...] *the counselor who has developed a working client hypotheses may unknowingly have his or her attention diverted to only certain types of client data. [...] In fact, it might even be said that the client hypothesis leads to the counselor’s expectancy for attending to and selecting only certain types of client data.*” (p. 469). Mit Ausnahme des Umstandes, dass den Probanden die Fallgeschichte vorlag, während sie die Informationen notierten, die zur Beantwortung der Frage relevant seien, war das Vorgehen weitgehend analog zu Experiment 1. Dasselbe gilt für die Ergebnisse: unabhängig von der Relation von konfirmatorischen zu disconfirmatorischen Informationen werden mehr konfirmatorische aufgelistet, was hier nicht als Gedächtniseffekt, sondern Erwartungseffekt zu interpretieren ist.

2.3.2.4 *Affirmation bias in der Inferenz aus der vorliegenden Evidenz*

Wie bereits beschrieben, handelt es sich beim *affirmation bias* um die Tendenz, stärker auf das Auftreten als das Ausbleiben bestimmter Evidenz zu reagieren, indem entweder die abwesende Evidenz weniger zielsicher wahrgenommen wird als die anwesende (*wahrnehmungsmäßiger affirmation bias*) oder die wahrgenommene Evidenz über abwesende

Ereignisse weniger gewichtet wird als die Evidenz über Anwesendes (*bewertungsmäßiger affirmation bias*). Dies kann dazu führen, dass eine Hypothese trotz einer objektiv geringen Zahl belegender Merkmale bestätigt wird.

Der *wahrnehmungsmäßige affirmation bias* steht im Einklang mit den Erkenntnissen von Newman, Wolff und Hearst (1980), die gezeigt haben, dass es Menschen leichter fällt, aus auftretenden Ereignissen eine Regel abzuleiten, statt aus nicht auftretenden Ereignissen. Beispielsweise mussten ihre Probanden herausfinden, wann Symbole einer guten vs. einer schlechten Kategorie angehörten. War das Kriterium durch die Abwesenheit eines Dreiecks gegeben, brauchten die Probanden deutlich länger, falls sie die Regel überhaupt herausfanden, als die Probanden, für die die guten Karten durch die Anwesenheit des Dreiecks gekennzeichnet war.

Auf empirischer Ebene haben sich Hodgins und Zuckerman (1993, s. auch Studie zur Überschätzung der a priori Wahrscheinlichkeit) auch mit dem *affirmation bias* beschäftigt, den sie jedoch – vermutlich an ihr Studiendesign angelehnt – *question bias* nennen. Sie ließen ihre Probanden zwei Fragen generieren und anschließend angeben, welchen Schluss sie aus der Bejahung (zu verstehen als anwesende Information) und der Verneinung (analog abwesende Information) hinsichtlich der Validität der Hypothese ziehen würden und wie sicher dieser Schluss sei. Die Probanden gaben mehrheitlich an, nach einer Bejahung ihrer Frage sicherer hinsichtlich einer Entscheidung zu sein als nach einer Verneinung – ein klarer *affirmation bias*, hier auf der Bewertungsebene.

Aus der Tatsache, dass das Ausbleiben von hypothesenkonsistenter Evidenz weniger gewichtet wird als das Eintreten konsistenter Evidenz, ist darüber hinaus abzuleiten, dass widersprüchlichen Informationen weniger Gewicht beigemessen wird als den bestätigenden Angaben. Denn das Ausbleiben einer konsistenten Angabe stellt unter Umständen einen Widerspruch dar, falls nämlich die betreffende Angabe eine notwendige Bedingung für das Zutreffen der Hypothese darstellen sollte. Die geringere Gewichtung inkonsistenter Informationen konnte unter anderem von Gadenne und Oswald (1986) gezeigt werden, die ihren Probanden eine Kriminalgeschichte vorgaben, die eine Hypothese über die Täterschaft einer bestimmten Person nahelegte. Anschließend gaben sie bestätigende und widerlegende Informationen vor. In ihrem ersten Experiment konnten sie zeigen, dass die Probanden die Relevanz der stark widersprechenden Information zwar erkennen, dieses Erkenntnis jedoch keinen Einfluss auf die Überzeugung hinsichtlich der Hypothese hatte. In einem zweiten Experiment wurde klar, dass es dafür zusätzlich eines Hinweises auf eine Alternativhypothese bedarf. Insgesamt zeigte sich folglich, dass widersprüchliche Informationen weniger gewichtet wurden als bestätigende Angaben, solange keine plausible Alternativerklärung präsent war.

Befunde aus der Forschung zur klinisch-psychologischen Diagnostik. Einen empirischen Befund zum *wahrnehmungsmäßigen affirmation bias* liefert die Studie von Potchen (2006), die hier trotz ihrer medizinischen statt psychologischen Ausrichtung Erwähnung finden soll. Diese Studienreihe zur *observer performance in chest radiology* baut im Wesentlichen darauf auf, dass eines von 60 zu beurteilenden Röntgenbildern eine Auffälligkeit beinhaltet, die sich durch das Fehlen eines Knochens auszeichnet (Schlüsselbein). Erfasst werden sollte die individuelle Fähigkeit, ein fehlendes statt eines pathogenen, anwesenden Merkmals zu erfassen. Es zeigte sich, dass 60% der Probanden diese (fehlende) Abnormalität nicht identifizierten und damit Schwierigkeiten offenbarten, die Nonexistenz eines Ereignisses bzw. des an sich normalen Zustands des vorhandenen Knochens zu erfassen. Erst durch einen Hinweis, dass die Bilder im Zusammenhang mit chirurgischen Eingriffen aufgrund von Metastasen entstanden wären, sank die Quote des Übersehens auf 17%.

In Bezug auf den *bewertungsmäßigen affirmation bias* sei auf die Argumentation von Turk und Salovey (1985) verwiesen, die darlegen, dass Kliniker in ihrem Arbeitskontext erwarten, pathologisches Verhalten zu sehen und diese Erwartung zu einer Überschätzung von tatsächlich pathologischem bzw. einer Unterschätzung von positiven, d.h. nicht pathologischen Eigenschaften führen kann. Dies kann durchaus im Sinne des *affirmation bias* interpretiert werden, d.h. die Anwesenheit von Symptomen (Pathologie) würde eher wahrgenommen bzw. stärker gewichtet als die Abwesenheit von Symptomen (Non-Pathologie). Diesen Effekt der ungleichen Gewichtung konnte Ganzach (2000) bei der Analyse des Gebrauchs von MMPI-Skalen zur Diagnosestellung zeigen.

2.3.3 Konfirmatorische Teststrategie

Zunächst wird im Abschnitt 2.3.3.1 die diagnostische Teststrategie von der pseudodiagnostischen abgegrenzt und diskutiert, wie die Sensitivität des Hypothesentesters für die Diagnostizität der Evidenz einzuschätzen ist. Es folgt die Abhandlung der positiven bzw. negativen Teststrategie (2.3.3.2), bevor im Kapitel 2.3.3.3 der konfirmatorische Einfluss der Kombination beider Vorgehensweisen erläutert werden kann.

2.3.3.1 Diagnostische vs. pseudodiagnostische Teststrategie

Die Unterscheidung, ob eine Teststrategie als diagnostisch oder als pseudodiagnostisch zu kennzeichnen ist, hängt maßgeblich davon ab, ob der Hypothesentester die Alternativhypothesen zu seiner Prüfhypothese mit in Betracht zieht oder nicht (vgl. Trope & Liberman, 1996). Eine diagnostische Teststrategie zeichnet sich in diesem Zusammenhang durch die strikte Berücksichtigung von drei Merkmalen aus: die Diagnostizität und die Reliabilität der Evidenz sowie die Basisrate der Prüfhypothese. Eben diese Merkmale bleiben beim pseudodiagnostischen Vorgehen jedoch unberücksichtigt. Der Testende richtet hier seine Aufmerksamkeit selektiv auf die zu testende Hypothese und ignoriert die alternative(n) Erklärung(en). Daraus resultiert vor allem, dass die vorhandene Evidenz lediglich auf die Konsistenz mit der zu testenden Hypothese überprüft wird, nicht aber auf die Konsistenz mit der Alternativhypothese. Die pseudodiagnostischen Schlussfolgerungen werden in dem Sinne gezogen, dass mit prüfhypothesenkonsistenter Evidenz gleichzeitig alternativhypotheseninkonsistente Evidenz gefunden wird. Es wird also lediglich geprüft, ob die Hypothese hinreichend für das Auftreten der Evidenz ist und nicht, ob die Hypothese auch die notwendige Bedingung darstellt. Auch Letzteres ist jedoch obligatorisch für die diagnostische Strategie.

Den zentralen und zugleich kognitiv anspruchsvollsten Aspekt der diagnostischen Strategie stellt die Berücksichtigung der Alternativhypothese(n) dar, die es erst ermöglicht, die Aussagekraft einer Information zu bewerten. Der (diagnostische) Wert einer Evidenz kann nur dann adäquat geschätzt werden, wenn dem Tester bekannt ist, wie wahrscheinlich das Auftreten des Merkmals unter der Prüf- und der entsprechenden Alternativhypothese ist. In diesem Sinne beschreiben Trope und Bassok (1983, p. 561) den diagnostischen Wert einer Evidenz bzw. die diagnostische Strategie als „. . . *extent to which the probabilities of the answers depend on whether or not the respondent possesses the hypothesized trait. In this diagnostic strategy, the lay interviewer searches . . . for behavioral features that are distinctively associated either with the hypothesized trait category or with its alternative(s).*“

Die Diagnostizität der Evidenz bezieht sich demnach auf das Ausmaß, in dem sich die Auftretenswahrscheinlichkeiten eines Ereignisses unter der Prüfhypothese und der Alternativhypothese unterscheiden. Diagnostische Informationen haben eine große Auftretenswahrscheinlichkeit unter der einen Hypothese und eine geringe Wahrscheinlichkeit unter der anderen Hypothese. Nicht diagnostische Informationen hingegen haben ähnlich große Auftretenswahrscheinlichkeiten unter beiden Hypothesen. Diese Informationen sind nicht sinnvoll einsetzbar für eine Unterscheidung zwischen der Prüfhypothese und der Alternativhypothese.

Die Diagnostizität der Evidenz kann formal-mathematisch in Analogie zur verbalen Umschreibung einer differentiellen Auftretenswahrscheinlichkeit unter verschiedenen Hypothesen als Differenz der bedingten Wahrscheinlichkeiten des betreffenden Merkmals A unter den Alternativhypothesen H_1 und H_2 , konzeptualisiert werden (vgl. Skov & Sherman, 1986; Trope & Bassok, 1983), d.h.: $\text{Diagnostizität}(A) = p(A|H_1) - p(A|H_2)$. Die ebenfalls gebräuchliche Auffassung als das gewichtete Mittel der Verhältnisse der bedingten Wahrscheinlichkeiten des Vorliegens bzw. der Abwesenheit des Merkmals A unter zwei Hypothesen H_1 und H_2 ist hingegen weniger anschaulich.

Im Beispiel der Bulimie-Diagnostik stellt sich der Aspekt folgendermaßen dar: es werden sowohl bei Bulimikern als auch Anorektikern (vom Binge-Eating/Purging Typus) Heißhungerattacken mit dem Gefühl von Kontrollverlust berichtet. Die Wahrscheinlichkeit, Heißhungerattacken aufzuweisen, wenn man unter Bulimie ($P(\text{Heißhunger}|Bulimie)$) oder Anorexie, Binge-Eating/Purging Typus ($P(\text{Heißhunger}|Anorexie, \text{ Binge-Eating/Purging})$) leidet, ist annähernd gleich groß. Gibt die Patientin an, unter Heißhungerattacken zu leiden, stellt dies eine nicht diagnostische Information dar (Differenz ≈ 0). Das Gewicht von Bulimikern und Anorektikern unterscheidet sich hingegen in der Regel. Die Wahrscheinlichkeit, substantielles Untergewicht aufzuweisen ist bei Anorektikern deutlich höher als bei Bulimikern. Gibt die Patientin an, Untergewicht zu haben, ist dies eine diagnostische Information: $P(\text{Untergewicht}|Anorexie) > P(\text{Untergewicht}|Bulimie)$.

Die prinzipielle Fähigkeit des Hypothesentesters, Alternativhypothesen zu berücksichtigen und die Aussagekraft der Evidenz im Sinne der Diagnostizität bewerten zu können, wurde vor allem von Trope und Kollegen untersucht (vgl. Trope & Bassok, 1983; Trope, Bassok & Alon, 1984; Trope & Mackie, 1987). Die Forschungsgruppe um Trope zeichnet dabei ein positives Bild vom Hypothesentester, da sie zeigen konnten, dass (explizit vorgegebene) Alternativhypothesen beachtet und diagnostische Fragen ausgewählt werden.

Denn die Probanden von Trope und Mackie (1987) passten ihre Fragensauswahl bzgl. des Berufs einer Zielperson an die verschiedenen Alternativhypothesen an, d.h. wenn zu überprüfen war, ob die Zielperson Professor ist und ein anderer lehrender Beruf die Alternative war, so wählten die Probanden forschungsbezogene Fragen, da diese im Gegensatz zu lehrbezogenen Fragen eine Unterscheidung ermöglichen würden. Das Verhältnis kehrte sich um, wenn ein anderer forschender Beruf die Alternative war. Analoge Resultate erzielten Kruglanski und Mayseless (1988, Experiment 1), die ihre Probanden nach Informationen suchen ließen, um zu entscheiden, ob eine Zielperson Architekt vs. Maler oder Architekt vs. Computerfachman ist.

In dieselbe Richtung weisen auch die Befunde von Skov und Sherman (1986), die ihre

Probanden auf eine Reise in den Weltraum schickten und dabei vor die Aufgabe stellten, zu entscheiden, zu welcher von zwei Spezies ein unsichtbares Weltallwesen gehört, dem sie dafür zwei Fragen stellen durften. Auch wurde ein klassisches Frage-Wahl-Paradigma umgesetzt, wobei konkret eine Liste mit zwölf Merkmalen und den zugehörigen bedingten Wahrscheinlichkeiten der Form $P(\text{Merkmal}_1|\text{Spezies}_1)$ und $P(\text{Merkmal}_1|\text{Spezies}_2)$ und vorgelegt. Wie zuvor dargelegt, kann aus diesen die Diagnostizität abgeleitet werden. Und in der Tat, die Probanden erkannten den Wert der Informationen und wählten häufiger Fragen, die hoch diagnostische Merkmale abfragten.

Aus diesen Befunden ist insgesamt ableitbar, dass Hypothesentester grundsätzlich in der Lage sind, diagnostisch vorzugehen, dies jedoch nicht in allen Situationen auch tatsächlich umsetzen. Ist keine explizite Alternativhypothese gegeben (Trope & Mackie, 1987) oder sind die zur Diagnostizitätsbestimmung benötigten Angaben nicht direkt vorgegeben (Doherty, Mynatt, Tweney & Schiavo, 1979), zeigt sich eine pseudodiagnostische Vorgehensweise. Ihre Probanden waren größtenteils nicht in der Lage, den Wert der bedingten Wahrscheinlichkeiten von sich aus zu erkennen und zu nutzen. Ihre Aufgabe war es, herauszufinden, von welcher Insel ein Tonartefakt stammte. Dafür konnten sie auswählen, welche Informationen über zwei Merkmale der Gefäße sie erhalten wollten. Es zeigte sich, dass die Probanden nicht die für die Bestimmung der Diagnostizität relevanten Angaben $P(\text{Merkmal}_1|\text{Insel}_1)$ und $P(\text{Merkmal}_1|\text{Insel}_2)$ suchten, sondern lieber Angaben zweier Merkmale einer Insel erfahren wollten. In diesem Fall gingen die Probanden dementsprechend pseudodiagnostisch vor, sie vernachlässigten die Alternativerklärung und ignorierten die Informationen, die zur Bestimmung der Diagnostizität der Merkmale nötig gewesen wären. Auch Kruglanski und Mayseless (1988, Experiment 2) konnten aufzeigen, dass die Rahmenbedingungen eine große Bedeutung für die tatsächliche Umsetzung eines diagnostischen Testens haben, denn Probanden, die von ihnen situativ unter Zeitdruck gesetzt wurden, stellten weniger diagnostische Fragen als Probanden denen eine Belohnung für die richtige Einschätzung in Aussicht gestellt wurde.

Zumindest ist jedoch unbestritten, dass ein diagnostisches Vorgehen dem pseudodiagnostischen unbedingt vorzuziehen ist, um Fehler zu vermeiden. Ob dies in Bezug auf das zweite Merkmal der Teststrategie (positiv vs. negativ) ebenso ist, wird zu diskutieren sein.

Befunde aus der Forschung zur klinisch-psychologischen Diagnostik. Da sämtliche Studien ihren Hauptfokus auf den zweiten Aspekt der Teststrategie legen, nämlich die Charakterisierung als positive vs. negative Teststrategie, werden die teststrategierelevanten Studien im nächsten Abschnitt referiert.

2.3.3.2 *Positive vs. negative Teststrategie*

Wie in Abschnitt 2.1.2 bereits erwähnt, ist für die Beantwortung der Frage, ob eine konkrete Teststrategie eine positive oder negative Ausrichtung hat, die Art der gesuchten Evidenz relevant. Als erste Annäherung an das Konzept sei daher formuliert, dass positives Testen das Suchen bestätigender bzw. konsistenter Informationen beinhaltet, weswegen Trope und Liberman (1996) auch von hypothesenkonsistentem Testen sprechen. Das negative Testen hingegen ist durch die Suche nach widerlegenden Informationen gekennzeichnet. Es folgen ausführliche Definitionen der beiden Ansätze.

Die positive Teststrategie wird von Klayman und Ha (1987, p. 212) wie folgt definiert: „... *in terms of a general positive test strategy, you test a hypothesis by examining instances in which the property or event is expected to occur (to see if it occurs), or by examining instances in which it is known to have occurred (to see if the hypothesized conditions prevail).*“ Folglich wird danach gefragt, ob das erwartete Merkmal unter den von der Hypothese vorhergesagten Bedingungen auftritt. In Abgrenzung zur positiven Teststrategie wird die *negative Teststrategie* von Oswald (in press) wie folgt definiert: „... *negative test strategy in which events which are expected not to occur are explicitly asked for...*“ Der Mieter im Mehrfamilienhaus würde eine negative Teststrategie verfolgen, wenn er danach fragen/suchen würde, ob der neue Nachbar eine Fernreise mit dem Flugzeug geplant hat, weil man Letzteres bei einer Person mit hoher ökologischer Verantwortlichkeit nicht vermuten würde. Der Psychologe würde im Sinne einer negativen Strategie das Vorliegen von substantiellem Untergewicht abklären, da dieses gegen eine Bulimie sprechen würde.

Im Rahmen der Forschung zur Verwendung der positiven und negativen Teststrategie hat sich wiederholt gezeigt, dass Probanden die positive Teststrategie der negativen vorziehen. Dieser Befund zeigt sich nicht nur in der Kognitionspsychologie (vgl. z.B. Wason, 1960; Wason & Johnson-Laird, 1972), sondern auch in der kognitiven Sozialpsychologie (vgl. z.B. Devine, Hirt & Gehrke, 1990; Evett et al., 1994; Hodgins & Zuckerman, 1993; Swann & Giuliano, 1987; Synder & Cantor, 1979).

Im Rahmen des Regelfindungsproblems des *2-4-6 tasks* der Kognitionspsychologie (Wason, 1960) wurde von den Probanden verlangt, die Regel zu entdecken, der das Zahlentripel 2-4-6 entspricht. Die zu findende Regel hinter diesem Tripel war *Reihe von ansteigenden Zahlen*. Die Probanden konnten für den Entdeckungsprozess weitere Zahlentripel nennen, über die ihnen mitgeteilt wurde, ob sie der gesuchten Regel entsprachen oder nicht. Sobald die Probanden sicher waren, die richtige Regel gefunden zu haben, sollten sie dem Versuchsleiter diese mitteilen. Nur 21 Prozent der Probanden

äußerten im ersten Versuch die korrekte Regel, so dass eine eingehendere Analyse der Hypothesen und Zahlentripel der Personen notwendig wurde, die mindestens eine falsche Annahme geäußert hatten. Diese zeigte an, dass die Probanden fast ausschließlich Zahlentripel nannten, die ihrer vermuteten Hypothese entsprachen. Wer beispielsweise die falsche Hypothese *Zahlen steigen um zwei an* im Kopf hatte, gab klassischerweise 4-6-8 als Testtripel an. Daraus schloss Wason, dass die Probanden dazu neigen, ihre Regel zu überprüfen, indem sie eine Bestätigung anstreben, anstatt die korrekte Regel über ein falsifikatorisches Vorgehen erschließen zu können, welches dadurch gekennzeichnet gewesen wäre, die Regel *Zahlen steigen um zwei an* durch ein Tripel von bspw. 1-2-3 zu prüfen, also einem Tripel, dass der vermuteten Regel nicht entspricht.

Ein analoges Vorgehen zeigte sich auch bei den Studien über das so genannte Vierkartenproblem (Wason & Johnson-Laird, 1972). In diesem Experiment wurden den Probanden vier Karten vorgelegt, die auf einer Seite eine Zahl und auf der anderen einen Buchstaben zeigten. Sichtbar waren die folgenden Symbole *E, K, 4, 7*. Die Probanden wurden nun vor die Aufgabe gestellt, durch das Umdrehen der notwendigen Karten folgende Regel zu überprüfen: *Wenn sich auf einer Seite einer Karte ein Vokal befindet, dann steht auf der anderen Seite eine gerade Zahl*. Es zeigte sich, dass am häufigsten die Karten mit dem Vokal (*E*) und der geraden Zahl (*4*) umgedreht wurden. Nach aussagenlogischer Betrachtung ist diese Lösung jedoch nicht korrekt. Die Regel entspricht der logischen Implikation *Wenn p, dann q*, d.h. *Wenn Vokal, dann gerade Zahl*, die nur dann falsifiziert ist, wenn gezeigt werden kann, dass eine Karte *p und nicht-q* (Vokal und ungerade Zahl) zeigt. Deshalb müssen die Karten für *p (Vokal = E)* und *nicht-q (ungerade Zahl = 7)* umgedreht werden. Die Karte mit *q (gerade Zahl = 4)* ist irrelevant, da die Regel hierfür keine Vorhersage macht. Nur vier Prozent der Probanden wählten die richtige Kombination der Karten *E* und *7* aus. Wason nimmt wiederum an, dass das Bestreben, die Regel zu bestätigen statt zu falsifizieren, zu der schlechten Leistung der Probanden geführt hat. An dieser Stelle soll aber nicht verschwiegen werden, dass diese Interpretation kontrovers diskutiert wurde, da bspw. von Griggs und Cox (1982) ein hohes Maß an Kontextabhängigkeit der Leistung gezeigt werden konnte. Stellten sie die Probanden vor die Aufgabe, die Regel zu prüfen *Wenn jemand Bier trinkt, dann muss er über 18 Jahre alt sein*, wählten mehr Personen die richtigen Prüfkarten. Eine Alternativerklärung zum konfirmatorischen Testen ist demnach nicht ausgeschlossen.

Des Weiteren konnte die Präferenz für die positive Teststrategie ebenso im sozialpsychologischen Forschungskontext des Testens sozialer Hypothesen über Persönlichkeitsmerkmale nachgewiesen werden (vgl. z.B. Devine et al., 1990; Evett et al., 1994; Hodgins & Zuckerman, 1993; Swann & Giuliano, 1987; Snyder & Cantor, 1979).

Exemplarisch sei die Arbeit von Snyder und Swann (1978) erneut erwähnt, in der Probanden, instruiert herauszufinden, ob eine unbekannte Person extravertiert/introvertiert sei, Fragen auswählen, die im Sinne der positiven Teststrategie die Merkmale der jeweiligen Hypothese abfragten. Im Rahmen dieses vielfach verwendeten Untersuchungsparadigmas wurde den Probanden eine Liste mit 26 Fragen vorgelegt, aus denen sie zwölf Exemplare auswählen mussten. Diese Fragen beschrieben 21 Kennzeichen von entweder typisch introvertierten oder extravertierten Personen, fünf waren irrelevant. Die Mehrheit der Probanden wählte solche Fragen aus, die Merkmale abfragten, die mit ihrer Hypothese korrespondierten – sie verfolgten also eine positive Teststrategie. Auch diese Studie ist vielfach kritisiert worden (vgl. u. a. Bassok & Trope, 1984; Effler, 1986; Gadenne, 1982; Klayman & Ha, 1987; Semin & Strack, 1980; vom Schemm, 2008a). Zentral für die vorliegende Arbeit ist jedoch ausschließlich der zu kritisierende Umstand, dass die Autoren das hypothesenkonsistente Testen mit dem *confirmation bias* gleichsetzen.

Ist die positive Teststrategie eine konfirmatorische Teststrategie? In den meisten Diskussionen der frühen Arbeiten wird einheitlich gefolgert, dass die positive Teststrategie für den *confirmation bias* verantwortlich zu machen ist. Die positive Teststrategie wird im Einklang damit auch häufig als konfirmatorische Strategie bezeichnet. Im Folgenden soll diskutiert werden, ob die positive Teststrategie tatsächlich eine konfirmatorische Strategie in dem Sinne ist, dass sie die Bestätigung der Hypothese systematisch begünstigt. Gegenargumente liefern dafür einerseits Trope und Liberman (1996), indem sie darlegen, dass die positive Teststrategie nicht mit den Konsequenzen des Testungsprozesses zusammenhängt, und andererseits Klayman und Ha (1987), die zeigen, dass eine positive Teststrategie auch zur Falsifikation genutzt werden kann. Abschließend wird darauf eingegangen, wie die Bedeutung der positiven Teststrategie im Rahmen dieser Arbeit eingeschätzt wird.

Trope und Liberman (1996) argumentieren, dass hypothesenkonsistente (synonym für positive) Fragen nicht grundsätzlich für die (systematische) fälschliche Bestätigung von Hypothesen verantwortlich gemacht werden können. Sie beschrieben vielmehr, dass sie sich im Hinblick auf die Konsequenzen für die Validität der Schlussfolgerung nicht von hypotheseninkonsistenten (negativen) unterscheiden. Die Validität der Entscheidung einer Hypothesenprüfung ist nach ihrer Auffassung vielmehr von drei Faktoren abhängig: der Diagnostizität der Frageinhalte, den Antworten auf die Fragen und der Frage, ob diagnostisch oder pseudodiagnostisch beim Schlussfolgern aus den Antworten vorgegangen wird.

Sie veranschaulichen die Konsequenzen eines Hypothesentestungsprozesses daher wie in Abbildung 3 dargestellt. Es fällt auf, dass v.a. das diagnostische vs. pseudodiagnostische Schlussfolgern im Sinne der Beachtung von Alternativhypothesen über die Korrektheit der Entscheidung bestimmt und es selbst im Falle nicht diagnostischer Fragen nicht zwangsweise zu einer falschen Bestätigung kommen muss.

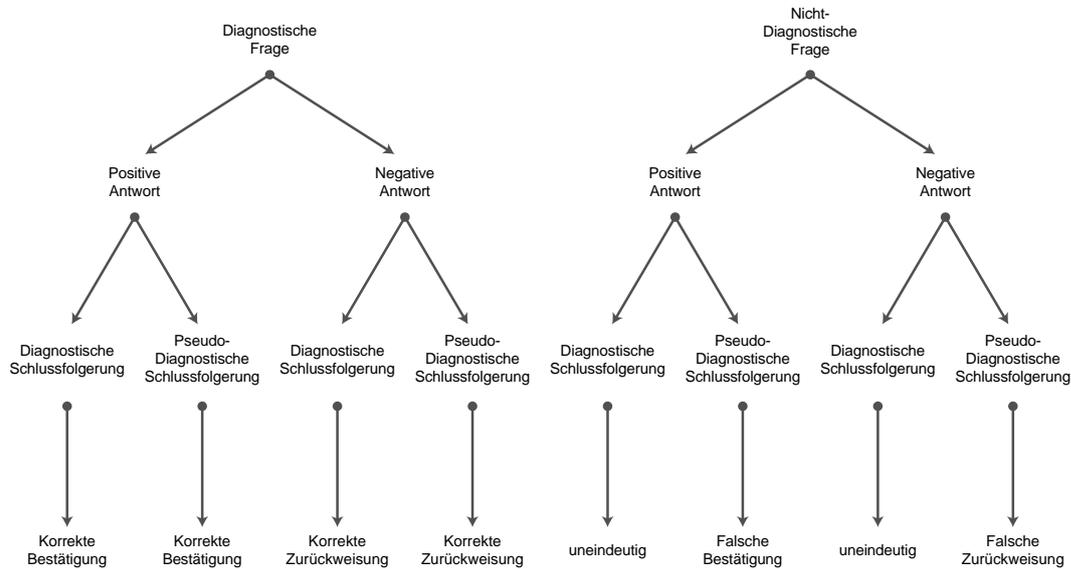


Abbildung 3. Konsequenzen eines Hypothesentestungsprozesses (angelehnt an Trope & Liberman (1996, p. 245))

Der Abbildung ist auch zu entnehmen, dass das Merkmal der Positivität vs. Negativität keinen Zusammenhang mit der Validität der Entscheidung aufweist – es ist konsequenter Weise nicht dargestellt. Insofern können Fragen im Sinne der positiven Teststrategie nicht grundsätzlich als selbstbestätigend verstanden werden. Diese Position steht im Einklang mit Oswald und Grosjean (2004), die in ihrer Definition des *confirmation bias* einen systematischen Einfluss hinsichtlich der Verringerung der Zurückweisungswahrscheinlichkeit erwarten. Ein Vorgehen im Rahmen der positiven Teststrategie führt dann zur Bestätigung, wenn das untersuchte Merkmal auch tatsächlich vorliegt. Eine Frage, formuliert im Sinne der positiven Teststrategie, würde einen Beleg für die Hypothese liefern, wenn der Befragte sie bejaht. Bei einer Verneinung liefert die Antwort widersprüchliche Evidenz. Eine Bestätigung ist somit nicht zwingend und ein (systematisch) konfirmatorischer Einfluss hiermit widerlegt.

Die Meinung, die positive Teststrategie sei ein konfirmatorisches Vorgehen, basiert auf der Annahme, die positive Strategie entspräche der generellen Verifikationsstrategie, im Rahmen derer Hypothesen getestet werden, indem versucht wird, sie zu belegen.

Analog wird das negative Vorgehen mit der Falsifikationsstrategie im Zusammenhang gesehen, nach der eine Prüfung durch den Versuch der Widerlegung vorgenommen wird. In der Psychologie wird jedoch vornehmlich der Position Poppers (Popper, 1959) folgend die Falsifikationsstrategie als Goldstandard angesehen, während die Verifikationsstrategie als weniger geeignet für den Erkenntnisgewinn erachtet wird. Popper argumentiert, dass die Aneinanderreihung von Hypothesenbestätigungen niemals zur Sicherheit über das Zutreffen der Annahme führen könne. Nur aus dem Ausschluss von falschen Hypothesen könnten Erkenntnisse gewonnen werden, so dass die Falsifikationsstrategie das Mittel der Wahl zur Überprüfung von Hypothesen sei. In diesem Lichte scheint es irrational und fehlerhaft, eine Strategie zu verfolgen, die zur Bestätigung der zu prüfenden Annahmen dient. Die positive Teststrategie wäre nun genau dann konfirmatorisch, wenn man mit ihr lediglich eine Bestätigung nicht aber die Falsifikation der Hypothese bewerkstelligen könnte. Dies ist allerdings mitnichten der Fall. Der Glaube, mit der positiven Teststrategie könne lediglich eine Bestätigung und keine Falsifikation einer Hypothese erlangt werden, ist nicht korrekt. Klayman und Ha (1987) zeigen in ihrer Analyse des *2-4-6 rule discovery tasks* von Wason (1960), dass durch die Wahl der Regel *ansteigende Zahlen* eine sehr breite Kategorie gewählt wurde, die viele Tripel von differenzierteren Regeln umfasst. In dieser Situation ist in der Tat keine Falsifikation möglich. Wäre die richtige Regel hingegen *aufeinanderfolgende gerade Zahlen* gewesen und hätte der Proband spontan angenommen, dass die Regel *drei gerade Zahlen* ist, hätte ein positiv konstruiertes Tripel (z.B. 2-8-14) zu einer Falsifikation geführt, weil der Versuchsleiter als Feedback eine Verletzung der echten Regeln *aufeinanderfolgende gerade Zahlen* angezeigt hätte. Es kann folglich durchaus Situationen geben, in der mit Hilfe der positiven Teststrategie eine Falsifikation erreicht werden kann. Die Wahl einer positiven bzw. negativen Teststrategie korrespondiert dementsprechend nicht notwendigerweise mit der Wahl eines verifizierenden vs. falsifizierenden Vorgehens. Der Erkenntnisgewinn bei positivem Testen muss nicht zwingend geringer sein.

Im Einklang mit Klayman und Ha (1987) wird die positive Teststrategie im Rahmen dieser Arbeit daher als Heuristik verstanden, die lediglich im Zusammenwirken mit anderen Faktoren einen konfirmatorischen Einfluss ausüben kann (vgl. Oswald & Grosjean, 2004). Die Autoren führen drei Konstellationen an: erstens die Situation, in der die korrekte Hypothese breiter ist als die vom Test angenommene (vgl. Diskussion von Klayman und Ha zum *rule discovery task*), zweitens eine derartige Gestaltung der Interviewsituation, die den Interaktionspartner zum affirmativen Antworten verleitet und drittens die Kombination mit einem pseudodiagnostischen Vorgehen.

Befunde aus der Forschung zur klinisch-psychologischen Diagnostik. Der Aspekt der Informationsgewinnung ist im Zusammenhang mit der klinisch-psychologischen Diagnostik v.a. im Sinne der Generierung von Fragen untersucht worden. Diverse Studien, deren Ausgangspunkt in der Arbeit von Snyder und Swann (1978) liegt und i.d.R. eine Übertragung auf den klinisch-psychologischen Kontext beinhaltet, zeichnen dabei ein durchweg positives Bild des Diagnostikers, der in diesem Sinne keineswegs hypothesenbestätigend agiere. Letzteres soll nicht unwidersprochen bleiben, aber zunächst werden die Argumente gegen das Auftreten des *confirmation bias* in der klinisch-psychologischen Diagnostik vorgebracht.

Zunächst replizierten Strohmer und Newman (1983) die Studie von Snyder und Swann (1978) an einer Gruppe von Studierenden, ergänzten die Liste der Auswahlfragen allerdings um selbstentwickelte, so genannte *unbiased* Fragen, die wie z.B. *Would you describe yourself as outgoing or shy?* mit gleicher Wahrscheinlichkeit konsistente wie inkonsistente Antworten hervorrufen können. Sie begegneten somit einem gewichtigen Kritikpunkt an der Originalarbeit und fanden, dass ca. 70% ihrer Probanden Fragen im Sinne der *unbiased strategy* bevorzugten. In einer zweiten Studie übertrugen sie die Situation auf einen Beratungskontext und ließen 40 *advanced doctoral students* ein Interview mit je einem *undergraduate* führen und dabei entweder die Hypothese mangelnder oder ausreichender Selbstkontrolle prüfen. Die Fragen wurden von Ratern hinsichtlich der Qualität als *self-controlled vs. uncontrolled vs. unbiased vs. irrelevant* eingeschätzt. Dass die Art der Hypothese keinen Einfluss auf die Häufigkeit der Verwendung der Fragen aus den vier verschiedenen Klassen hatte, werteten die Autoren als weiteren Beleg für die Präferenz einer unvoreingenommenen Strategie.

In dieselbe Richtung scheint der Befund von Wilson Dallas und Baron (1985) zu weisen, die einen weiteren Kritikpunkt an Snyder und Swann (1978) aufgreifend zusätzlich zu einer Replikationsbedingung mit Psychotherapeuten eine Bedingung realisierten, in der nicht zwölf Fragen aus der Liste zu wählen, sondern selbst zu generieren waren, die später von zwei Ratern in die Kategorien *extroverted vs. introverted vs. neutral* eingeteilt wurden. In der realitätsnäheren Bedingung der selbstgenerierten Fragen zeigte sich nun keinerlei Präferenz für die jeweils hypothesenkonsistenten Fragen. Einen etwas schalen Beigeschmack hinterlässt allerdings der Befund, dass in der Bedingung der Fragenauswahl auch bei den Psychotherapeuten eine Präferenz für die hypothesenkonsistenten Fragen gefunden werden konnte, obwohl er wegen der geringeren Realitätsnähe als weniger bedeutsam eingestuft werden muss.

Das positive Bild des unvoreingenommenen Hypothesentesters hält außerdem den weiteren Versuchen von Strohmeyer und Chiodo (1984) stand, den Fragegenerierungsprozess unter stärker verzerrenden Einflüssen zu untersuchen. Dafür überprüften sie bspw., ob Personen, die aufgefordert werden, selbst eine Hypothese über einen zuvor geschilderten Fall zu formulieren und mithilfe von selbst generierten Fragen zu testen, weiterhin unvoreingenommen vorgehen, obwohl sie mehr in die Hypothese involviert sind als Personen, denen diese vom Versuchsleiter vorgegeben wurde (Experiment 1). Auch in dieser Situation wurden am häufigsten unvoreingenommene Fragen generiert und zudem zeigten sich keine Mittelwertunterschiede in der Nutzung der verschiedenen Kategorien. Es muss allerdings festgehalten werden, dass nicht dokumentiert wurde, ob in irgendeiner Weise abgefragt wurde, welche Hypothese von den jeweiligen Probanden überhaupt verfolgt wurde oder ob es überhaupt einen Unterschied in dem *involvement* mit der zu testenden bzw. entwickelten Hypothese gab. Ein fehlender Unterschied in der Präferenz für Frageformulierungen könnte demnach auch ein methodisches Artefakt sein, da die Gruppen sich nicht hinreichend unterschieden.

Desweiteren liefern die Ergebnisse von Hayden (1987) erneut einen Hinweis auf die Stabilität der Präferenz für eine unvoreingenommene Teststrategie, indem zusätzlich darauf geachtet wird, die Präferenz für Frageformen über den Verlauf eines (vermeintlichen) Gesprächs zu untersuchen und auch die Reaktionen auf die Fragen in den Prozess mit einzubeziehen. Hierfür nutzt Hayden in seinem ersten Experiment eine Computersimulation – vergleichbar mit der Operationalisierung der Frageauswahl von Dreger (2006) - und benutzte hierfür die Materialien von Snyder und Swann (1978) bzw. die modifizierten von Strohmeyer und Newman (1983). Die Präferenz für die *unbiased* Teststrategie im Sinne von Strohmeyer und Newmann (1983) wurde bestätigt. Außerdem konnte durch die kontinuierliche Aufzeichnung gezeigt werden, dass die Teststrategie zeitlich über das Interview stabil ist.

Insgesamt zeigen die Befunde daher ein (scheinbar) erfreuliches Bild des unvoreingenommenen hypothesentestenden Diagnostikers. Strohmeyer et al. (1990) resümieren den Forschungsstand daher wie folgt: „*when counselors develop a client hypothesis and ask questions in order to test the hypothesis, counselors are not subject to the tendency to confirm their hypotheses*“ (p. 466). Allerdings ist an dem Studiendesign in Anlehnung an das Extraversions-Introversions-Paradigma trotz seiner Erweiterungen durch die oben genannten Autoren einiges in Frage zu stellen. Auch Haverkamp (1993) ist der Überzeugung, dass die Forschungsfrage gerade für den diagnostischen Kontext noch nicht hinreichend geklärt ist, und führt methodische Kritikpunkte auf drei Ebenen an.

Erstens kritisiert sie die Operationalisierung der Teststrategie anhand des Paradigmas von Snyder und Swann (1978), da mehrere Besonderheiten des diagnostischen Kontextes unberücksichtigt bleiben, wenn der Einsatz einer konfirmatorischen Teststrategie durch die Anzahl der ausgewählten, vorab kategorisierten Fragen zu einer vorgegebenen Hypothese operationalisiert wird. Einerseits wird durch die *ausschließliche Fokussierung auf die Fragen* des Diagnostikers außer Acht gelassen, dass seine Informationssuche beispielsweise auch para- und nonverbale Hinweise beinhalten wird. Die bisherigen Studien bilden den Suchprozess daher nur eindimensional ab. Und durch die besonderen Sozialisationsbedingungen während der psychologischen Ausbildung könnte andererseits die Tendenz, hauptsächlich offene Fragen zu stellen (offene Fragen werden als Einsatz einer neutralen Strategie gewertet), weniger ein unmittelbarer Ausdruck der Unvoreingenommenheit sein, sondern vielmehr die *Demonstration eines gelernten Zielverhaltens* darstellen, denn das Stellen von offenen Fragen wird (zu Recht) als Goldstandard vermittelt. Desweiteren kritisiert die Autorin die *Vernachlässigung des Frageinhalts* zugunsten der Frageform. Dies ist problematisch, denn eine verzerrte Informationssuche kann sich auch durch die Begrenzung auf bestimmte Inhaltsbereiche ausdrücken, z.B. nur Abfrage von familiären Zusammenhängen bei Außerachtlassung des schulischen Kontexts, wenn die Hypothese getestet wird, dass eine Kindesmisshandlung für den Leistungsabfall in der Schule verantwortlich ist. Darüber hinaus weist Haverkamp (1993) auch auf eine *Konfundierung von Frageform und Informationsgehalt* hin. Denn während die (dis-)konfirmatorisch formulierte Frage nur die einzelne Hypothese anspricht und damit den Probanden wahrscheinlich weniger Informationen produzieren lässt, als bei den neutralen Fragen (z.B. *I would first like to know if you enjoy talking with people or enjoy being quiet around people*; Hayden, 1987, p. 151), die meist Informationen über sowohl die Test- als auch Alternativhypothese liefern werden. Haverkamp stellt nun in Frage, ob die Probanden tatsächlich einer unvoreingenommenen Strategie folgten oder vielmehr einer Strategie, die beinhaltet, die meisten Informationen zu produzieren.

Darüber hinaus ist Haverkamp (1993) skeptisch, dass die Instruktionen der Probanden in den zugrundeliegenden Arbeiten geeignet waren, einen realitätsgetreuen Gedankenprozess der Diagnostiker auszulösen, da vielfach die Begriffe *test*, *confirm* oder *disconfirm* verwendet wurden. Daher ist die Alternativhypothese salienter geworden, was einerseits vom realistischen Setting abweicht und damit ein untypisches Denkmuster produziert und andererseits genau eine der Rahmenbedingungen ist, die diagnostisches Vorgehen begünstigen. Zwar wäre dies zu begrüßen, aber die Verfasserin schließt sich Haverkamps (1993) Kritik an, dass aufgrund dieser verringerten externen Validität die Aussagekraft des Befunds des unvoreingenommenen Diagnostikers in Frage zu stellen und weiter zu untersuchen ist.

Ein dritter Kritikpunkt Haverkamps (1993) betrifft die *Quelle der zu testenden Hypothese*, die ebenfalls mit einem Mangel an Übereinstimmung mit einem naturalistischen Setting assoziiert ist, wenn diese vom Versuchsleiter vorgegeben wird (*experimenter provided*) statt vom Diagnostiker selbst generiert wird (*self generated*). Da in den bisherigen Studien nicht abgefragt wurde, welche Hypothese überhaupt verfolgt wird, sondern stets unterstellt wurde, die vorgegebene Hypothese werde getestet, kann streng genommen nicht ausgeschlossen werden, dass die ausgewählten (dis-/konfirmatorischen) Fragen eigentlich eine andere Strategie verfolgen, als ausgewertet.

In ihrer eigenen Studie (Haverkamp, 1993), bei der sie *counseling trainees* ein 15 minütiges Video einer Klient-Berater-Interaktion präsentierte, verschleierte sie daher den Hintergrund der Studie, indem sie vorgab, zu untersuchen, wie der Klient seinem Berater unter der Nutzung von verbalen und nonverbalen Kanälen Hinweise auf sein Problem gibt. Darüber hinaus gab es eine fünfteilige Messung des Testungsverhaltens, das sowohl für die *experimenter provided* als auch die *self generated hypothesis* durch unabhängige Rater ausgewertet wurde: 1.) das identifizierte Problem, 2.) eine Auflistung von relevant erachteten verbalen und nonverbalen Hinweisen, 3.) fünf Adjektive zur Beschreibung für den Vorgesetzten, 4.) fünf Fragen zur weiteren Einkreisung des Problems und 5.) eine Hausaufgabe für den Klienten.

Ihre Analyse der Antworten ergab, dass unter der Betrachtung der *self generated hypothesis* 64% als konfirmatorisch, d.h. konsistent mit dem selbst identifizierten Problem zu werten waren. 21% waren neutral, die restlichen 15% diskonfirmatorisch. Das konfirmatorische Vorgehen beim Testen einer selbst generierten Hypothese war daher robust auf allen Maßen und zeigt ein entgegengesetztes Bild zu den bisherigen Forschungsergebnissen (Wilson Dallas & Baron, 1985; Hayden, 1987; Strohmer & Chiodo, 1984; Strohmer & Newman, 1983). Die Ergebnisse unter der Betrachtung der *experimenter provided hypothesis* sollen an dieser Stelle nicht weiter kommentiert werden, weil das unerwartete Ergebnis eines in sich inkonsistenten Ergebnismusters nach Angabe von Haverkamp (1993) mit einem Materialeffekt zusammen zu hängen scheint.

Es ist daher festzuhalten, dass die vorliegende Untersuchung einen weiteren Forschungsbedarf angezeigt hat, der eine Öffnung des Untersuchungsparadigmas über die Fra-genauswahl hinaus notwendig macht. Kritisch bzgl. der Arbeit von Haverkamp (1993) ist jedoch anzumerken, dass auch sie im Einklang mit Snyder und Swann (1978) das hypothesenkonsistente Testen (im Sinne der positiven Teststrategie) als *confirmation bias* konzeptualisiert. Wie oben argumentiert, handelt es sich dabei jedoch eher um eine Heuristik, die lediglich in Kombination mit einem pseudodiagnostischen Vorgehen als konfirmatorisch zu beschreiben ist. Dieses gilt es als Nächstes zu veranschaulichen.

2.3.3.3 *Kombination von positiver und pseudodiagnostischer Teststrategie*

Bei dieser Kombination kumulieren sich die folgenden Umstände: es wird erstens bevorzugt nach Merkmalen gesucht, die die Prüfhypothese vorhersagt (positive Teststrategie), zweitens bleibt die Diagnostizität der Evidenz unbeachtet und drittens wird die Reliabilität der Evidenz nur unzureichend mit einbezogen. Wenn im Rahmen der Informationssuche Fragen bejaht werden, die diese Charakteristika aufweisen, kann es zur (dreifachen) Selbstbestätigung der Hypothese kommen.

Zur Veranschaulichung sei ein Beispiel des Mieters im Mehrfamilienhaus herangezogen. Der Mieter befrage dazu Frau Krause, eine Nachbarin, die von ihrer Wohnung aus auf die Mülltonnen gucken kann. Er gehe dabei davon von der Evidenzregel aus *Wenn mein Nachbar die Grünen wählt, dann sollte er den Müll trennen*. Er stelle dazu die Frage *Trennt der neue Nachbar den Müll korrekt?* Die Nachbarin antworte mit *Ja*. Die Bejahung der Frage, die eine Frage im Sinne der positiven Teststrategie darstellt, liefert dem Mieter dann einen ersten Beleg für die Grünen-Hypothese. Würde darüber hinaus nicht in Betracht gezogen werden, dass auch Wähler anderer Parteien korrekt den Müll trennen und man nicht annehmen muss, dass sich die Wahrscheinlichkeit, Müll zu trennen bei Grünenwählern und Wählern anderer Parteien unterscheidet (pseudodiagnostisches Vorgehen), erhärtete die Antwort den Verdacht weiter. Dieser Effekt würde unter Umständen noch verstärkt werden, wenn der Mieter sich keine Gedanken darüber machen würde, dass die befragte Nachbarin schon über 80 Jahre alt ist und stark sehbehindert ist (Vernachlässigung der Reliabilität der Evidenz). Dieses Beispiel macht deutlich, dass die Kombination der Strategien eine große Gefahr zur fälschlichen Selbstbestätigung einer Hypothese birgt.

Um das Risiko für Selbstbestätigungen abschätzen zu können, stellt sich daher die Frage, wie wahrscheinlich es ist, dass die beiden Merkmale der Teststrategie so ungünstig kombiniert werden. In diesem Zusammenhang interessiert auch die Frage, ob Personen eigentlich eine Präferenz des diagnostischen über das positive Testen (oder umgekehrt) haben. Es wurden daher etliche Studien über die relative Bedeutung der beiden Merkmale der Teststrategie durchgeführt (vgl. z.B. Devine et al., 1990; Skov & Sherman, 1986; Trope & Bassok, 1983; Trope et al., 1984).

Konsistent hat sich gezeigt, dass die Probanden primär eine Präferenz für die diagnostische Teststrategie aufweisen, wenn man sie vor die Wahl stellt, hoch diagnostisch und negativ zu suchen oder nicht diagnostisch und positiv zu fragen (Skov & Sherman, 1986). Dieser Befund scheint zunächst erleichternd, weil er anzeigt, dass die kritische

Kombination erkannt und seltener genutzt wird als die günstige. Sekundär zeigte sich jedoch auch eine klare Präferenz für die positive gegenüber der negativen Strategie, wenn die Diagnostizität kontrolliert wurde (z.B. vgl. Devine et al., 1990 im Paradigma der Testung von Persönlichkeitsmerkmalen). Das heißt insgesamt, dass am ehesten solche Fragen gewählt werden, die hoch diagnostisch und positiv sind. Wie oben argumentiert ist diese Kombination an sich noch nicht kritisch, allerdings ist die Aufrechterhaltung eines diagnostischen Vorgehens über den gesamten Ablauf der Hypothesentestung kognitiv und motivational sehr herausfordernd und so führen Trope und Liberman (1996) aus, dass es durchaus vorkommen kann, dass ein diagnostisch begonnenes Vorgehen mit differenziert abgeleiteten Alternativhypothesen bei der Informationssuche in ein pseudodiagnostisches Vorgehen umschlägt. Die Fragilität des diagnostischen Vorgehens wurde zudem bereits aufgrund der oben dargestellten Befunde zur Sensitivität für Alternativhypothesen und Diagnostizität angedeutet, da die Probanden ohne explizite Alternativhypothese (vgl. Trope & Mackie, 1987) oder ohne explizite Vorgabe der zur Diagnostizitätsbestimmung benötigten Angaben (vgl. Doherty et al., 1979), eine pseudodiagnostische Vorgehensweise zeigten. Da man zudem in einer realen Situation von deutlich höheren Anforderungen als in einer kontrollierten Laborstudie zur Entscheidung der Spezieszugehörigkeit von Weltraumwesen ausgehen muss, scheint die Fragilität der tatsächlichen Umsetzung der diagnostischen Strategie bei einer recht robusten Vorliebe für positives Testen ein hinreichenden Nährboden für den Einsatz der Kombination einer pseudodiagnostischen und positiven Strategie zu liefern und die anfänglich aus rationaler Sicht unwahrscheinliche Überlegung, Menschen könnten tatsächlich diese ungeeignete Kombination wählen, wird nun nachvollziehbar.

2.3.4 Motivationale Prozesse

Zum Verständnis dieses Verzerrungsmechanismus ist es zunächst erforderlich darzulegen, warum Menschen nach Ansicht der Autoren überhaupt das kognitiv und zeitlich aufwändige Hypothesentesten auf sich nehmen anstatt die Richtigkeit ihrer Hypothese gleich von vornherein anzunehmen. Nach Trope und Liberman (1996) liegt die Motivation für das Hypothesentesten in dem Bedürfnis, Fehler zu vermeiden bzw. die Wahrscheinlichkeit von Fehlern zu minimieren, um damit Konsequenzen bzw. Kosten zu vermeiden, die aus diesen resultieren könnten. Tabelle 1 (siehe folgende Seite) veranschaulicht die richtigen und falschen Entscheidungen bei einem Hypothesentestungsprozess am Beispiel der Bulimie-Diagnostik.

Tabelle 1. Veranschaulichung der Validität eines Urteils in Abhängigkeit des wahren Zustandes.

		<i>In Wahrheit: Bulimie...</i>	
		<i>liegt vor</i>	<i>liegt nicht vor</i>
<i>Der Psychologe befindet: Bulimie...</i>	<i>liegt vor</i>	richtige Annahme	falsche Annahme
	<i>liegt nicht vor</i>	falsche Zurückweisung	richtige Zurückweisung

Eine Schlussfolgerung kann dementsprechend auf zwei Arten korrekt sein: eine richtige Annahme (richtig positiv) und eine richtige Zurückweisung (richtig negativ) der Hypothese. Die Schlussfolgerung kann jedoch auch auf zwei Arten falsch sein: die falsche Annahme (falsch positiv) und die falsche Zurückweisung (falsch negativ) einer Hypothese. Die Kosten einer fälschlichen Annahme, die Patientin leide unter einer Bulimie (vs. keine Erkrankung) führt zu unnötigen Behandlungen, die neben der finanziellen Belastung des Gesundheitssystems zu einer (negativen) Etikettierung der Patientin führen könnte. Die Kosten bei fälschlicher Zurückweisung der Hypothese, die Patientin leide unter Bulimie, wären die physische und psychische Belastung, wenn sie unbehandelt bliebe. Insofern spielen sowohl die persönlichen Kosten des Hypothesentesters als auch die Folgen für die ebenfalls betroffenen Personen eine Rolle für die Motivation.

Trope und Liberman (1996) nehmen nun an, dass sich der Tester zwei Schwellenwerte setzt: eine Schwelle für die Akzeptanz und eine für die Zurückweisung der Hypothese. Die Schwelle für die Akzeptanz beschreibt die minimale Überzeugung die notwendig ist, bevor die Hypothese akzeptiert wird. Die Höhe der Schwelle hängt dabei von der Höhe der antizipierten Kosten ab. Je größer die vermuteten Kosten sind, desto höher wird der Schwellenwert ausfallen. Werden die Kosten einer falschen Annahme anders bewertet als die Kosten einer falschen Zurückweisung, fallen auch die Schwellenwerte unterschiedlich aus. Wählt der Tester nun eine niedrige Schwelle für die Annahme der Hypothese und eine höhere für deren Zurückweisung, weil er die Kosten für eine fälschliche Zurückweisung höher beurteilt als die Folgen einer fälschlichen Annahme der Hypothese, so reichen ihm zur Annahme der Hypothese weniger Informationen als zu deren Zurückweisung. Dies stellt eine weitere potentielle Begünstigung der zu testenden Hypothese und damit einen Verzerrungsmechanismus dar.

In einer Studie im klassischen Paradigma zur Untersuchung von Falschbeschuldigen wegen sexuellen Kindesmissbrauchs von Carstens (2001) wurden die Schwellen für die Annahme bzw. Ablehnung der Missbrauchshypothese manipuliert, indem eine Hälfte der

Probanden einen Bericht las, der die Konsequenzen von sexuellem Kindesmissbrauch salient machte. Die andere Hälfte las diesen Bericht nicht. Orthogonal dazu erhielt eine Hälfte der Probanden einen Bericht über die Konsequenzen von Falschanklagen, die andere Hälfte nicht. Hatten die Probanden ausschließlich den Falschanklagebericht gelesen, so gelangten sie zu signifikant weniger Falschbeschuldigungen (nun ca. 19% statt 27% in den anderen drei Bedingungen). Daraus lässt sich schlussfolgern, dass die Salienz der Konsequenzen für eine falsche Annahme der (Missbrauchs-)Hypothese die Schwelle für die Annahme abgesenkt hat und in diesem Zusammenhang weniger Fehlurteile resultiert sind. Der Einfluss der salient gemachten Konsequenzen auf die Schwellenwahl und Entscheidung konnte somit empirisch gezeigt werden.

Nachdem sich die bisherige Vorstellung der sieben potentiellen Verzerrungsmechanismen nach dem Modell des konfirmatorischen Hypothesentestens von Schulz-Hardt und Köhnken (2000) eng an den Standpunkten der Autoren orientiert hat, soll abschließend der Blick für Ergänzungsmöglichkeiten geweitet werden.

2.4 Ergänzungsmöglichkeiten

Grundsätzlich sind verschiedene Ergänzungsmöglichkeiten in Bezug auf das Modell des konfirmatorischen Hypothesentestens nach Schulz-Hardt und Köhnken (2000) denkbar und die folgende angerissene Diskussion erhebt keinen Anspruch auf Vollständigkeit. Vielmehr werden die potentiellen Ergänzungsmöglichkeiten vorgestellt, zu denen die Verfasserin im Rahmen ihrer Beschäftigung mit dem konfirmatorischen Hypothesen eigene Vorarbeiten umgesetzt hat. Die vorzustellenden Optionen lassen sich danach gliedern, ob die Faktoren innerhalb der Person des Hypothesentesters liegen oder durch die Art der Hypothese bedingt sind.

2.4.1 Faktoren innerhalb der Person des Hypothesentesters

An die Person des Hypothesentesters bzw. an die interindividuellen Unterschiede von Hypothesentestern geknüpft sind die Ausführungen zum Einfluss von Persönlichkeitsmerkmalen, dem intern-repräsentierten Anfangsverdacht sowie der Expertise des Hypothesentesters.

2.4.1.1 Persönlichkeitsmerkmale

Vom Schemm (2008a) hat sich näher mit Erweiterungsmöglichkeiten für das Modell des konfirmatorischen Hypothesentestens beschäftigt und thematisiert hier v.a. Persönlichkeitsvariablen, da ihrer Ansicht nach in allen Stufen des Modells Unterschiede zwischen Personen wirksam werden könnten. Sie setzt sich daher näher mit dem Einfluss der Merkmale Ambiguitätsintoleranz, *need for cognitive closure* und Kontrollüberzeugungen auseinander.

Der inhaltlichen Nähe der beiden Konstrukte geschuldet, fasst sie für Ambiguitätsintoleranz und *need for cognitive closure* den Forschungsstand dahingehend zusammen, dass die jeweilige Ausprägung davon dafür mitverantwortlich gemacht werden kann, wie Menschen mit mehrdeutigen Reizen oder uneindeutigen Situationen umgehen, wobei jeweils eine hohe Merkmalsausprägung mit einer Tendenz zu bestätigenden Fehlern in Verbindung gebracht werden kann.

Kontrollüberzeugungen werden von vom Schemm (2008a) in Anlehnung an Amelang, Bartussek, Stemmler und Hagemann (2006) als generalisierte Erwartung hinsichtlich der Instanz verstanden, die für die Konsequenzen bzw. Verstärkung des eigenen Verhaltens verantwortlich gemacht werden. Die Autorin hält die Kontrollüberzeugungen im Zusammenhang mit der Informationssuche für relevant, da Personen, die sich selbst für verantwortlich halten (internale Kontrollüberzeugung) gezielter nach Lösungen von Problemen und Aufgaben suchen und diese besser nutzen können (vgl. Pines & Julian, 1972).

In einer eigenen Vorarbeit (vgl. auch Jochumsen, 2009) wurde der Einfluss von Persönlichkeitsmerkmalen auf die Ausbildung eines Verdachts (hier Schuld vs. Unschuld bzgl. einer schweren Brandstiftung mit Todesfolge) im Paradigma des konfirmatorischen Hypothesentestens direkt untersucht. Neben der Berücksichtigung der Merkmale, die vom Schemm (2008a) vorschlägt, wurden zudem kognitive Komplexität (erfasst mit Hilfe der deutschen Version des *object sorting tests* von Scott (1969), *need for cognition* (verkürzte deutsche Version der Need for Cognition-Skala von Bless, Wänke, Bohner, Fellhauer & Schwarz, 1994), Autoritarismus (AUS, Autoritarismusskala; Dalbert, 1996), Offenheit für Erfahrungen (NEO-FFI; Borkenau & Ostendorf, 2008), Konfliktbereitschaft (Eigenentwicklung) und Empathie (E-Skala, Fragebogen zur Erfassung von Empathie; Leibetseder, Laireiter, Riepler & Köller, 2001) untersucht. Den einzigen überzeugenden Zusammenhang mit dem Anfangsverdacht wies allerdings das Merkmal *need for cognition* (zu verstehen als Freude an und Engagement bei Denkaufgaben) mit einer negativen Korrelation $r = -0.26$ ($p = .01$) auf, d.h. es erwies sich als protektives Merkmal, da ein hohes Maß an *need for cognition* seltener mit einem Anfangsverdacht einherging.

2.4.1.2 Intern-repräsentierter Anfangsverdacht

An dieser Stelle sei auf die klassischen Arbeiten zum konfirmatorischen Hypothesentesten verwiesen (u. a. Ammann, 1999; Baer, 2002). Als Beleg für die Selbstbestätigung wird in diesem Zusammenhang stets angeführt, dass mehr Personen in der Experimentalgruppe mit induziertem Verdacht zu einer Falschbeschuldigung kommen als in der Kontrollgruppe, wofür korrekterweise der Verdacht zur Verantwortung gezogen wird. Darüber hinaus zeigen diese Studien jedoch neben den Belegen für den Einfluss des Missbrauchsverdachts auch zwei unerwartete Phänomene: einerseits das Auftauchen eines Verdachts ohne vorherige Induktion und das nicht Übernehmen eines induzierten Verdachts.

Verdacht ohne vorherige Induktion. In den genannten Studien kann beobachtet werden, dass auch in der Kontrollgruppe ca. 15% der Probanden ein Schuldurteil bzgl. des Sportlehrers bzw. Zauberers fällen. Das heißt, es findet sich unabhängig von der Stichprobe ein ähnlicher Prozentsatz von Personen, die trotz Fehlens eines direkten Verdachts den Zauberer beschuldigen, obwohl ihnen mitgeteilt wurde, dass sie Vergleichsmaterial beurteilen, dass nicht mit Missbrauch im Zusammenhang steht.

Keine Falschbeschuldigung trotz Verdacht. Ebenfalls beachtenswert erscheint der Umstand, dass trotz Vorgabe des Verdachts in der Experimentalbedingung nicht alle Personen eine Falschbeschuldigung aussprechen. Nützlich erscheint in diesem Zusammenhang die Arbeit von Haverkamp (1993), die darauf hinweist, dass das Vorliegen bzw. die Vorgabe einer Hypothese bzw. eines Verdachts nicht notwendigerweise dazu führt, dass die Probanden diese internalisieren und weiter verfolgen. So unterschied sie in ihrer Studie zwischen *experimenter provided* sowie *self generated hypotheses* und konnte feststellen, dass die Inhalte, die die Probanden als Kernproblem angaben (Messung der selbstgenerierten Hypothese) von den Aussagen des fiktiven Patienten auf dem zuvor gesehenen Videoband (Umsetzung der *experimenter provided hypothesis*) abwichen. Die Probanden übernahmen also nicht unbedingt die vorgegebene Annahme, sondern schufen ihren eigenen Interpretationsrahmen. Da diese vor dem Hintergrund absolut identischer Informationsgrundlage geschah, ist ein Einfluss von Persönlichkeitsmerkmalen naheliegend, d.h. ein Einfluss der Person des Testers.

Das Konzept des intern-repräsentierten Anfangsverdachts. Es liegt daher die Vermutung nahe, dass Personen sich in der Bereitschaft unterscheiden, eine induzierte bzw. nahegelegte Hypothese anzunehmen. Es erscheint daher sinnvoll, zwischen der von außen induzierten Hypothese und einer, vorläufig als *intern-repräsentiertem Anfangsverdacht* bezeichneten, individuellen Annahme zu unterscheiden. Die Verfasserin schlägt daher vor, den intern-repräsentierten Anfangsverdacht als vermittelnde Variable zwischen induziertem Verdacht

und konfirmatorischem Testen in die theoretischen Überlegungen mit aufzunehmen. Jochumsen (2009) veranschaulicht diesen Zusammenhang nach Erarbeitung mit der Verfasserin wie folgender Abbildung 4 entnommen werden kann:

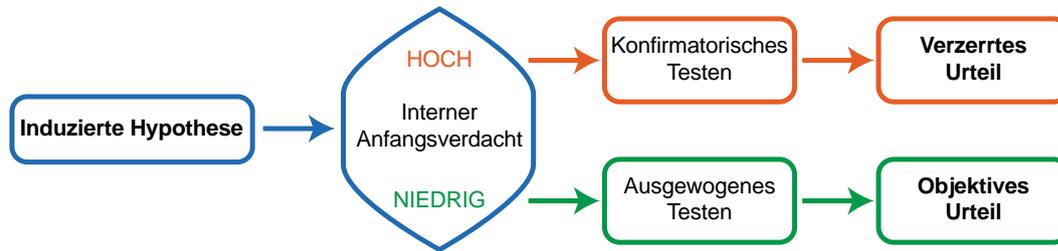


Abbildung 4. Unterscheidung zwischen induzierter Hypothese und internem Anfangsverdacht

Es kann dementsprechend angenommen werden, dass es bei einer hohen Stärke des intern-repräsentierten Anfangsverdachts zu einem konfirmatorischen Testen kommt, während ein niedriger Verdacht eher ein ausgewogenes Testen nach sich ziehen könnte. Welcher Zusammenhang dabei zwischen der Hypothese und dem Anfangsverdacht besteht, d.h. wann aufgrund der identisch dargebotenen Hypothese ein hoher und wann ein niedriger Verdacht resultiert, bleibt zu untersuchen. Allerdings ist es mindestens plausibel, dass an dieser Stelle Persönlichkeitsmerkmale einen Einfluss ausüben könnten, wie ihn bspw. vom Schemm (2008a) für drei Konstrukte expliziert.

Ebenso könnte der intern-repräsentierte Anfangsverdacht erklären, warum Personen in der Kontrollgruppe konfirmatorisch testen und eine Falschbeschuldigung aussprechen – wobei das grundlegende Problem, warum sich der Verdacht entwickeln konnte, davon unberührt bleibt und auch Persönlichkeitsmerkmale hier nicht ausreichend sind – wenn man im normalpsychologischen Bereich bleibend unterstellt, dass es keine stabile und situationsinvariante Tendenz gibt, hinter jeder Interaktion zwischen Sportlehrern bzw. Zauberern und Kindern einen Missbrauch anzunehmen.

2.4.1.3 Der Experte als Hypothesentester

Eine Person ist nicht nur durch ihre Persönlichkeitsmerkmale, sondern auch durch ihr Wissen und ihre bereichsspezifischen Fertigkeiten gekennzeichnet. Man kann sich also die Frage stellen, welche Rolle die Expertise eines Hypothesentesters im Zusammenhang mit der Selbstbestätigungstendenz spielt. Ist die Annahme von Smith und Kida (1991)

tatsächlich zutreffend, dass man die Generalisierbarkeit der Ergebnisse hinsichtlich des *confirmation bias* auf Experten anzweifeln darf, da in den zitierten Studien, die teilweise der Grundlagenforschung entspringen, jeweils äußerst abstrakte und den Probanden unvertraute Aufgabenstellungen eingesetzt wurden (vgl. *rule discovery task* und *four card problem*)? Im Einklang mit Edwards (1983) sowie Hogarth (1981) fordern sie daher, den Einsatz von Heuristiken bzw. das Auftreten von *biases* bei der Bearbeitung von realitätsgetreuen und den Experten vertrauten Aufgaben zu untersuchen, um Übergeneralisierungen zu vermeiden.

Ihre positiv wirkende Hypothese ist daher, dass die Verzerrungen bei realistischeren und vertrauteren Aufgaben weniger deutlich auftreten. Untermuert wird diese Annahme u. a. von den Befunden von Borgida und DeBono (1989). Sie greifen den aus der Expertiseforschung bekannten Fakt auf, dass Experten nicht nur mehr Wissen haben als Novizen, sondern dass dieses Wissen auch anders strukturiert ist (z.B. höhere Differenzierung, stärkere Strukturierung und Hierarchisierung). Es stellt sich nun die Frage, ob Experten dieses Wissen auch anders einsetzen als Novizen. Untersucht wird daher, ob Experten beim Testen einer Hypothese aus ihrem Wissensbereich weniger dazu neigen, konfirmatorische Informationen zu suchen als Novizen – bei Themen außerhalb des Expertenbereich sollten sich keine Unterschiede ergeben. Dafür replizierten sie das Experiment von Snyder und Cantor (1979; s. Abschnitt 2.3.2.2), wobei sie die Gedächtnisleistung von Bibliothekaren (als Experten für die Eignung als Bibliothekarin) mit derjenigen von Studierenden (als Novizen) verglichen. Während die Novizen unabhängig von der Hypothesenbedingung eher bestätigende Hinweise erinnerten, zeigten sich Experten in ihrem Arbeitsbereich sensibler für widersprüchliche Informationen.

Die folgenden Befunde stimmen jedoch nachdenklich. Schweizer (2005) wies in einer Fragebogenuntersuchung die Bestätigungstendenz auch bei Richtern nach. Kreamer und Zierer (1994) sind im Rahmen einer Studie der Frage nachgegangen, ob Experten gegen den *confirmation bias* gefeit sind, indem sie im Bereich der klinischen Diagnostik untersuchten, inwieweit die Bestätigungstendenz durch bereichsspezifisches Wissen und/oder diagnostische Kompetenz des Urteilers vermieden oder vermindert werden kann (s. auch Kap. 2.5). Dabei zeigte sich, dass das Expertentum den Hypothesentester keineswegs automatisch gegen Selbstbestätigungstendenzen bzw. Urteilsfehler immunisiert.

Insofern muss geschlussfolgert werden, dass die Beschäftigung mit dem konfirmatorischen Hypothesentesten auch bei Experten relevant ist. In Abschnitt 2.5 werden hierzu ausführlich Befunde referiert.

2.4.2 Art der zu prüfenden Hypothese

In Bezug auf die Art der Hypothese, die mutmaßlich einen Einfluss auf den Testungsprozess nehmen kann, wird zunächst die Valenz der Prüfhypothese betrachtet, bevor das umfassendere Konzept der Devianzhypothese vorgestellt wird, welches im Zusammenhang mit der Übertragbarkeit des Modells des konfirmatorischen Hypothesentestens diskutiert werden wird.

2.4.2.1 Valenz der Prüfhypothese

Den bisherigen Bezugsrahmen des Modells des konfirmatorischen Hypothesentestens noch weiter verlassend soll an dieser Stelle auf eine weitere potentielle Einflussgröße eingegangen werden. Es handelt sich dabei um die Valenz der zu prüfenden Hypothese, d.h. ob ein allgemein als positiv oder als negativ bewerteter Sachverhalt zu testen ist. Denn dieses Merkmal der zu testenden Hypothese könnte eine Einflussgröße sein, die die Implementierung der induzierten Hypothese als intern-repräsentierten Anfangsverdacht erschwert oder erleichtert.

In einer eigenen Vorarbeit (vgl. Hagemann, 2008; Limbrecht, 2008) sollte der Einfluss der Valenz einer induzierten Hypothese auf den Prüfprozess untersucht werden. Für diesen Zweck wurde eine an das Paradigma der Studien von Ammann (1999) und Baer (2002) angelehnte Studie entwickelt, die jedoch nicht auf den Missbrauchskontext bezogen war, sondern eine Schuleignungsdiagnostik thematisierte. Hierzu wurde eine Fallvignette über einen Viertklässler konstruiert, der aufgrund der insgesamt vorliegenden Informationen eine Eignung für die Realschule aufwies; was den Probanden selbstverständlich nicht bekannt war. Als UV fungierte die Existenz einer Schulempfehlung seitens der Grundschullehrkräfte, die in den drei Abstufungen *negative Valenz* (operationalisiert durch die Vorgabe Hauptschule), *positive Valenz* (Gymnasium) und *keine Hypothese* (Kollegium kann sich nicht einigen) variiert wurde. Das Urteil der Probanden wurde in Form der ausgesprochenen Schulartempfehlung erhoben, nachdem sukzessive alle relevanten Informationen vorgelegt worden waren.

In Form eines so genannten Spontanurteils wurde ausgewertet, ob die Probanden nach der Verarbeitung anfänglicher, allerdings non-diagnostischer Informationen die vorgegebene Hypothese übernahmen oder eine andere Schulform präferierten. Tabelle 2 ist zu entnehmen, dass die Valenz der Hypothese hierauf einen großen Einfluss hatte. Probanden, denen die negativ valente Hauptschulhypothese induziert wurde, übernahmen diese nur zu 17%, wohingegen 65% der Probanden mit positiv valenter Gymnasialhypothese das Gymnasium als Spontanurteil angaben (Freitextformat!).

Tabelle 2. Häufigkeit der Eignungsurteile in Abhängigkeit von der Valenz der Hypothese

		<i>UV: Valenz der Hypothese</i>			
		<i>Negativ Hauptschule</i>	<i>Keine Valenz Keine Empf.</i>	<i>Positiv Gymnasium</i>	<i>Gesamt</i>
<i>Spontanurteil: erste Schulempfehlung</i>	<i>Hauptschule</i>	4	0	0	4
	<i>Realschule</i>	19	16	8	43
	<i>Gymnasium</i>	0	6	15	21
	<i>Gesamt</i>	23	22	23	68

Entgegen der Erwartung, dass die in der Diagonale repräsentierten Zellen, die eine Übereinstimmung zwischen induzierter Hypothese und Spontanurteil andeuten, immer am höchsten besetzt sind, zeigte sich dies für die Bedingung mit der negativ valenten Hypothese nicht. Aufgrund des analogen Studienmaterials, das sich mit Ausnahme des Wortes für die Schulform nicht unterschied, sind die Gründe für diese Abweichung in dem Inhalt der Hypothese zu suchen. Um einen weiteren Hinweis zu finden, dass eine negative Bewertung der Schulart Hauptschule vorlag, wurde eine Nachbefragung der Probanden durchgeführt. Diese ergab eine eindeutig negative Einstellung gegenüber dieser Schulform, indem 84% der Befragten angaben, ihr Kind im Zweifelsfall eher auf die Realschule als die Hauptschule zu schicken.

Insofern kann angenommen werden, dass eine negative Valenz der Hypothese verhindern könnte, dass die Hypothese als intern-repräsentierter Anfangsverdacht übernommen wird. Ein erster Faktor bzw. eine erste Grenze für die Übernahme von induzierten Hypothesen scheint daher gefunden. Aufgrund der Tatsache, dass die beiden Gruppen mit negativer und ohne Valenz sich hinsichtlich des intern-repräsentierten Anfangsverdachts nicht unterscheiden, ist eine weitere Auswertung der erhobenen Variablen für das abschließende Urteil sowie weitere Größen, die die einzelnen Mechanismen des Modells des konfirmatorischen Hypothesentestens darstellen, nicht angezeigt, weil die Gruppen sich nicht wie beabsichtigt unterschieden. Für den Vergleich der positiv valenten Gruppe mit derjenigen ohne Valenz sei an dieser Stelle auf die Arbeit von Limbrecht (2008) verwiesen.

2.4.2.2 *Devianzhypothesen: Übertragbarkeit auf andere Anwendungskontexte*

Dass mit der Valenz der Prüfhypothese ein Merkmal der Hypothese selbst vorliegt, das den Testungsprozess beeinflusst, wird die mit dem Konzept der Devianzhypothese vertrauten Leser wenig überraschen. Vogelgesang, Schulz-Hardt und Köhnken (2006) postulieren, dass sich das Modell von Schulz-Hardt und Köhnken (2000) auf eine ganze Klasse von Hypothesen generalisieren lässt, die als *Devianzhypothesen* bezeichnet werden. Im Folgenden soll dargelegt werden, was unter dem Konzept der Devianzhypothese zu verstehen ist und warum hier ein besonderes Selbstbestätigungspotential anzunehmen ist. Nachfolgend soll eine eigene Vorarbeit zur Übertragbarkeit des Modells des konfirmatorischen Hypothesentestens nach Schulz-Hardt und Köhnken (2000) auf den diagnostischen Prozess bei der Überprüfung einer klinisch-psychologischen Hypothese vorgestellt werden.

Devianzhypothesen. Eine Devianzhypothese liegt vor, wenn der in der Hypothese unterstellte Sachverhalt eine Abweichung von der Norm darstellt. Damit verbunden besteht die Besonderheit von Devianzhypothesen darin, dass sie auf vier Dimensionen einen Extrempol darstellen, der sie von Standardhypothesen unterscheidet. Die Autoren führen in diesem Zusammenhang die Anzahl und Spezifität der Alternativhypothesen, die Basisrate von Devianz- und Prüfhypothese, die Valenz des hypostasierten Ereignisses und die emotionale Involviertheit des Testers an.

Anzahl und Spezifität der Alternativhypothesen. Wie im Zusammenhang mit dem Modell des sozialen Hypothesentestens nach Trope und Liberman (1996) bereits beschreiben, werden nicht nur die zu testende Hypothese, sondern im besten Fall auch eine oder mehrere Alternativhypothesen aufgestellt. Die Anzahl denkbarer Alternativhypothesen variiert dabei von Hypothese zu Hypothese, außerdem können die Alternativen unterschiedlich konkret bzw. spezifisch sein. Eine Devianzhypothese zeichnet sich dadurch aus, dass der Prüfhypothese (z. B. des juristischen und moralischen Normverstoßes: *Person X hat einen Kindesmissbrauch begangen*) nur die unspezifische Alternativhypothese der Normalität gegenüber steht (*X hat keinen Kindesmissbrauch begangen*). Ein potentiell selbstbestätigender Effekt ergibt sich laut Autoren nun daraus, dass bei solch einseitig spezifischen Hypothesen nur die Hypothese, nicht aber die Alternativhypothese ein Interpretationsmuster für uneindeutige Evidenz bereitstellt. Sie illustrieren dies anhand des folgenden Beispiels: Die Beobachtung, dass ein Kind schüchtern ist, ist im Sinne des Missbrauchs so interpretierbar, dass das Kind durch den Missbrauch Angst vor Fremden

hat. Warum das Kind schüchtern sein sollte, obwohl kein Missbrauch vorliegt, könnte viele Ursachen haben und ist daher nicht konkret greifbar: vielleicht liegt es an seiner Persönlichkeit im Allgemeinen, vielleicht ist es traumatisiert, weil die Eltern sich haben scheiden lassen, die Oma gestorben ist oder die Katze überfahren wurde. Vielleicht hat auch die Lehrerin in der Schule mit dem Kind geschimpft, so dass es seit Längerem ein niedriges Selbstwertgefühl hat und sich nicht traut, in der Gegenwart von Erwachsenen zu sprechen. Das Interpretationsmuster im Missbrauchskontext ist um einiges greifbarer.

Basisraten von Hypothese und Alternativhypothesen. Da Normabweichungen sich definitionsgemäß auf seltene Ereignisse beziehen, begünstigt diese niedrige Basisrate die Selbstbestätigung einer Hypothese, wenn man pseudodiagnostisch und zugleich positiv testet.

Valenz des hypostasierten Ereignisses. Der zu prüfende Sachverhalt kann für die testende Person positiv, neutral oder negativ besetzt sein (vgl. Kap. 2.4.2.1). Dies kann nun nicht nur die Annahme einer induzierten Hypothese beeinflussen, sondern nach deren Annahme u. a. die Schwellen für Annahme und Ablehnung der Hypothese: Bei negativen Sachverhalten wird üblicherweise eine niedrige Annahme- und eine hohe Ablehnungsschwelle gesetzt, denn so lässt sich der Fehler der falschen Zurückweisung minimieren, der andernfalls bedeuten würde, dass der negative Zustand unentdeckt weiter fortbesteht.

Emotionale Involviertheit des Testers. Hier ist ein enger Zusammenhang mit der Valenz der Hypothese zu erwarten, da eine emotionale Involviertheit des Testers vor allem bei positiv und negativ valenten Sachverhalten auftreten sollte. Bei Devianzhypothesen, die generell eine negative Valenz aufweisen (z. B. Straftatverdacht), sollte sie dementsprechend hoch sein. Ihren verzerrenden Einfluss übt die emotionale Involviertheit wieder über eine erhöhte gedankliche Beschäftigung mit der Hypothese, aus, die – wie in Abschnitt 2.3.2.1 ausgeführt – die Selbstbestätigung begünstigt.

Die Autoren schließen daraus, dass eine Generalisierung der Selbstbestätigungsmechanismen auch auf anderen Themengebiete denkbar ist, wenn es sich um eine Testung von Devianzhypothesen handelt.

Übertragbarkeit auf den diagnostischen Prozess. In einer eigenen Vorarbeit (vgl. Dreger, 2006) wurden die angenommenen Effekte der Verzerrungsmechanismen des Modells des konfirmatorischen Hypothesentestens im klinisch-psychologischen Kontext der Diagnostik

einer Bulimie bzw. allgemein einer Essstörung überprüft. Die zentrale Forschungsfrage bezog sich darauf, ob Personen mit dem Auftrag, das Vorliegen einer spezifischen (objektiv falschen) psychischen Störung zu überprüfen, häufiger fälschlicherweise diese Störung diagnostizieren als Personen, die einen unspezifischen Auftrag erhalten. Darüber hinaus sollte die Existenz eines verzerrten Informationsverarbeitungsmusters im Sinne der oben dargestellten Mechanismen untersucht werden.

Die Vorgabe einer psychischen Störung als Hypothese trägt dem Konzept der Devianzhypothese Rechnung, denn es handelt sich hier um eine Normabweichung hinsichtlich des Erlebens und Verhaltens von Menschen. Darüber hinaus bietet die Hypothese *Störung* ein eindeutiges Interpretationsmuster für geschildertes Verhalten, während die Alternativhypothese *Gesundheit* kein solches bietet. Es ist daher von einer geringen Anzahl und Spezifität der Alternativhypothesen auszugehen, was das erste Merkmal von Devianzhypothesen darstellt. Außerdem sind die Basisraten von psychischen Störungen gegenüber Gesundheit jeweils als gering zu bewerten und beinhalten stets einen negativ valenten Sachverhalt. Letzteres kann einerseits damit begründet werden, dass die einschlägigen Diagnosesysteme stets das Vorhandensein von Leiden und Einschränkungen als notwendige Begleiterscheinungen fordern. Andererseits gelten psychische Störungen auch heute vielfach noch als Tabuthema und werden mit Verrücktheit in Zusammenhang gebracht. Hiermit kann eine negative gesellschaftliche Bewertung von psychischen Störungen bzw. Menschen mit psychischen Störungen in Verbindung gebracht werden. Aufgrund dieser negativen Valenz ist zudem eine emotionale Involviertheit zu erwarten, so dass das Konzept der Devianzhypothese erfüllt scheint.

In der PC-gestützten Studie wurde ein einfaktorieller Versuchsplan mit der UV *Spezifität der Hypothese* und den beiden Bedingungen *falsch-spezifische Hypothese* (Experimentalgruppe, Verdacht auf Bulimie) und *unspezifische Hypothese* (Kontrollgruppe, Verdacht auf Essstörung) umgesetzt. Den $N=60$ Studierenden der Psychologie ab dem 7. Fachsemester wurde eine eigens konstruierte Fallvignette einer jugendlichen Anorexie-Patientin, Subtypus Binge-Eating/Purging (ICD-10: F50.01) ausgehändigt, die durch sukzessiv ergänzte Informationen vervollständigt wurde, bis ein eindeutiges Urteil möglich war.

Die Hypothesen konnten in weiten Teilen bestätigt werden: Personen mit einer falschen spezifischen Hypothese stellen häufiger die falsche Diagnose (Bulimie) als Personen mit einer unspezifischen Annahme und wiesen dabei sowohl ein stärker ausgeprägtes selektives Erinnern konsistenter Informationen, ein deutlicher pseudodiagnostisches Testen und eine größere Tendenz zur hypothesenkonsistenten Interpretation uneindeutiger Evidenz auf. Eine Präferenz für die positive Teststrategie konnte hingegen nicht gefunden werden (für eine ausführliche Diskussion s. Dreger (2006)).

Eine deutliche Einschränkung der Interpretierbarkeit bzw. Generalisierbarkeit der Ergebnisse entsteht allerdings aus der Wahl studentischer Diagnostiker, die zwar über die notwendigen theoretischen Kenntnisse verfügten, jedoch über kaum bis gar keine praktischen Kompetenzen der klinisch-psychologischen Diagnostik. Auf den weiteren Studienbedarf an erfahrenen Praktikern wurde daher hingewiesen. Der folgende Abschnitt der vorliegenden Arbeit widmet sich außerdem genau diesem Umstand.

2.5 Confirmation bias, Berufserfahrung und Fachwissen

Nachdem die Annahme, die Resultate der Forschung an Laien seien nicht ohne Weiteres auf Experten bei der Bearbeitung realistischer Aufgaben zu übertragen (Edwards, 1983; Hogarth, 1981; Smith & Kida, 1991) widerlegt wurde, da auch Diagnostiker im Training (z.B. Parmley, 2006; Strohmer et al., 1990) bzw. erfahrene Diagnostiker bei Vorstellung eines (schauspielenden) Patienten in ihrer eigenen Praxis (Strohmer & Shivy, 1994) konfirmatorisch verzerrte Urteilsprozesse aufzeigen, ist eine genauere Auseinandersetzung mit dem *confirmation bias* in Abhängigkeit von der Berufserfahrung und dem spezifischen Fachwissen nötig. Denn dies könnte erste Ansatzpunkte darstellen, um Gegenmaßnahmen gegen das konfirmatorische Testen abzuleiten und in Trainings zu implementieren.

Bezüglich des Zusammenhangs von Berufserfahrung und *confirmation bias* sind (zumindest) auf Ebene der Plausibilität zwei gegensätzliche Szenarien denkbar. Einerseits könnte man annehmen, dass Berufsanfänger im Gegensatz zu Berufserfahrenen eher eine konfirmatorische Strategie einsetzen, weil die Aufgabe für sie kognitiv noch herausfordernder ist als für die Berufserfahrenen. Aufgrund der Befunde von Stangor und McMillan (1992) bzgl. der Selektivität des Gedächtnisses unter kognitiv herausfordernden Bedingungen sowie diejenigen von Doherty et al. (1979) zur Vernachlässigung der Diagnostizität von Informationen im Falle nicht komplett vorgegebener, sondern zu erschließender Angaben zur Bestimmung derselben, liegt diese Betrachtung durchaus nahe.

Andererseits könnte man auch vermuten, dass die Berufserfahrenen eher einen *confirmation bias* aufweisen als die Berufsanfänger, wenn man unterstellt, dass Erstere eine höhere subjektive Sicherheit bei all ihren Entscheidungen aufweisen. Denn dies würde u.a. die Asymmetrie der Schwellenwahl beeinflussen. Berufserfahrene Diagnostiker haben u.U. bereits ein klares Bild über eine spezifische Symptomatik im Kopf, die durch verschiedenste Fälle veranschaulicht wird und sind daher womöglich bei der Einordnung gefundener Informationen sicherer als Berufsanfänger. Letztere könnten in dieser Situation eher geneigt sein, die Unsicherheit dadurch zu reduzieren, dass die Kongruenz mit verschiedenen

Alternativen geprüft wird, was dem *confirmation bias* entgegenwirken würde. Weitere Kompensationsmechanismen wie die umfängliche Nutzung von Diagnoseschemata mit veranschaulichter Differentialdiagnostik sind denkbar. Geringe Berufserfahrung könnte über die Strategie der Nutzung unsicherheitsreduzierender Maßnahmen daher auch einen protektiven Einfluss haben.

Während die bislang referierten Studien daher untersuchten, ob Berufserfahrung per se ein protektiver Faktor ist, und dafür i.d.R. ausschließlich eine Gruppe (mehr oder weniger) erfahrener Diagnostiker analysierten, interessieren im Folgenden Ergebnisse, aus einem Vergleich von weniger und mehr erfahrenen Diagnostikern.

In der eigenen Vorarbeit (Dreger, 2006; s. Abschnitt 2.4.2.2) bezüglich einer Essstörungsdiagnostik zeigte sich, dass der Ausbildungsvorteil der Erfahreneren nur dann mit einer korrekten Diagnose einherging, wenn ihnen keine spezifische Hypothese vorgegeben wird, sondern eine Störungskategorie, die die Alternativdiagnosen impliziert. Erfahrenere Probanden konnten demnach in einer für den *confirmation bias* weniger anfälligen Situation Vorteile vor den weniger Erfahrenen erzielen. Diese Erfahrung schützte jedoch nicht länger, wenn eine spezifische Hypothese vorgegeben wurde. Beide Gruppen zeigten ein äquivalentes Diagnosemuster mit vielen konfirmatorisch verzerrten Diagnosestellungen. Die unterschiedliche Erfahrung führte also unter den verzerrenden Bedingungen nicht zu Unterschieden. Allerdings ist aufgrund der Wahl der Probanden interpretatorische Vorsicht in Bezug auf die Generalisierung geboten, da alle Probanden Psychologie-Studierende waren und der Erfahrungsunterschied sich nur aus dem unterschiedlichen Fachsemester ergab.

Bei Haverkamps Probanden (1993, s. Abschnitt 2.3.3.2), die im Mittel drei Jahre Berufserfahrung aufwiesen, da sie in einschlägigen Trainingsprogrammen rekrutiert wurden, zeigte sich die Berufserfahrung weder als Anzahl der Jahre im Trainingsprogramm noch als Anzahl der Monate von Berufserfahrung als Prädiktor eines konfirmatorischen Testens. Problematisch bleibt jedoch der Umstand, dass die Probanden aufgrund der Rekrutierung aus den Ausbildungsprogrammen heraus nicht allzu heterogen hinsichtlich ihrer Berufserfahrung sein konnten. Eine größere Variabilität erzielte Parmley (2006, S. Abschnitt 2.6), deren Probanden 1 bis 52 Jahre Berufserfahrung berichteten. Trotz dieses großen Ranges hatten ca. 50% eine Berufserfahrung von maximal sechs Jahren, so dass von einem großen Anteil von Berufsanfängern ausgegangen werden muss. Auch hier konnte keine Vorhersagekraft der Berufserfahrung für das Auftreten des *confirmation bias* feststellen werden. Außerdem hatten Haverkamp (1993) und Parmley (2006) den Effekt nur anhand von exploratorischen Analysen in Form von Zusammenhängen der demographischen Variable Berufserfahrung mit dem Urteil vorgenommen.

Einen systematischen Vergleich stellte hingegen Martin (2000) an, indem er das

Informationsverarbeitungsmuster von je 40 Teilnehmern untersuchte, die entweder mindestens in ihrem zweiten Jahr des *clinical or counseling psychology doctoral program* waren und im Mittel ca. 550 Stunden an psychotherapeutischer Arbeit berichteten, oder zu den *undergraduate psychology students* gehörten. Hier besteht demnach ein deutlich größerer Unterschied zwischen den Probandengruppen hinsichtlich ihrer Berufserfahrung. Martin (2000) verglich in seiner Arbeit daher Personen mit fortgeschrittenem Training bzw. Berufserfahrung (Experten) mit Personen ohne Training bzw. Berufserfahrung (Naive) hinsichtlich ihrer Strategien zur Hypothesentestung, wobei er sich auf Unterschiede in der Häufigkeit von konfirmatorischen, neutralen und diskonfirmatorischen Fragen konzentrierte, die die Probanden zu drei Zeitpunkten bei einer Interaktion mit einer (schauspielenden) Klientin notierten. Konform mit Martins Erwartungen (2000) stellten die Experten mehr diskonfirmatorische Fragen als die Naiven. Keine Unterschiede ergaben sich bei der Anzahl der konfirmatorischen und neutralen Fragen. Insofern ließ sich nur eine Teilbestätigung für den von Martin angenommenen Vorteil für Berufserfahrene finden.

Während demnach mehrere Hinweise aus Studien mit postgraduierten Diagnostikern vorliegen, die andeuten, dass die Berufserfahrung keinen oder zumindest nur einen geringen Beitrag zur Vermeidung des *confirmation bias* leistet, so fehlt es zudem vollständig an aussagekräftigen Ergebnissen für Personen mit einer ausgeprägten Berufserfahrung von mindestens zehn Jahren. Ob und wenn ja, wie sich Berufsanfänger von Personen mit mehreren Jahrzehnten Erfahrung unterscheiden, bleibt weiterhin ungeklärt. Um methodische Schwächen von Einzelstudien zu umgehen, sollen im Anschluss die Befunde eines Reviews (Garb, 1989) sowie einer aktuellen Meta-Analyse (Spengler et al., 2009) referiert und diskutiert werden.

Vorab sei erwähnt, dass beide Beiträge eine Unterscheidung zwischen Berufserfahrung und Training machen. Um mit den Worten von Spengler et al. (2009) zu sprechen, kann man (zumindest theoretisch) zwischen einer *clinical experience* und einer *educational experience* unterscheiden. *Clinical experience* als Berufserfahrung kann verstanden werden als die Erfahrungen, die durch den wiederholten Kontakt mit der relevanten Klientenpopulation die Gelegenheit bieten, existierende Strategien über die Zeit hinweg zu wiederholen und damit zu optimieren. Im Unterschied dazu ist die *educational experience* auch als Trainingseffekt zu verstehen, im Rahmen dessen u.U. die Qualität des Urteilsprozesses verbessert werden kann, indem neue Strategien vermittelt werden oder durch Supervision Fehler erkannt und korrigiert werden können. Es geht hier also um stärker institutionalisierte Aspekte der Anreicherung der fachlichen Kompetenzen. Ob

u.U. unterschiedliche Effekte mit den beiden Erfahrungsarten verbunden sind, wird in den beiden Darstellungen der Arbeiten angesprochen.

Das narrative Review von Garb (1989). Garb (1989) analysierte 55 Studien, die den Zusammenhang von Berufserfahrung sowie Training und der Validität der gefällten Urteile untersuchten. Dabei formulierte er folgende Einschlusskriterien: die Studien mussten erstens mindestens eine Gruppe beinhalten, die professionell trainiert war, zweitens einen Urteilsprozess untersuchen, der für die klinische Psychologie relevant ist, drittens die Validität der Urteile bewerten und viertens eine Relation zwischen der Validität der Urteile und dem Ausmaß von Erfahrung bzw. Training herstellen.

Die Befunde von Garb (1989) bzgl. des Zusammenhangs von *Berufserfahrung* und Urteilsgüte fallen ernüchternd aus: so zeigte sich bspw. bezüglich der Einschätzung von Persönlichkeitseigenschaften - sei es auf Basis von projektiven Testverfahren, des *Minnesota Multiphasic Personality Inventory* (MMPI) oder direkten Beobachtungen - keine Überlegenheit der erfahrenen gegenüber den wenig erfahrenen Klinikern. Desweiteren konnten die Kliniker im Vergleich zu Studierenden keine Überlegenheit demonstrieren; ein wiederum robuster Befund unabhängig von der Vorgabe von Interviewdaten, Therapieprotokollen, biographischen Informationen oder projektiven Testverfahren.

Die Befundlage zum Zusammenhang von *Training* und Urteilsgüte ist hingegen inkonsistenter und Garb (1989) interpretiert abschließend, dass es bestimmte Bedingungen gibt, unter denen trainierte Personen besser abschneiden als untrainierte Laien. Sie zeigten sich bspw. bei der Interpretation von *WAIS*-Ergebnissen (*WAIS*: Wechsler Adult Intelligence Scale) hinsichtlich einer Schizophrenie und bei Nutzung von biographischen Daten psychiatrischer Personen sowie *MMPI*-Profilen den Laien (i.d.R. *undergraduates*) überlegen. Im Zusammenhang mit Interviewdaten berichtet Garb (1989) von gemischten Resultaten, wohingegen es bei der Nutzung von projektiven und neuropsychologischen Daten keine Überlegenheit der Trainierten gegeben habe.

Die Meta-Analyse von Spengler et al. (2009). Ein positiveres Bild vom Zusammenhang der Berufserfahrung (sowohl im Sinne der *clinical* als auch *educational experience*) und der Urteilsgüte zeichnen hingegen Spengler et al. (2009) in ihrer Metaanalyse im Rahmen des großangelegten *Meta-Analysis of Clinical Judgment Project* (MACJ), wofür die Autoren einen Studienpool von 1135 zwischen 1970 und 1996 erschienenen Studien aus über 35000 potentiell relevanten Publikationen herausgefiltert haben, die für verschiedene Forschungsfragen genutzt wurden. Für die vorliegende Meta-Analyse wurden 75 Studien

davon berücksichtigt, die den folgenden Kriterien genügten: erstens hat die Studie einen Fokus auf klinischer Urteilsbildung bzw. damit zusammenhängenden Urteilsverzerrungen, zweitens mussten *mental health issues* oder psychologische Konstrukte involviert sein. Ein drittes Einschlusskriterium bezog sich auf den Status der Probanden, der entweder *mental health professional* oder *graduate-level trainees* sein musste. Das vierte Kriterium bezog sich auf die Möglichkeit, anhand der berichteten Daten Effektstärken berechnen zu können.

Die Autoren ermittelten eine mittlere Effektstärke von $d=0.12$, die anzeigt, dass es einen kleinen aber reliablen, positiven Zusammenhang zwischen Erfahrung und Zuverlässigkeit des Urteils gibt. Moderatorvariablen spielten bei den Analysen zudem kaum eine Rolle, was die Robustheit des Erfahrungsaspekts unterstreicht. Interessanterweise fanden die Autoren auch keinen Hinweis auf einen differentiellen Einfluss von *clinical vs. educational experience*, wie man ihn bei Garb (1989) ableiten könnte.

Wie können nun die Unterschiede zwischen den Befunden von Garb (1989) und Spengler et al. (2009) erklärt werden? Zunächst ist auf eine unterschiedliche Methodik hinzuweisen, da es sich einerseits um ein narratives Review und andererseits um eine quantitative Metaanalyse handelt. Ein methodischer Vorteil ist demnach auf der Seite von Spengler et al. (2009) zu verzeichnen, die zudem eine breitere Literatursuche vorgenommen haben als Garb (1989). Entscheidend ist jedoch auch der Unterschied hinsichtlich der berücksichtigten Studien. Während Spengler und Kollegen erst Studien ab den 1970igern in ihr MACJ-Projekt aufgenommen haben, spielen in Garbs Review auch viele Studien aus den 1950er und 1960er Jahren eine Rolle. Die Studienpools der beiden Arbeiten unterscheiden sich demnach systematisch nach dem Alter. Nicht auszuschließen sind daher auch Einflüsse der Ausbildungsorganisation und theoretischen Ausrichtung der betrachteten Diagnostiker, da die Psychologie mit dem Wandel von der psychoanalytischen hin zur verhaltenstherapeutischen Schule eine große Veränderung vorgenommen hat. Da der Studienpool von Spengler et al. (2009) in diesem Sinne der aktuellen Situation der Psychologie angemessener ist als derjenige von Garb mit einem hohen Anteil an psychodynamisch orientierter Diagnostik - mit all ihren eigenen Validitätsproblematiken - sieht die Verfasserin mehr Argumente auf der Seite des vorsichtig optimistischen Blickwinkels von Spengler et al. (2009), wobei noch einmal betont werden soll, dass der vorgefundene Effekt nach den gängigen Klassifikationen (Cohen, 1988) als klein zu beurteilen ist. Nach Ansicht der Verfasserin sollte dies nicht deshalb weniger beachtet werden, weil es in anderen Bereichen in- und außerhalb der Psychologie auch kleine Effekte gibt, die eine große Beachtung erfahren. Es ist vielmehr anzunehmen, dass es weitere entscheidende Einflussgrößen gibt, die die Validität der Urteile beeinflussen und

in dieser Konsequenz auch einen Zusammenhang mit dem *confirmation bias* aufweisen könnten.

Wenn es alleine die Berufserfahrung mit der Gelegenheit zum Kennenlernen der Klientenpopulation, den einschlägigen Diagnosemethoden und der edukativen Komponente von Weiterbildung und Supervision nicht ist, liegt die Vermutung nahe, es könnte das einschlägige Fachwissen bzgl. des konkreten Urteilsgegenstands sein, also z.B. das individuelle Fachwissen über die interessierende Störungskategorie bei der Diagnosestellung. Krems und Zierer (1994) sind genau dieser Frage nachgegangen.

Die Abhängigkeit des confirmation bias vom Fachwissen – die Studie von Krems und Zierer (1994). Konkret interessierten sich Krems und Zierer (1994) dafür, inwieweit die Bestätigungstendenz durch bereichsspezifisches Wissen und Können des Urteilers vermieden oder vermindert werden kann. Es ging ihnen daher um die Untersuchung, ob für die Fehlervermeidung das fallspezifische Wissen, die allgemeine diagnostische Kompetenz (die Berufserfahrung im obigen Sinne) oder beides ausschlaggebend ist.

Die Autoren legten dafür vier Probandengruppen (Spezialisten vs. Nicht-Spezialisten orthogonal zu erfahrenen vs. unerfahrenen Diagnostikern) eine Verdachtsdiagnose sowie sukzessive Informationen zu Symptomen eines potentiellen Krankheitsbildes vor, die für die Verdachts-, eine Alternativ- oder beide Diagnosen sprachen. Aufgabe der Probanden war sodann die Überprüfung der Korrektheit der Verdachtsdiagnose anhand der vorgelegten Symptome und gegebenenfalls die Nennung einer Alternativdiagnose. Die Verdachtsdiagnose war dabei in allen Fällen inkorrekt, eine Beibehaltung ist gleichzusetzen mit dem *confirmation bias*. Anzumerken ist an dieser Stelle, dass aufgrund der Fallkonstruktion und Probandenwahl grundsätzlich alle Probandengruppen über das notwendige Wissen verfügten, die Fälle korrekt zu beurteilen.

Als zentrales Merkmal für die Güte des Urteilsprozess nahmen Krems und Zierer (1994) die Flexibilität an, mit der Hypothesen modifiziert bzw. aufgegeben werden. In diesem Sinne ist bei hoher Flexibilität von einer niedrigen Wahrscheinlichkeit eines *confirmation bias* auszugehen. Es stellt sich nun die Frage, ob die Flexibilität v.a. durch ein episodisches, fallbezogenes Wissen als Erfahrungen mit unterschiedlichen Krankheitsbildern im Berufsalltag resultiert (Fachwissen) oder eine allgemein charakteristische Fähigkeit eines diagnostischen Problemlösens ist (Berufserfahrung), da Letzteres bedeuten würde, dass auch Personen außerhalb ihres fallbezogenen Wissens bei Diagnoseaufgaben hinreichend flexibel sind und valide urteilen können. Falls die Flexibilität hingegen unmittelbar vom fallspezifischen Wissen abhängt, so ist bei diagnoseerfahrenen Probanden nur innerhalb ihres Spezialgebiets zu erwarten, dass schon bei einer vergleichsweise geringen Anzahl

an widersprüchlichen Befunden die Diagnose korrigiert wird.

Es zeigte sich, dass die Spezialisten tatsächlich jeweils weniger Informationen benötigten, bis sie erstmals die Differentialdiagnose in Betracht zogen, die Flexibilität war ausschließlich vom fallspezifischen Wissen abhängig, nicht von allgemeiner Diagnoseerfahrung. Außerdem beeinflusste auch ausschließlich der Faktor Spezialisierung die Konsequenz, mit der die Alternativdiagnose beibehalten wurde. Krems und Zierer (1994) schlussfolgern daraus, dass in der Gruppe der Spezialisten daher die Bestätigungstendenz deutlich geringer ausgeprägt ist als bei Probanden, die zwar die für die Diagnose notwendigen Grundkenntnisse, aber kein zusätzliches Spezialwissen ausgebildet haben. Erfahrene Nicht-Spezialisten waren hingegen nicht von unerfahrenen Spezialisten zu unterscheiden und konnten von ihrem Vorsprung im diagnostischen Problemlösen innerhalb anderer Teilgebiete nicht profitieren.

Für die Bewertung des Stellenwerts der Studie müssen allerdings noch einige Informationen nachgereicht werden. Einerseits ist die Stichprobengröße von $N = 32$, d.h. je 8 Probanden pro Probandengruppe als äußerst gering einzustufen, so dass Replikationen geboten sind. Andererseits blieb bislang unerwähnt, dass die Studie sich thematisch nicht mit dem Bereich der klinisch-*psychologischen* Diagnostik beschäftigte, sondern der medizinischen Diagnostik, da je acht Fachärzte der Inneren Medizin bzw. Neurologie (Facharzt = Erfahrener mit im Mittel ca. 13-14 Jahren Berufserfahrung) sowie je acht Ärzte in Facharztausbildung Innere Medizin bzw. Neurologie (in Facharzt-Ausbildung = Unerfahrene mit im Mittel ca. 3 Jahren Berufserfahrung) mit Krankengeschichten aus der Inneren Medizin konfrontiert wurden (Personen in bzw. mit abgeschlossener Facharztausbildung Innere Medizin = Spezialisten).

Eine Untersuchung an einer hinreichend großen Stichprobe im Kontext einer klinisch-psychologischen Diagnosestellung scheint daher angezeigt, um die Robustheit dieses äußert interessanten Effekts zu prüfen.

2.6 Maßnahmen zur Vermeidung des confirmation bias

Während alle bisher herangezogenen empirischen Arbeiten die Entstehung des *confirmation bias* dadurch untersuchten, dass dessen Auftreten durch experimentelle Manipulationen provoziert werden sollte, ist noch ein anderer Zugang denkbar: auch durch den Nachweis, wann eine Selbstbestätigung nicht bzw. weniger stark ausgeprägt auftritt, können Rahmenbedingungen genauer aufgeklärt werden. Insofern sollen auch diejenigen Studienbefunde zusammengefasst werden, die sich mit Maßnahmen zur Vermeidung des

confirmation bias beschäftigen, da hierdurch indirekt Belege für Einflussmechanismen geliefert werden können. Im Folgenden werden sowohl hilfreiche wie weniger hilfreiche Ansätze diskutiert, die auch als *debiasing techniques* bezeichnet werden.

Da die Zunahme der allgemeinen Urteilsgüte mit zunehmender Berufserfahrung nur in begrenztem Maß vorhanden ist (vgl. Spengler et al., 2009) und unmittelbare Zusammenhänge der Berufserfahrung mit dem *confirmation bias* in den zitierten Arbeiten teilweise gar nicht (vgl. Haverkamp, 1993; Parmley, 2006) oder nur auf einzelnen Maßen nachweisbar waren (vgl. Martin, 2000), ist es keine angemessene Maßnahme, rein auf den Faktor Zeit zu setzen, vielmehr muss mit gezielten Maßnahmen gekontert werden. Nach der Argumentation von Dawes (1994) dürfe man auch gar nicht erwarten, dass die Urteilskompetenzen mit zunehmender Berufserfahrung steigen, da die relevanten Aufgaben so mehrdeutig sind und Feedback selten mit so großer Sicherheit wie zum Beispiel in der Medizin erfolgen könne.

Be unbiased. Die einfachste denkbare Intervention ist es, die Diagnostiker über den *confirmation bias* aufzuklären und sie darauf hinzuweisen, diese Selbstbestätigungstendenz bitte zu vermeiden, d.h. die Aufforderung *be unbiased* zu transportieren. Parmley (2006) untersuchte in ihrer Studie, inwieweit eine ausführliche Aufklärung über den *confirmation bias* zu dessen Reduktion beitragen kann. Der Hälfte ihrer 102 Versuchspersonen, die alle lizenzierte Therapeuten waren, legte sie daher eine sehr ausführliche Beschreibung des *confirmation bias* vor, im Rahmen derer sie erstens die Selbstbestätigungstendenz definierte, zweitens die wichtigsten Forschungsergebnisse zusammenfasste, indem sie bspw. Arkes und Harkness (1980), Haverkamp (1993) sowie Temerlin (1968) zitierte. In einem dritten Schritt gab sie Hinweise, wie der *confirmation bias* zu vermeiden sei. Sie empfahl daher nach widerlegenden Informationen zu suchen, erst so spät wie möglich eine Diagnose zu stellen und dabei zu berücksichtigen, dass es eine Arbeitshypothese sei, die im Lichte weiterer Informationen anzupassen sei. Außerdem sollten umfangreiche Notizen gemacht werden, um nicht auf das fehleranfällige Gedächtnis zurückgreifen zu müssen. Die andere Hälfte der Probanden erhielt eine solche Instruktion nicht, bevor alle nach einer Woche Pause den zweiten Teil einer Fallbeschreibung zu lesen bekamen. Anders als angenommen wiesen die instruierten Probanden jedoch keinen geringeren Anteil des *confirmation bias* auf, sondern zeigten ihn mit einer vergleichbaren Häufigkeit. Die Wirkungslosigkeit der Maßnahme steht im Einklang mit der Bewertung von Arkes (1981): „*One technique that has been proven to be absolutely worthless is telling people what a particular bias is and then telling them not to be influenced by it.*“ (p. 326). Wenngleich die von Parmley (2006) in ihrer Aufklärung formulierten Hinweise, wie dem *confirmation*

bias begegnet werden kann, sinnvoll erscheinen und unmittelbar aus Studienergebnissen abgeleitet sind, wie auch Garb (1989) es vorschlägt, um einer schnellen Festlegung auf eine Hypothese (vgl. Sandifer et al., 1970) und den Erinnerungsverzerrungen durch eine Diagnose (vgl. Arkes & Harkess, 1980) vorzubeugen, so ist der reine Hinweis nicht ausreichend, damit die Strategien eingesetzt werden. Garb (1989) fordert dafür ein gezieltes Training, nähere Ausführungen macht er allerdings nicht.

Konkretisierungen lassen sich jedoch u. a. bei Arkes (1981), Haverkamp (1994), Kurpius, Benjamain und Morran (1985) sowie Miller (1985) finden. Die Autoren formulieren weitestgehend übereinstimmende *debiasing techniques*, die überblicksartig in Tabelle 3 veranschaulicht sind.

Tabelle 3. Überblick über *debiasing techniques* beim Hypothesentesten

<i>Debiasing technique</i>	<i>Kurzbeschreibung</i>
<i>Generieren alternativer Hypothesen</i>	Es soll ein möglichst großer Pool an Erklärungen gefunden werden, die die vorliegende Situation erklären könnten und daher zu überprüfen sind. Bestenfalls werden die Gründe hierfür verschriftlicht.
<i>Reduktion der Anforderungen an das Gedächtnis</i>	Gemeint ist die Nutzung jeglicher erdenklicher Hilfsmittel, die die Bedeutung des Gedächtnisses bei der Informationssuche und Bewertung reduzieren (z.B. halbstandardisierte Interviewleitfäden, damit keine Themen im Gespräch vergessen werden; Aufzeichnung des Gesprächs auf Tonband statt eines Gedächtnisprotokolls etc.)
<i>Suchen von alternativen Erklärungen/ Gegenargumenten</i>	Sobald Informationen gefunden sind, sollte man hinterfragen, welche Erklärungen alle für sie verantwortlich sein könnten. Zu gezogenen Schlussfolgerungen sollen stets Gegenargumente gesucht werden, d.h. vorliegende Daten, die in eine andere Richtung weisen.

Generieren alternativer Hypothesen. Das Generieren alternativer Hypothesen steht auf theoretischer Ebene in engem Zusammenhang mit dem diagnostischen Vorgehen beim Hypothesentesten, welches sicherstellt, dass gefundene Informationen u.a. hinsichtlich ihrer Auftretenswahrscheinlichkeit unter den anderen Hypothesen betrachtet werden. Empirische Hinweise, dass Instruktionen bzw. allgemeiner ein Training dazu beitragen können, Urteilsverzerrungen zu reduzieren, kommen aus der Forschung zum *hindsight bias*. So zeigten einerseits die Probanden von Slovic und Fischhoff (1977) eine weniger verzerrte Reaktion, wenn sie angehalten waren, die verschiedenen Ausgänge

des Ereignisses vor ihrem Urteil zu durchdenken. Analog reduzierte sich der *bias* bei Probanden, die Gründe für und gegen vorgegebene Alternativen notieren mussten (vgl. Koriat, Lichtenstein & Fischhoff, 1980). Diese Befunde lassen vermuten, dass die Beschäftigungen mit Alternativen vor dem Urteilen abmildernd wirkt. Anzumerken ist auch, dass es sich jeweils um relativ kurze Instruktionen bzw. Aufgaben gehandelt hat, die in einem Zeitraum von einigen Minuten bearbeitet wurden. Größere Effekte bei zeitintensiveren Interventionen wären daher zumindest plausibel.

Reduktion der Anforderungen an das Gedächtnis. Dieser Umstand leitet sich unmittelbar aus Arkes und Harkness' (1980) Befund ab, dass Personen nach der Diagnosestellung öfter fälschlicherweise Symptome wiedererkennen, die störungskonsistent sind, dem beschriebenen Patienten aber nicht zugeschrieben worden waren (s. auch Abschnitt 2.3.2.2).

Suchen von alternativen Erklärungen bzw. Gegenargumenten. Im Unterschied zum Generieren von alternativen Hypothesen, welches eher zu Beginn des Testungsprozesses relevant ist, bezieht sich dieser Ansatz auf Situationen, in denen bereits Informationen vorliegen, die beurteilt werden sollen. Zu konkreten Ereignissen sollen daher andere Erklärungen gesucht werden oder nach einer Entscheidung werden unmittelbar Argumente betrachtet, die dagegen sprechen könnten. Die Strategie, gezielt Möglichkeiten und Annahmen in Betracht zu ziehen, die den eigenen Ansichten bzw. dem gegenwärtigen Standpunkt entgegengesetzt sind, werden auch als Technik des *consider the opposite* bezeichnet. Der Urteiler muss sich demnach zumindest zeitweise von seiner derzeitigen Ansicht lösen und die Datenbasis aus einer anderen Perspektive betrachten. Positive Effekte dieses Vorgehens werden aus zahlreichen Anwendungsfeldern berichtet.

Klauer (1988) erreichte mit dem *consider the opposite* Vorgehen zum Beispiel eine Reduktion des *Halo-Effekts*, indem er einen Teil seiner Probanden nach einer für alle Teilnehmer obligatorischen Einschätzung des politischen Verhaltens eines Abgeordneten bspw. mit der Aufgabe betraute, sich „so gut es geht in die Gedankengänge eines Menschen einzufühlen, der die entgegengesetzte politische Meinung vertritt“ (S. 603), um anschließend ein zweites Urteil abzugeben, wie eine Person es täte, die den Politiker genau konträr zur Probandensicht bewertet würde.

Auch der Ankereffekt lässt sich reduzieren, wenn zwischen der Gabe des Ankers und der eigenen Schätzung Argumente für einen (zu) hohen bzw. (zu) niedrigen Wert rekapituliert werden mussten (vgl. Mussweiler, Strack & Pfeiffer, 2000). Mumma und Wilson (1995) reduzierten den Einfluss einer eingangs vermittelten Information über ein

extremes Merkmal auf der Neurotizismus-Dimension, indem die Probanden in der *consider the opposite*-Bedingung vor der Urteilsbildung angeleitet wurden, an eine ihnen vertraute Person mit genau entgegengesetzten Eigenschaften zu denken, was der Aktivierung des gegensätzlichen Persönlichkeitsschemas diene.

Einen Beitrag unmittelbar zur Reduktion des *confirmation bias* leisten Lord, Lepper und Preston (1984) in einer Erweiterung der Studie von Lord et al. (1979), die die Meinungspolarisation am Beispiel der Todesstrafe gezeigt hat (s. auch Abschnitt 2.3.2.3). So wurde einerseits das Experiment von 1979 im Sinne einer Kontrollgruppe repliziert und mit zwei Experimentalgruppen kontrastiert. Eine Gruppe entsprach der *be unbiased* Manipulation und wurde angehalten, die beiden Studien so unvoreingenommen und objektiv wie möglich zu bewerten. Ihnen wurde dafür die Rolle des Richters zur Veranschaulichung vorgeschlagen. Außerdem wurden die Probanden vorgewarnt, dass die beiden zu lesenden Studien möglicherweise abweichende Ergebnisse liefern würden. In der dritten Gruppe wurden die Probanden über die Entstehung und Vermeidung von Urteilsfehlern aufgeklärt. Um Letzteres umsetzen zu können, sollten sie sich bei jedem abzugebenden Urteil fragen, ob sich ihr Urteil verändern würde, wenn die Methodikstudie genau das entgegengesetzte Ergebnis geliefert hätte. Die Probanden dieser *consider the opposite*-Bedingung mussten daher alternative Ausgänge bzw. Gegenargumente bedenken, bevor tatsächlich ein Urteil abgegeben werden sollte.

Im Einklang mit den Ergebnissen von Parmley (2006) unterschieden sich Kontrollgruppe und die *be unbiased*-Bedingung nicht von einander. Einzig im Falle des *consider the opposite* resultierten weniger einstellungskonsistente Bewertungen, d.h. eine geringer ausgeprägte Meinungspolarisation. Die Ergebnisse ließen sich zudem auf den Kontext des sozialen Hypothesentestens im Sinne von Snyder und Swann (1978) generalisieren. Bei Lord et al. (1984) testeten alle Probanden die Extraversionshypothese, während sie entweder der Kontrollgruppe (weitestgehend unter Replikationsbedingungen), dem *be unbiased* oder dem *consider the opposite* zugeteilt waren. Das *consider the opposite* wurde allerdings auf eine indirektere Weise umgesetzt, indem der scheinbar unzuverlässige Versuchsleiter sich entschuldigte, den Probanden lediglich das Introversion-Profil vorlegen zu können, da er das Extraversionsprofil nicht finden könne. Er halte dieses ersatzweise aber auch für hilfreich, denn schließlich sei Introversion das genaue Gegenteil von Extraversion. Der Testungsprozess der Extraversionshypothese war daher notwendigerweise von der Introversion beeinflusst, was im Ergebnis zu einer erwartungskonformen Reduktion der Präferenz für extraversionsbezogene Fragen bei Informationssuche führte. Die Generalisierung der Wirksamkeit auf ein zentrales Paradigma der *confirmation bias* Forschung war gelungen.

Es kann daher von einem relativ robusten Effekt ausgegangen werden, dessen Einsatz in weiteren Urteilsbereichen, so z.B. im Bereich der klinisch-psychologischen und (laien-)richterlichen Urteilsbildung, zumindest vielversprechend erscheint. Denkt man an die Diagnosestellung in der Psychologie bzw. Psychiatrie, so fällt jedoch unmittelbar ein Unterschied auf. Häufig geht es nicht so sehr um die Frage, ob eine Störung vorliegt oder nicht (dichotomes Ereignis, das zutreffen und nicht zutreffen kann), sondern vielmehr um die Abgrenzung, welches aus einer Reihe von Störungsbildern das Symptombild am besten beschreibt. Die Abgrenzung zur Normalität ist weniger problematisch als die Differenzialdiagnostik. Insofern ist es jedoch schwierig, von einem Gegenteil zu sprechen. Das Gegenteil einer Bulimie-Erkrankung ist natürlich nicht die Anorexie-Erkrankung, sie ist vielmehr eine Alternative.

Ob das Paradigma des *consider the opposite* im Bereich der Störungsdiagnostik tatsächlich hilfreich ist oder nicht vielmehr eine Instruktion im Sinne eines *consider an(y) alternative* angebracht ist, muss in Frage gestellt werden. Hirt und Markmann (1995) beschäftigten sich im Rahmen der Reduktion des *explanation bias* daher mit der Frage, ob es möglicherweise bereits ausreichend ist, irgendeine Alternative zu berücksichtigen und dadurch die Aktivierung des exakten Gegenteils als Voraussetzung zu umgehen, und fanden erste bestätigende Hinweise. Limitierend erwies sich jedoch die Plausibilität der Alternativen. Unter Berücksichtigung dieses Erfordernisses scheint ein korrekativer Einfluss des *consider the opposite* bzw. allgemeiner *consider an alternative* auch im Bereich der klinisch-psychologischen Diagnostik untersuchenswert.

Allen vorangegangenen Maßnahmen ist gemeinsam, dass sie in Form von einmaligen Instruktionen gegeben wurden und alles in allem eine Interventionsdauer von einigen Minuten bis maximal der Studiendauer aufwiesen. Ob sich hieraus Langzeiteffekte ergeben und ein Transfer auf andere Urteilsbereiche möglich ist, muss dahin gestellt bleiben. Diese Ziele stehen jedoch im Mittelpunkt der beiden nachfolgend vorzustellenden Ansätze.

Scientist-practitioner model. Spengler, Strohmer, Dixon und Shivy (1995) stellen in ihrem *scientist-practitioner model* den Grundgedanken in den Vordergrund, dass der praktisch arbeitende Diagnostiker genau dann weniger anfällig für fehlerhaftete Urteile ist, wenn er wissenschaftliche Methoden des Hypothesentestens in seiner täglichen Arbeit einsetzt. Dies steht im Einklang mit den Überlegungen, die von Trope und Liberman (1996) bzw. Hager und Weißmann (1991) ausgehend von der Verfasserin in Abschnitt 2.2.3 dargestellt wurden.

Spengler et al. (1995) greifen allerdings konkret den Ansatz von Pepinsky und Pepinsky (1954) auf und erweitern deren Modell des *counselor-as-scientist* um diejenigen sinnvoll erscheinenden Komponenten, die durch die Forschungsbemühungen der Zwischenzeit zu Tage getreten sind. In Anlehnung an das Modell der Pepinskys schlagen die Autoren vor, dass der diagnostische Prozess mit einer Beobachtung des Klienten beginnt (*observation*), die nach einem schlussfolgernden Prozess (*inference*) in einer Annahme resultiert. Letzteres ist im Folgenden in prüfbare Hypothesen zu übersetzen und zu testen (*hypothesis testing*). Aus einer Serie von Testungen entsteht sodann ein vorläufiges Arbeitsmodell über den Klienten (*tentative working client model*), auf dem Behandlungsentscheidungen aufgebaut und Vorhersagen gemacht werden können.

Allerdings halten Spengler et al. (1995) auch mehrere einstellungsbezogene Variablen im Sinne einer Grundhaltung für erforderlich, die eine wissenschaftliche Orientierung wiedergeben. So sollten die Diagnostiker sensibel für ihre eigenen Werte und Präferenzen sein, Voreingenommenheiten im Zusammenhang mit ihrer eigenen theoretischen Orientierungen reflektieren und sich eine Einstellung der Offenheit und Wissbegierde bewahren.

Außerdem explizieren sie den Prozess der Hypothesentestung, indem angeleitet wird, neben den *confirmatory hypotheses* auch *disconfirmatory* und *alternative hypotheses* zu testen, die wesentliche *debiasing techniques* darstellen. Die Autoren konzeptualisieren und begründen diese Schritte analog zu der oben von der Verfasserin skizzierten Argumentation.

Spengler et al. (1995) plädieren für eine kontinuierliche Vermittlung dieses Modells des Diagnostikers in der Aus- und Weiterbildung. Für zentral erachten sie, dass unmittelbar die Fertigkeit geschult wird, den Prozess der Modellbildung zu explizieren und den Diagnostiker in die Lage zu versetzen, sich in Form eines inneren Dialoges selbst zu prompten und wenn nötig zu korrigieren. Die pure Präsentation des Modells scheint ihnen nicht hinreichend. 2000 berichten Stoltenberg, Pace, Kashubeck-West, Biever, Patterson und Welch davon, dass die Position des *scientist-practitioner* zum Kern der Identität der Disziplin der *counseling psychology* geworden sei und allen akkreditieren Trainingsprogrammen zu einem nicht unerheblichen Teil zugrundeliege.

Dass die Konzeptionierung des Hypothesentestungsprozesses im Sinne des *scientist-practitioner models* im Einklang mit den Forschungen nach einer diagnostischen statt pseudodiagnostischen Vorgehensweise steht, die Alternativerklärungen berücksichtigt, ist unmittelbar einsichtig. Ihre Implementierung in die formalen Trainingsprogramme von *counseling psychologists*, die notwendigerweise ist einer bereichsspezifischen Auslegung

resultiert, alleine um den Lerneffekt zu erhöhen, macht dieses Konzept für andere Berufsgruppen, wie bspw. polizeiliche Ermittler unzugänglich. Vom Schemm (2008b) stellte in diesem Zusammenhang einen Mangel an Trainingsprogrammen für Experten fest, die erstens speziell auf den *confirmation bias* zugeschnitten sind, zweitens *consider the opposite*-Strategien implementieren und drittens auf andere Tätigkeitsbereiche übertragen werden können. Hieraus resultierte die Entwicklung des Trainingsprogramms MiA – *Möglich ist Alles*.

Möglich ist Alles. Das Trainingsprogram MiA (vom Schemm, 2008b) umfasst vier Unterrichtseinheiten. Das erste Modul hat einerseits die Demonstration der Bestätigungstendenz selbst zum Ziel, die anhand der Befragung einer instruierten Person veranschaulicht wird, und andererseits die eindrückliche Schilderung der potentiell vernichtenden Konsequenzen wie bspw. im Zusammenhang mit den Falscheschuldigungen wegen sexuellen Missbrauchs im so genannten *Montessori-Prozess*.

Eine zweite Einheit beschäftigt sich mit der Vermittlung des *Teufelsmodells*, welches in sprachlich vereinfachter Weise das Modell des konfirmatorischen Hypothesentestens nach Schulz-Hardt und Köhnken (2000) repräsentiert. Dieses wird an einem berufsspezifisch zu wählenden Beispiel veranschaulicht, bevor in der dritten Sitzung das *Engelsmodell* eingeführt wird, welches eine ergebnisoffene Grundhaltung (die den Wahlspruch *Möglich ist Alles* repräsentiert) und das unbedingte Einbeziehen von Alternativhypothesen beinhaltet.

Konkret werden vier einzelne Strategien vorgestellt (s. Tabelle 4), die in der vierten Unterrichtseinheit vertiefend unter Nutzung von Checklisten zur Organisation des Vorgehens eingeübt und abschließend diskutiert werden, um die Wahrscheinlichkeit eines Transfers in den Alltag zu erhöhen. Dies wird ebenfalls mit der Möglichkeit der optionalen Einheit angestrebt, die Raum für berufsspezifische Anwendungen bzw. Didaktik vorhandener Konzepte gibt.

In einer ersten Evaluationsstudie (Konitzky, 2009) wurde die Wirksamkeit des Trainings untersucht. Dafür wurde analysiert, ob die MiA-Teilnehmer ($n = 18$) anschließend einerseits über ein größeres Maß an Wissen über konfirmatorische Prozesse verfügen und andererseits seltener Bestätigungsmechanismen bzw. Urteilsverzerrungen aufweisen als Nicht-Teilnehmer ($n=18$). Die folgenden Annahmen konnten bestätigt werden: Die MiA-Teilnehmer verfügen über größeres Wissen, stellten mehr Evidenzregeln mit diagnostischen Merkmalen auf und weniger mit hypothesenkonsistenten Inhalten. Außerdem wiesen sie im Endurteil seltener einen *confirmation bias* auf. So lassen sich Hinweise darauf finden, dass die MiA-Teilnehmer eher eine günstige Teststrategie einsetzen. Replikationen

sind aber aus methodischen Gesichtspunkten (z.B. geringe Teilnehmerzahl) sinnvoll. Nichtsdestotrotz liegt mit diesem Training ein vielversprechendes und flexibel zu nutzendes Programm vor, dem ein Potential zur Vermeidung von Bestätigungstendenzen nicht abgesprochen werden kann.

Tabelle 4. Maßnahmen zur Erreichung einer ergebnisoffenen Grundhaltung und der Einbeziehung von Alternativhypothesen im Trainingsprogramm MiA (vom Schemm, 2008b)

<i>Maßnahme</i>	
<i>A</i>	Die Vermeidung des <i>confirmation bias</i> muss bereits bei der Aufstellung der Evidenzregeln sensu Trope und Liberman (1996) beginnen
<i>B</i>	Die Informationssuche muss durch die Beachtung der Diagnostizität der Merkmale charakterisiert sein
<i>C</i>	Das Stellen offener Fragen im Interview ist oberstes Gebot
<i>D</i>	Der Gesprächsstil soll dem ethischen Gesprächsstil (Shepherd, 1988) entsprechen

3 Ableitung der Fragestellung

Abstrakt formuliert beschäftigt sich die vorliegende Arbeit erstens mit der Frage, ob das Phänomen des *confirmation bias* in dem Bereich der klinisch-psychologischen Diagnostik mit einem verzerrten Informationsverarbeitungsmuster einhergeht, wie es einerseits das Modell des konfirmatorischen Hypothesentestens sensu Schulz-Hardt und Köhnken (2000) und andererseits dasjenige von Hager und Weißmann (1991) vorhersagen und welchen Einfluss die Berufserfahrung sowie das bereichsspezifische Wissen auf das Auftreten des *confirmation bias* ausüben. Im Folgenden soll die Ableitung der Fragestellung aus den angeführten empirischen Arbeiten begründet werden, um die zu manipulierenden, zu messenden und zu kontrollierenden Variablen zu rechtfertigen. Anschließend erfolgt die Darstellung der inhaltlichen Hypothesen.

3.1 Einflussvariablen

Aus den Erkenntnissen der referierten empirischen Arbeiten bzw. den aufgrund der angeführten Kritik offenbleibenden Forschungsfragen lassen sich drei Einflussvariablen im Zusammenhang mit dem *confirmation bias* ableiten, die für die vorliegende Arbeit zentral sind: erstens die *Überzeugtheit von der Prüfhypothese*, zweitens die *Berufserfahrung* und drittens das *Fachwissen*.

3.1.1 Überzeugtheit von der Prüfhypothese

Um die Auswirkungen einer Prüfhypothese auf den Prozess des Hypothesentestens und die Validität der Schlussfolgerung analysieren zu können, muss eine Situation konstruiert werden, in der eine Hypothese salient gemacht wird. Dies kann in Anlehnung an die Arbeiten um die Arbeitsgruppen um Snyder und Swann (z.B. Snyder & Cantor, 1979; Snyder & Swann, 1978; Swann & Guiliano, 1987) sowie Strohmer (vgl. Strohmer & Chiodo, 1984; Strohmer & Newman, 1983; Strohmer et al., 1990) durch eine direkte Vorgabe der Prüfhypothese realisiert werden. Im Einklang mit der Argumentation von Haverkamp (1993) ist allerdings zu kritisieren, dass nicht sichergestellt ist, ob die Probanden tatsächlich die vorgegebene Hypothese, ihr Gegenteil oder eine ganz andere Hypothese testen. Da Haverkamp (1993) in ihrer Arbeit (s. Abschnitt 2.3.3.2) einerseits das Testungsverhalten hinsichtlich der

induzierten Hypothese und andererseits dasjenige zur selbstgenerierten Hypothese erhob und Unterschiede feststellte, die v.a. ein konfirmatorisches Vorgehen in Bezug auf die selbstgenerierte Hypothese nahelegen, müssen diese Bedenken ernst genommen werden. Außerdem steht ihr Befund im Einklang mit den Überlegungen der Verfasserin zum intern-repräsentierten Anfangsverdacht, der eine engere Verknüpfung mit dem konfirmatorischen Hypothesentesten zu haben scheint als die induzierte Hypothese selbst. Daraus ist abzuleiten, dass keinesfalls unterstellt werden kann, dass Personen, denen eine Hypothese induziert wird, diese auch im selben Maße subjektiv übernehmen und verfolgen. In der vorliegenden Arbeit soll daher die subjektive Betrachtung im Mittelpunkt stehen, die bei dem Hypothesentester durch das Stimulusmaterial ausgelöst wird. Dies kann nun einerseits in einem qualitativen Sinne gemeint sein, d.h. eine Person ist dadurch gekennzeichnet, *welche* Hypothese sie generiert. Andererseits kann eine quantitative Betrachtung vorgenommen werden, d.h. *wie stark* der Hypothesentester von einer Hypothese überzeugt ist. Dem Vorschlag von Hager und Weißmann (1991) folgend und im Einklang mit der Variablen Hypothesenstärke, wie Gadenne und Oswald (1986) sie umsetzen, wird daher das quantitative Maß der *Überzeugtheit von der Prüfhypothese* als relevante Einflussvariable angesehen. Hager und Weißmann (1991) argumentieren, dass es einer Situation, in der man u.a. die Veränderung der subjektiven Sicherheit einer Person in eine Hypothese untersuchen will, sinnvoller Weise so vorgeht, dass das Stimulusmaterial in den Versuchsgruppen Hypothesen mit unterschiedlichem Maß an subjektiver Sicherheit auslöst.

Im Sinne der ökologischen Validität wird daher von außen eine Prüfhypothese im Sinne einer bestimmten Verdachtsdiagnose induziert werden, denn im klinisch-psychologischen Kontext erhält der Psychologe, Psychiater oder Psychotherapeut i.d.R. mit der Überweisung des Patienten auch eine Verdachtsdiagnose des überweisenden Arztes mitgeteilt. Relevant ist jedoch, die subjektive Überzeugtheit von der salient gemachten Prüfhypothese.

Entscheidend ist darüber hinaus, dass die Prüfhypothese bei objektiver Betrachtung aller zur Verfügung stehender Informationen *falsch* ist, d.h. dass die Verdachtsdiagnose objektiv nicht zutrifft und die Bestätigung dieser Annahme inkorrekt ist. Wäre die Prüfhypothese zutreffend, ergäben sich keine Fehlentscheidungen und man könnte nicht von Verzerrungen sprechen, geschweige denn Verzerrungsmechanismen untersuchen.

3.1.2 *Berufserfahrung und Fachwissen*

In der vorliegenden Arbeit soll darüber hinaus der Zusammenhang zwischen dem *confirmation bias*, der Berufserfahrung und dem Fachwissen untersucht werden. Dies begründet sich einerseits aus der Tatsache, dass der von Spengler et al. (2009) berichtete kleine Effekt zu Gunsten der Erfahrenen auf die allgemeine Urteilsgüte und nicht speziell den *confirmation bias* bezogen war. Andererseits weisen die einschlägigen Studien in der Regel methodische Mängel auf, z.B. waren die Probanden zu homogen in ihrer Berufserfahrung (vgl. Haverkamp, 1993) oder wiesen eine nicht allzu große Differenz der Berufserfahrung auf (vgl. Martin, 2000) oder die untersuchten Stichproben waren zu klein (Krems & Zierer, 1994). Da aber gerade die Grundidee einer Differenzierung in Berufserfahrung vs. Fachwissen vielversprechend ist, soll das Design von Krems und Zierer (1994) im klinisch-psychologischen Bereich repliziert werden.

Daher wird einerseits die Einflussvariable *Berufserfahrung* mit den Ausprägungen *Erfahrene vs. Unerfahrene* über die Dauer ihrer praktisch-diagnostischen Tätigkeit realisiert. Andererseits ist ein Urteilskontext zu wählen, der es einer Personengruppe ermöglicht, ihr Fachwissen einzusetzen, während eine andere Gruppe daran gehindert wird (*Einflussvariable Fachwissen: relevant vs. irrelevant*).

3.2 *Abhängige Variablen*

Die zu messenden Variablen sind einerseits aus dem Modell des konfirmatorischen Hypothesentestens von Schulz-Hardt und Köhnken (2000) abgeleitet und beziehen sich andererseits auf die differenzierten Bestätigungstendenzen nach Hager und Weißmann (1991). Sie werden nach den beiden Modellvorstellungen getrennt erläutert.

3.2.1 *AVn nach dem Modell des konfirmatorischen Hypothesentestens*

Neben der *Schlussfolgerung* bzgl. der Prüfhypothese werden Maße für die *hypothesenkonsistente Interpretation uneindeutiger Evidenz* sowie für die *positive und pseudodiagnostische Informationssuche* erhoben, was im Folgenden zu begründen ist. Abschließend wird Stellung dazu genommen, warum die anderen Variablen des Modells nicht betrachtet werden.

Schlussfolgerung. Die Schlussfolgerung stellt im konkreten Anwendungsfall der klinisch-psychologischen Diagnostik die Diagnosestellung dar. Im Gegensatz zu den sozialpsychologischen Arbeiten (z.B. Snyder & Cantor, 1979; Snyder & Swann, 1978), die i.d.R. nur einen Ausschnitt des Hypothesentestungsprozesses untersuchten, indem bspw. zwar Fragen ausgewählt, aber nicht gestellt und dementsprechend nicht beantwortet wurden und daher kein abschließendes Urteil der Probanden hinsichtlich der Extraversion der Zielperson ermöglichen, soll in der vorliegenden Arbeit der gesamte Prozess abgebildet werden. Damit ist gemeint, dass alle Schritte im Ablauf des Testens von Hypothesen nach Trope und Liberman (1996) sowie Hager und Weißmann (1991) und analog dem diagnostischen Prozess Berücksichtigung finden. Nach Wissen der Verfasserin wird somit erstmalig auch das Aufstellen von Evidenzregeln (s.u.) sensu Trope und Liberman (1996) untersucht.

Es werden daher alle notwendigen Informationen präsentiert, um die Diagnose vor dem Hintergrund des *Diagnostischen und Statistischen Handbuchs Psychischer Störungen* in 4. Auflage (DSM-IV; Saß et al., 1998) stellen zu können, was der *Schlussfolgerung* (Trope & Liberman, 1996) bzw. der *Prüfung der Hypothese aufgrund der Informationen* (Hager & Weißmann, 1991) bzw. der *Beantwortung der Fragestellung* (Jäger, 1982) entspricht.

Hypothesenkonsistente Interpretation uneindeutiger Evidenz. Die empirischen Hinweise, dass mehrdeutige Informationen bevorzugt im Sinne der zu testenden Hypothese interpretiert werden, sind in der Grundlagenforschung zahlreich (vgl. Darley & Gross, 1983; Evett et al., 1994; Lord et al., 1979; Snyder & Swann, 1978). Außerdem liegt eine Arbeit von Strohmer et al. (1990; s. Abschnitt 2.3.2.3) für die klinisch-psychologische Diagnostik vor. Auch hier wurde allerdings isoliert untersucht, ob sich die erfahrenen Probanden eher auf hypothesenkonsistente Informationen fokussieren, wenn sowohl diese als auch hypothesensinkonsistente vorliegen, was in der Summe eine uneindeutige Situation darstellt. Eine Schlussfolgerung in Form einer Diagnosestellung wurde nicht erhoben.

Deswegen soll in der vorliegenden Arbeit die hypothesenkonsistente Interpretation uneindeutiger Evidenz als Repräsentant für die Prozessstufen *Identifizierung der Evidenz* (Trope & Liberman, 1996), *Bewertung und Interpretation der Informationen zur Eignung* (Hager & Weißmann, 1991) sowie der *Datenauswertung* (Jäger, 1982) erhoben werden. Konkret wird dabei die Aufwertung und Umdeutung der Aussagekraft von einzelnen Informationen betrachtet, d.h. eine abweichende Beurteilung vorgelegter Informationen, die für hypothesenkonsistente Informationen zu einer größeren Gewichtung führt.

Konfirmatorische Teststrategie. Die Informationssuche, die mit verschiedenen Teststrategien verbunden sein kann, stellt das Herzstück des diagnostischen Prozesses dar. Qualitätsverluste durch ungünstiges Suchverhalten durch den Einsatz einer positiven und zugleich pseudodiagnostischen Teststrategie können auch durch eine noch so sorgfältige Interpretation nicht ausgeglichen werden. Aufgrund der Wichtigkeit der Informationssuche ist dieser Aspekt empirisch schon gut untersucht (s. Abschnitt 2.3.3.2). Für das Auswählen und Generieren von Fragen im Interviewkontext, der sicherlich eine große Bedeutung für die Hypothesentestung hat, liegen demnach umfassende Erkenntnisse vor. Haverkamp (1993) weist allerdings zu Recht daraufhin, dass noch Forschungsbedarf für die Informationssuche jenseits der reinen Frageformulierung besteht. Auch nach Ansicht der Verfasserin ist der Aspekt der Teststrategie nicht mit der Informationssuche alleine verknüpft. Genauso gut kann sich die Tendenz, eher konsistente als inkonsistente Informationen zu suchen, darin manifestieren, dass in den aufgestellten Evidenzregeln mehrheitlich Merkmale berücksichtigt werden, die von der Prüfhypothese vorhergesagt werden. Entsprechend könnte die pseudodiagnostische Teststrategie sich manifestieren, indem nur Evidenzregeln zur Prüfhypothese aufgestellt werden und Alternativerklärungen unberücksichtigt bleiben.

In der vorliegenden Arbeit soll der Einsatz der Teststrategien daher in Bezug auf die Evidenzregeln nach Trope und Liberman (1996) untersucht werden, indem deren Güte bestimmt wird, welche sich darin ausdrückt, dass auch Evidenzregeln zu prüfhypotheseninkonsistenten Inhalten formuliert werden, die zudem optimaler Weise diagnostisch sind.

Aufgrund dieser Ausrichtung ist die Verknüpfung mit dem Prozess des sozialen Hypothesentestens nach Trope und Liberman (1996) definitionsgemäß am engsten. Wie oben argumentiert korrespondiert dieser Akt jedoch auch in nicht unerheblichem Maße mit der *Ermittlung von Prüfinstanzen sowie deren Auswahl und Erzeugung* (Hager & Weißmann, 1991) bzw. der *Untersuchungsplanung und Untersuchungsdurchführung* (Jäger, 1982).

Nicht umgesetzt. Nach dem Modell des konfirmatorischen Hypothesentestens (Schulz-Hardt & Köhnken, 2000) bleiben die drei kognitiven Mechanismen *Überschätzung der a priori Wahrscheinlichkeit*, die *selektive Speicherung bzw. der selektive Abruf hypotheseinkonsistenter Informationen* sowie *affirmation bias* und der motivationale Aspekt der *asymmetrischen Schwellenwahl* unberücksichtigt.

Nach Ansicht der Verfasserin ist der Ausschluss des Verzerrungsmechanismus *Überschätzung der a priori Wahrscheinlichkeit der Hypothese* in dem gewählten Kontext der klinisch-psychologischen Diagnostik begründet. Die a priori Wahrscheinlichkeit

bezieht sich hier auf die Prävalenzrate der Störung, die als Verdachtsdiagnose im Rahmen der Prüfhypothese vorgegeben wird. Aufgrund der Tatsache, dass die epidemiologischen Studien regelmäßig Abschätzungen der Prävalenzraten der psychischen Störungen liefern, ist davon auszugehen, dass Schätzungsprozesse in diesem spezifischen Kontext keine große Rolle spielen. Eine unbewusste Überschätzung ist daher in dem professionellen Kontext nicht anzunehmen. Deswegen wird auf die Untersuchung verzichtet.

Desweiteren werden keine Maße hinsichtlich selektiver Gedächtniseffekte erhoben, da nach Ansicht der Verfasserin aufgrund der Befunde von u.a. Arkes und Harkness (1980) zum Effekt der Diagnosestellung auf die spätere Wiedererkennensleistung der Diagnostiker schon entsprechende Hinweis aus einem ganzheitlichen Kontext vorliegen. Außerdem wäre die Umsetzung einer entsprechenden Variable im beabsichtigten Kontext einer Onlinestudie (s. Abschnitt 4.4) mit erheblichen methodischen Problemen verbunden gewesen, die nach Einschätzung der Verfasserin eine Dateninterpretation nicht erlaubt hätten.

Auf die Erfassung des *affirmation bias* in der vorliegenden Studie wurde verzichtet, da aufgrund der Ergebnisse der eigenen Vorarbeiten (Dreger, 2006; Hagemann, 2008; Limbrecht, 2008) die Vermutung naheliegt, dass es sich bei dem *affirmation bias* um ein solch allgemeines Phänomen handelt, dass nahezu jede Person dieser Tendenz unterliegt und man nicht unbedingt annehmen muss, dass die Tendenz z.B. von der *Überzeugtheit der Prüfhypothese* noch moderiert wird. Möglicherweise liegt demnach eine Art Deckeneffekt vor, der den Nachweis der Abhängigkeit erschwert. Aufgrund der Komplexität die mit dem Versuch der Abbildung des gesamten Hypothesentestungsprozesses einhergeht, wurde auf gezielte Anstrengungen zur Vermeidung eines solchen Deckeneffekts verzichtet, weil die negativen Konsequenzen bei einer kognitiven Überforderung der Probanden für alle AVn schwerer gewichtet wurde als der Erkenntnisverlust bzgl. des *affirmation bias*, der sich bereits als sehr robuste Tendenz gezeigt hat.

Die Motivation für die Außerachtlassung der *asymmetrischen Schwellenwahl* als motivationalem Prozess des Hypothesentestens ist wiederum mit dem Kontext der klinisch-psychologischen Diagnostik verknüpft. Zum einen ist es Teil der psychologischen Ausbildung, sowohl auf die Vermeidung des Fehlers erster als auch zweiter Art zu achten, und die damit verbundenen Konsequenzen sind zudem durchaus salient, so z.B. die Mortalitätsraten bei unbehandelten Anorektikern. Es ist daher davon auszugehen, dass die Abwägungen auf der Ebene einer bewussten Verarbeitung passieren. Schulz-Hardt und Köhnken (2000) beziehen sich jedoch vornehmlich auf die unbewussten Prozesse. Es ist daher in Frage zu stellen, ob ein entsprechender Effekt überhaupt in der klinisch-psychologischen Diagnostik zu unterstellen ist. Ähnlich wie beim *affirmation bias* wurde zugunsten der stärker begründeten Variablen auf die Erhebung des motivationalen Aspekts verzichtet.

3.2.2 AVn bzgl. der Bestätigungstendenzen nach Hager & Weißmann (1991)

Nach Kenntnisstand der Verfasserin sind die Bestätigungstendenzen nach Hager und Weißmann (1991) noch nicht im Zusammenhang mit dem *confirmation bias* im Bereich klinisch-psychologischer Diagnostik untersucht worden. Die Arbeitsgruppe um Spengler und Strohmmer dürfte hiervon weitgehend unabhängig gearbeitet haben, da dort (wie in den anderen im Rahmen dieser Arbeit referierten Quellen) keine Verweise auf die Arbeit von Hager und Weißmann (1991) gefunden werden konnten, was womöglich an deren deutschsprachiger Publikation liegt.

Schlussfolgerung. Auch Hager und Weißmann (1991) schlagen vor, festzustellen, inwieweit die Probanden die Prüfhypothesen am Ende des Testungsprozesses als bestätigt ansehen. Im Sinne der *Schlussfolgerung* wird daher die Einschätzung des Zutreffens der Hypothese erfasst, was der Variable *Schlussfolgerung* nach dem Modell des konfirmatorischen Hypothesentestens nach Schulz-Hardt und Köhnken (2000) entspricht. Alle Anmerkungen gelten daher analog.

Subjektive Sicherheit. Hager und Weißmann (1991) betrachten den Prozess des Testens einer Hypothese immer eingebettet in einen größeren Kontext und gehen daher davon aus, dass es bereits vor dem Hypothesentestungsprozess ein gewisses Maß an Vertrauen in die Hypothese gab. Nach Abschluss der konkreten Überprüfung ist dieses in Form der subjektiven Sicherheit bzgl. des Zutreffens der Hypothese zu messen.

Dieser Aspekt entspricht am wenigsten einem Schritt im Testen sozialer Hypothesen bzw. im diagnostischen Prozess und stellt vielmehr ein Spezifikum nach Hager und Weißmann (1991) dar (s. auch Abschnitt 2.2.3).

Bestätigungstendenz B. Um die Bestätigungstendenz B, d.h. die Bevorzugung von konsistenten gegenüber inkonsistenten Informationen im Bewertungsprozess, untersuchen zu können, müssen nach Hager und Weißmann (1991) entsprechende Informationen vorgelegt und seitens der Probanden eingeschätzt werden. Einerseits geschieht dies hinsichtlich der *Wichtigkeit der Informationen* für die Bewertung des Falls und andererseits im Hinblick auf die *Veränderung der subjektiven Sicherheit* in Bezug auf das Zutreffen der Hypothese bei isolierter Betrachtung der Information. Es wird von einer Aufwertung gesprochen, wenn die Wichtigkeit bzw. Änderung der Sicherheit zugunsten der konsistenten Items ausfällt.

Es handelt sich hierbei um einen Bewertungsprozess, der analog zur *hypothesen-konsistenten Interpretation uneindeutiger Evidenz* als Repräsentant für die Prozessstufen *Identifizierung der Evidenz* (Trope & Liberman, 1996), *Bewertung und Interpretation der Informationen zur Eignung* (Hager & Weißmann, 1991) sowie der *Datenauswertung* (Jäger, 1982) zu interpretieren ist.

Bestätigungstendenz I (Uminterpretation). Letztgenanntes gilt auch für die Bestätigungstendenz I, die sich im Sinne einer Uminterpretation darauf bezieht, dass Widersprüche nicht als solche erkannt oder benannt werden. Das Ausmaß einer solchen Selektion bei der Vorgabe inkonsistenter Informationen ist daher zu erheben.

Nicht umgesetzt. Die Bestätigungstendenz A wird in der vorliegenden Arbeit nicht erhoben. Nach Hager und Weißmann (1991) sind ganz spezielle Anwendungsvoraussetzungen zu beachten, um ein konkretes Testverhalten als Bestätigungstendenz A zu klassifizieren. Um einem Hypothesentester das gezielte Aufsuchen von hypothesenbestätigenden Prüfinstanzen nachzuweisen, muss sichergestellt sein, dass derjenige nicht nur hypothesenkonsistente Information sucht (wie im Falle der positiven Teststrategie), sondern dass er bereits vorab mit einer hinreichenden Sicherheit davon ausgehen kann, dass er tatsächlich eine Bestätigung finden wird. Als potentiell Beispiel für die Bestätigungstendenz A nennen Hager und Weißmann (1991) ein bewusst eingesetztes suggestives Fragemuster, weil aufgrund des suggestiven Charakters angenommen werden muss, dass entsprechende Fragen eine bestätigende Antwort liefern. Da jedoch im Zusammenhang mit der Variablen *konfirmatorische Teststrategie* argumentiert wurde, dass in der vorliegenden Arbeit nicht auf den Prozess der Informationssuche fokussiert werden soll, die durch eine Interviewsequenz operationalisiert werden könnte, entfällt diese Option.

3.2.3 AV Verhaltenskonsequenz in Anlehnung an Haverkamp (1993)

Im Zusammenhang mit dem Einsatz der positiven Teststrategie in der klinisch-psychologischen Diagnostik (s. Abschnitt 2.3.3.2) wurde bereits auf die Kritik von Haverkamp (1993) hinsichtlich der nahezu ausschließlichen Fokussierung auf die Frageformulierung bei der Informationssuche hingewiesen. In ihrer eigenen Studie ließ sie ihre Probanden daher nicht nur Fragen generieren, sondern wie bereits erwähnt auch verbale und nonverbale Reize seitens des Klienten notieren bzw. eine Hausaufgabe formulieren.

Auch aus Sicht der Verfasserin scheint es interessant, ob die Verzerrungen in der Informationsverarbeitung sich neben dem Handeln im diagnostischen Prozess – wie bspw. der Frageformulierung im Interview – auch im darüber hinausgehenden Verhalten manifestieren. Denn streng genommen hat eine falsche Bestätigung einer Prüfhypothese nur dann tatsächlich Konsequenzen, wenn dementsprechend gehandelt wird, d.h. indem daraus resultierend bspw. die falschen Therapiemaßnahmen ergriffen werden oder indem eine falsche Anklage erhoben wird. Die Variable *Verhaltenskonsequenz* soll daher den Umgang mit einem Patienten nach der Diagnosestellung abbilden und beinhaltet die Integration von Elementen in die Therapie (z.B. in Form von Gesprächsinhalten oder Hausaufgaben), die für die tatsächlich vorliegende Störung gar nicht indiziert sind, sondern ihre Berechtigung v.a. bei Gültigkeit der Prüfhypothese haben.

3.3 Störvariablen

In jeder Studiensituation gibt es Variablen, die die Messung des Einflusses der bedingenden Variablen auf die bedingten stören, weil sie ihn entweder verschleiern oder vortäuschen, was bestenfalls auszuschließen, minimal aber zu kontrollieren ist. In der Folge sollen insbesondere diejenigen Variablen erläutert werden, die aus inhaltlichen Gesichtspunkten einen störenden Einfluss ausüben könnten. Diejenigen störenden Einflüsse, die sich hingegen aus der konkreten Umsetzung der Studie ergeben und daher methodischen Charakter haben, werden im Rahmen der Operationalisierung jeweils dort diskutiert, wo sie sich aufgrund der getroffenen Entscheidungen manifestieren.

Bewusstsein hinsichtlich konfirmatorischer Prozesse. Ein störender Einfluss bei der Untersuchung des Einflusses der *Überzeugtheit von der Prüfhypothese* auf den Prüfungsprozess ist anzunehmen, wenn die Hypothesentester sich über die untersuchten konfirmatorischen Prozesse bewusst sind, da sie das Modell des konfirmatorischen Hypothesentests (Schulz-Hardt & Köhnken, 2000) oder die Bestätigungstendenzen nach Hager und Weißmann (1991) kennen und in der konkreten Untersuchungssituation aktiv versuchen, diese zu vermeiden. Das Verhalten der betreffenden Person, die um die geschilderten Gefahren weiß, könnte durch Vorsicht und Einsicht beeinflusst werden. Selbstverständlich wäre es wünschenswert, dass das Wissen um die Gefahr der potentiell konfirmatorischen Auswirkungen einer Prüfhypothese ihnen entgegenwirkt, für die Untersuchung der Urteilsverzerrungen ist dies jedoch hinderlich. Es ist daher zu vermeiden, dass ein Zusammenhang zwischen dem theoretischen Hintergrund der Arbeit und der konkreten Versuchsanordnung gesehen werden kann.

Der diagnostische Prozess als soziale Interaktion. Die empirische Forschung zum Einfluss der sozialen Kognition im diagnostischen Prozess nahm ihren Ursprung, als sich die Erkenntnis durchsetzte, dass Beratung und Therapie (mit inbegriffener Diagnostik) als interpersoneller Einflussprozess zwischen Berater/Therapeut und Klient zu verstehen ist (Strong, 1968). Der *confirmation bias* ist in diesem Kontext selbstverständlich nicht die einzige Einflusskomponente. So spielen auch Faktoren innerhalb der Person des Diagnostikers eine Rolle, die in der sozialen Interaktion wirksam werden. Dies betrifft bspw. dessen kognitive Komplexität (Garb & Lutz, 2001; Spengler & Strohmmer, 1994; Walker & Spengler, 1995) sowie Geschlecht und ethnische Zugehörigkeit (z.B. Pottick et al., 2007), um nur einige wenige zu nennen. Auch der Sympathie zwischen Berater/Therapeut und Klient sowie deren Ähnlichkeit ist ein (im Rahmen dieser Studie) störender Einfluss zuzuschreiben. Einige Autoren von Untersuchungen zum sozialen Hypothesentesten in einer konkreten Interviewsituation (vgl. Barnikol, 2004; Burwitz, 2005; Modes, 2005) weisen ausdrücklich auf diese Problematik hin. Zur Wahrung der internen Validität ist daher eine weitestgehende Ausschaltung der sozialen Interaktion erforderlich, was v.a. deshalb vertretbar scheint, weil gerade die Forschungsbemühungen von der Arbeitsgruppe um Strohmmer zeigen, dass sich die Effekte, die in Laborstudien oder bei postalischen Befragungen erzielt werden, auch auf realere Situationen mit (instruierten) Patienten übertragen lassen (Strohmmer & Shivy, 1994).

3.4 Inhaltliche Hypothesen

Im Folgenden werden die aus dem theoretischen und empirischen Hintergrund abgeleiteten psychologischen Hypothesen (PH) zum Auftreten des *confirmation bias* in der klinisch-psychologischen Diagnostik expliziert. Dies erfolgt differenziert nach den bedingten Variablen. Es resultiert daher ein Hypothesenkomplex A_i bzgl. des *verzerrten Urteils*, Komplex B_i im Zusammenhang mit *asymmetrischen Evidenzregeln*, Hypothese C bzgl. der *hypothesenkonsistenten Interpretation uneindeutiger Evidenz*, Hypothese D bezogen auf die *Bestätigungstendenz I*, der Komplex E_i im Zusammenhang mit der *Bestätigungstendenz B* und schlussendlich die Hypothese F für die *Verhaltenskonsequenzen*.

Überzeugtheit von der Prüfhypothese. Alle Hypothesen bzgl. dieser Variablen unterstellen eine zunehmende Verzerrung des Testungsprozesses und eine steigende Wahrscheinlichkeit für ein selbstbestätigendes Urteil bei zunehmender Überzeugtheit von der Prüfhypothese.

Berufserfahrung und Fachwissen. Aufgrund der konzeptionellen Ähnlichkeit zur Arbeit von Krems und Zierer (1994) wird einerseits erwartet, dass erfahrene Personen sich im Hinblick auf die Häufigkeit eines konfirmatorisch verzerrten Urteils und der verzerrten Informationsverarbeitung nicht von unerfahrenen Personen unterscheiden, während Personen mit relevantem Fachwissen diesen Tendenzen seltener unterliegen als Personen ohne relevantes Fachwissen.

Im Rahmen von Tabelle 5 (siehe folgende Seite) ist nun für jede oben abgeleitete Variable die ungünstige Ausprägung beschrieben und nach Themenkomplexen geordnet entweder dem Modell des konfirmatorischen Hypothesentestens nach Schulz-Hardt und Köhnken (2000) oder dem Konzept von Hager und Weißmann (1991) zugeschrieben. Hypothesenkomplexe X_{1a} beziehen sich auf den Einfluss der *Überzeugtheit der Prüfhypothese* und Komplexe X_{1c} auf denjenigen des *Fachwissens*. Für die analogen Hypothesen der X_{1b} im Zusammenhang mit der *Berufserfahrung* wurde auf eine differenzierte Darstellung verzichtet und hypothesenübergreifend angegeben, dass keine Unterschiede erwartet werden.

Hypothese F. Die Hypothese F besagt, dass *die Wahrscheinlichkeit, nicht indizierte Elemente in die Therapie zu integrieren mit steigender Überzeugtheit von der Prüfhypothese zunimmt* und bezieht sich auf die in Anlehnung an Haverkamp (1993) abgeleiteten Verhaltenskonsequenzen in Abhängigkeit vom *confirmation bias*. Sie ist nicht in Tabelle 5 integriert, da sie anders als alle anderen bedingten Variablen ausschließlich in Abhängigkeit der *Überzeugtheit der Prüfhypothese* untersucht wird. Eine Berücksichtigung in Bezug auf die Variable *Berufserfahrung* scheint trivial, da die Probanden sich gerade im Wissen um die Indikation bestimmter Maßnahmen bei einzelnen Störungen unterscheiden. Eine Interpretation bei *irrelevantem Fachwissen* verbietet sich sogar aus sachlogischen Gründen.

Tabelle 5. Psychologische Hypothesenkomplexe zur Überzeugtheit von der Prüfhypothese, der Berufserfahrung und dem Fachwissen

Hypothesenkomplex	Je höher die subjektive Überzeugtheit von der Prüfhypothese, desto [spezifischer Aspekt der Hypothese]	Für alle PH-X _{ib} :	Personen mit relevantem Fachwissen ... [spezifischer Aspekt der Hypothese] als Personen, deren Fachwissen irrelevant ist.
PH-A _i Verzerrtes Urteil	A _{1a} ... höher die Wahrscheinlichkeit einer Selbstbestätigung der Prüfhypothese. A _{2a} ... höher die subjektive Sicherheit.	Personen mit Berufserfahrung unterscheiden sich nicht von Personen ohne Berufserfahrung.	A _{1e} ... bestätigen seltener die Prüfhypothese ... A _{2e} ... sind sich im Urteilen subjektiv weniger sicherer ...
PH-B _i Asymmetrische Evidenzregeln	B _{1a} ... größer die Präferenz für positive Evidenzregeln. B _{2a} ... größer die Präferenz für Evidenzregeln mit nicht diagnostischen Inhalten B _{3a} ... größer die Präferenz für positive Evidenzregeln mit nicht diagnostischen Inhalten.		B _{1e} ... zeigen eine geringere Präferenz für positive Evidenzregeln ... B _{2e} ... zeigen eine geringere Präferenz für Evidenzregeln mit nicht diagnostischen Inhalten B _{3e} ... zeigen eine geringere Präferenz für positive Evidenzregeln mit nicht diagnostischen Inhalten
PH-C: Hypothesen-konsistente Interpretation	B _{4a} ... weniger Alternativhypothesen werden in den Evidenzregeln berücksichtigt. B _{5a} ... größer die Präferenz für prüfhypothesenbezogene Evidenzregeln.		B _{4e} ... berücksichtigen in ihren Evidenzregeln mehr alternative Hypothesen ... B _{5e} ... zeigen eine geringere Präferenz für prüfhypothesenbezogene Evidenzregeln ...
PH-D: Bestätigungstendenz I	C _a ... ausgeprägter ist die hypothesenkonsistente Interpretation uneindeutiger Informationen. D _a ... geringer ist die Erkennensleistung für hypotheseninkonsistente Informationen.		C _e ... interpretieren uneindeutige Informationen weniger stark hypothesenkonsistent ... D _e ... erkennen hypotheseninkonsistente Informationen häufiger...
PH-E: Bestätigungstendenz B	E _{1a} ... größer die Wichtigkeit konsistenter gegenüber inkonsistenter Informationen. E _{2a} ... mehr übersteigt die Bekräftigung der Prüfhypothese durch konsistente Angaben die Entkräftigung durch inkonsistente.		E _{1e} ... schätzen die Wichtigkeit konsistenter Informationen in Relation zu inkonsistenten weniger hoch ein ... E _{2e} ... schätzen die Bekräftigung der Prüfhypothese durch konsistente Angaben in Relation zur Entkräftigung seitens der inkonsistenten weniger hoch ein ...

4 Operationalisierung

Im vorliegenden Kapitel wird die Umsetzung der zu manipulierenden, zu messenden und zu kontrollierenden Variablen beschrieben. Außerdem wird die Realisierung als Online-Studie begründet. Bevor die Umsetzung der Variablen im Einzelnen erläutert wird, soll zur Erleichterung der Einordnung der Details kurz ein grober Überblick über die Studie vermittelt werden: Angelehnt an das Studienkonzept von Krems und Zierer (1994) werden sowohl erfahrenen als auch unerfahrenen Diagnostikern speziell konstruierte Fallinformationen präsentiert, zu deren Beurteilung in Form einer Online-Studie sie entweder auf ihr spezifisches Fachwissen zurückgreifen können oder nicht.

Zu Beginn erhalten alle Probanden eine Fallbeschreibung, die sich nach der Zugehörigkeit zur Bedingung *Fachwissen* unterscheidet: die vorgelegte Fallgeschichte in der Gruppe *Fachwissen relevant* bezieht sich auf eine jugendliche Patientin mit dem Verdacht auf das Vorliegen einer Bulimie (zur Konstruktion s. Abschnitt 4.1.3.1). In der Bedingung *irrelevant* wird den Probanden eine Fallgeschichte bzgl. einer mutmaßlichen schweren Brandstiftung mit Todesfolge vorgelegt (s. Abschnitt 4.1.3.2). Durch die Gestaltung der Fallbeschreibung wird die jeweilige Prüfhypothese (Bulimie vs. schuldig) vermittelt und anschließend die subjektive Überzeugtheit erhoben.

Im Anschluss daran formulieren die Probanden Evidenzregeln im Sinne von Trope und Liberman (1996) und bewerten weitere Fallinformationen, welches der Erhebung der AVn zur Informationsverarbeitung dient. Nach der Vorgabe aller relevanten Informationen, die in irrelevante eingebettet sind, fällen alle Probanden ein abschließendes Urteil, d.h. stellen in der Bedingung *Fachwissen relevant* (*Essstörung*) eine Diagnose und erklären den Angeklagten in der *Bedingung irrelevant* (*Brandstiftung*) für schuldig oder unschuldig.

4.1 Einflussvariablen

Es werden die drei Einflussvariablen *Überzeugtheit von der Prüfhypothese*, *Berufserfahrung* und *Fachwissen* berücksichtigt.

4.1.1 Überzeugtheit von der Prüfhypothese

Für die Erzeugung der Variablen *Überzeugtheit von der Prüfhypothese* soll der natürliche Unterschied in der Stärke, mit der die vorgegebene Hypothese übernommen wird, genutzt werden. Demnach liegt ein quasi-experimentelles Vorgehen vor, da keine Zuteilung auf

Versuchsbedingungen vorgenommen wird, sondern die subjektive Überzeugung für die Einordnung zu verwenden ist. Aus inhaltlichen Gründen ist dieses Vorgehen dem rein methodisch überlegenen experimentellen Ansatz vorzuziehen (s. Abschnitt 3.1.1).

Nach dem Lesen der Fallbeschreibung, die mit dem Hinweis einer statushohen Person auf eine konkrete Prüfhypothese verbunden ist, geben die Probanden auf einer siebenstufigen Skala von *0=sehr unwahrscheinlich* und *6=sehr wahrscheinlich* an, für wie wahrscheinlich sie die von der statushohen Person vermutete Annahme halten. In der *Bedingung Fachwissen relevant* geben die Probanden daher an, für wie wahrscheinlich sie die Verdachtsdiagnose Bulimie seitens des überweisenden Hausarztes halten. Dies findet in der *Bedingung Fachwissen irrelevant* seine Entsprechung in dem Schuldvorwurf seitens der Staatsanwaltschaft.

Um keinen Informationsverlust zu erleiden, wird keine (künstliche) Dichotomisierung vorgenommen, sondern über die Wahl geeigneter Auswertungsverfahren die gesamte Varianz in der Variablen als intervallskaliert angenommener Variable genutzt.

4.1.2 Berufserfahrung

Um den Zusammenhang des *confirmation bias* mit der Berufserfahrung zu untersuchen, werden erfahrene Diagnostiker (*Erfahrene*) mit unerfahrenen Diagnostikern (*Unerfahrene*) verglichen. Da es sich bei der Fallbeschreibung in der *Bedingung Fachwissen relevant* um eine klinisch-psychologische Fragestellung handeln wird, bei der eine Störungsdiagnostik im Mittelpunkt steht, ist v.a. die Erfahrung im klinischen bzw. psychotherapeutischen Bereich von Bedeutung, da Symptome erkannt und vor dem Hintergrund verschiedener alternativer Störungen bewertet werden müssen.

Als *Erfahrene* werden daher diejenigen Probanden angesehen, die über eine mindestens dreijährige Erfahrung im klinischen Bereich verfügen, was bspw. durch eine Approbation zum Psychologischen Psychotherapeuten bzw. Kinder- und Jugendlichenpsychotherapeuten oder eine Facharztausbildung im Bereich Psychiatrie angezeigt wird. Neben dem Psychologie- bzw. Medizinstudium verfügen diese Personen aufgrund der Rahmenbedingungen ihrer Zusatzausbildungen über mindestens drei Jahre Berufserfahrung in der diagnostischen wie therapeutischen Arbeit mit Menschen mit psychischen Störungen.

Unerfahrene sind hingegen diejenigen Probanden, die zum Zeitpunkt der Studienteilnahme noch Psychologie studieren. Mit Ausnahme von Praktika, die jedoch höchstens einige Wochen umfassen und nicht notwendigerweise in einer Klinik absolviert

werden müssen, verfügen diese Personen über keine Berufserfahrung im Sinne einer eigenständigen Diagnostik realer Personen.

Durch diese Operationalisierung wird einerseits sichergestellt, dass die Gruppe der Erfahrenen nicht aus noch zu trainierenden Personen besteht, da sie alle aufgrund der Approbation die staatliche Zulassung zur Ausübung der Heilkunde bereits aufweisen. Ihnen ist also eine hinreichende Erfahrung zu unterstellen. Anders als bspw. bei Haverkamp (1993) oder auch Krems und Zierer (1994) wird auf die Untersuchung von in Ausbildung befindlichen Psychologen/Psychotherapeuten verzichtet. Neben dem höheren Niveau an absoluter Berufserfahrung als in den genannten Studien resultiert hieraus zudem eine schärfere Abgrenzung zur Gruppe der Unerfahrenen, da aufgrund der Definition der Gruppen mindestens drei Jahre Berufserfahrung zwischen diesen liegen.

Versuchsplanerisch handelt es sich hierbei um ein quasiexperimentelles Vorgehen, da die Probanden den Gruppen selbstverständlich nicht randomisiert zugeteilt werden können, sondern aufgrund ihrer eigenen Angaben zum beruflichen Status in eine der beiden Gruppe eingeteilt werden. Dies erfolgt unmittelbar zu Beginn der Befragung, indem die Probanden aus einer Liste mit sechs Möglichkeiten diejenige Situation auswählen, die ihren Ausbildungsstand am besten beschreibt (s. Abbildung 5).

Da sich das Projekt OPUS ausdrücklich an spezielle Berufsgruppen/Studenten richtet, würden wir gerne schon an dieser Stelle von Ihnen erfahren, welche Aussage auf Sie zutrifft:

- Ich studiere derzeit Psychologie
- Ich befinde mich derzeit in der Ausbildung zum/r Psychologischen Psychotherapeuten/-in bzw. Kinder- und Jugendlichenpsychotherapeuten/-in
- Ich habe mein Psychologie-Studium abgeschlossen und arbeite im klinischen Bereich (ohne Therapieausbildung)
- Ich bin approbierte/r Psychologische/r Psychotherapeut/-in bzw. Kinder- und Jugendlichenpsychotherapeut/-in
- Ich bin Arzt/Ärztin im psychotherapeutischen Bereich (Auch Ärzte/-innen in Weiterbildung)
- Auf mich trifft keine der genannten Beschreibungen zu

Abbildung 5. Bildschirmansicht zur Einordnung in die Berufsgruppe

Erfahrene und *Unerfahrene* unterscheiden sich dementsprechend in der Menge ihrer berufsrelevanten Erfahrungen. Keinen Unterschied darf es hingegen im Hinblick auf die minimal notwendigen Kenntnisse über die studienspezifische Prüfhypothese geben. Sofern nicht sichergestellt ist, dass auch die *Unerfahrenen* über das notwendige Wissen verfügen (können), das für eine korrekte Beurteilung der Prüfhypothese notwendig ist, wäre ein etwaiger Unterschied nicht auf die Berufserfahrung an sich attribuierbar, sondern auf ein spezifisches Wissensdefizit. Außerdem kann nicht ausgeschlossen werden, dass auch innerhalb der Gruppe der Erfahrenen Wissensunterschiede bzgl. der Prüfhypothese vorzufinden sind, da einige Psychotherapeuten möglicherweise in Spezialkliniken

arbeiten und dementsprechend nur mit einer geringen Anzahl von Störungsbildern in Kontakt kommen, zu denen die gewählte Prüfhypothese ggf. nicht gehört. Es ist daher unabdingbar, ein Minimum an prüfhypothesenrelevantem Wissen für alle Teilnehmer sicherzustellen.

Deswegen erhalten die Probanden neben der Fallbeschreibung einen Hintergrundtext, der die wichtigsten Informationen zur Prüfhypothese und der Alternativhypothese beinhalten, und auf den die Teilnehmer während der gesamten Studie zugreifen können. Der Hintergrundtext kann in Anhang A (Fachwissen relevant) bzw. Anhang B (Fachwissen irrelevant) nachgelesen werden.

4.1.3 *Fachwissen*

Um den Zusammenhang des *confirmation bias* mit dem spezifischen Fachwissen zu untersuchen, werden Personen, die ihr spezifisches Fachwissen bei der Urteilsbildung einsetzen können, mit solchen verglichen, die daran gehindert werden. Begrüßenswert wäre ein Design in Anlehnung an Krems und Zierer (1994), die die Leistung von Fachärzten verschiedener Fachrichtungen anhand eines spezifischen Fachgebiets untersuchten, und damit Fachärzte anderer Fachrichtungen zu Personen ohne spezifisches Fachwissen machten. Die postgraduale Ausbildung von Psychologen ist jedoch nicht in dem Maße spezialisiert wie in der Medizin. D.h. eine derart formale Unterscheidung des Fachwissens aufgrund der gewählten Fortbildungsrichtung ist in der Psychologie nicht zu finden, da im Rahmen der Psychotherapieausbildung das Handwerkszeug für den Umgang mit allen Störungen gelegt wird. Dementsprechend gibt es nicht den „Fachpsychotherapeuten für Essstörungen“, der mit dem „Fachpsychotherapeuten für Depressionen“ verglichen werden könnte.

Eine prinzipiell denkbare Option wäre es ersatzweise die Gruppe der Psychotherapeuten mit Psychologen anderer Fachrichtungen (z.B. Fachpsychologen für Rechtspsychologie oder Experten aus dem Bereich der Arbeits- und Organisationspsychologie) zu vergleichen. Hier wäre (je nach Wahl des konkreten Urteilsgegenstands) vermutlich ein Unterschied hinsichtlich des störungsspezifischen Wissens gegeben, allerdings muss befürchtet werden, dass es auch systematische Unterschiede in zugrundeliegenden diagnostischen Kompetenzen gibt, die jedoch nicht unter dem Fachwissen verstanden werden sollen.

Um also auf Unterschiede im Wissen über Fallspezifika und nicht auf Unterschiede im allgemeinen diagnostischen Vorgehen rückschließen zu können, wurde statt dessen

ein zweiter, fachfremder Urteilsgegenstand in die Studie aufgenommen, über den die Probanden (zumindest aus beruflicher Sicht) *kein* spezifisches Wissen aufgebaut haben konnten. In der Bedingung *Fachwissen relevant* bearbeiteten die Probanden daher eine klinische Störungsdiagnostik, während in der Bedingung *Fachwissen irrelevant* eine laienrichterliche Urteilsbildung vorgenommen werden musste, indem aus der Rolle des Schöffen vor Gericht heraus die Schuld bzw. Unschuld eines wegen schwerer Brandstiftung mit Todesfolge Angeklagten beurteilt werden musste.

Die Zuteilung zu einer der beiden Bedingungen der UV *Fachwissen* erfolgt automatisiert durch das Computerprogramm, dass durch einen Algorithmus die Gleichverteilung auf die Gruppen der UV *Berufserfahrung* wahrt.

4.1.3.1 Fallmaterialien in der Bedingung *Fachwissen relevant* (Essstörung)

An dieser Stelle ist zunächst zu begründen, warum die Wahl auf den Essstörungskontext fiel, um im Anschluss die Konstruktion der Fallbeschreibung sowie der Zusatzinformationen zu erläutern.

4.1.3.1.1 Wahl des Essstörungskontextes

Die klinisch-diagnostische Fragestellung verlangt von den Probanden konkret die Abklärung einer Essstörung bei einer fiktiven 16 jährigen Patientin. Hier werden eine anfängliche Fallvignette sowie später sukzessive weitere Einzelinformationen präsentiert. Die Gesamtheit der ausgegebenen Informationen ist dabei so gewählt, dass die Symptomatik der Diagnose einer *Anorexie vom Binge-Eating/Purging Typus* nach dem DSM-IV (Code: 307.1; Saß et al., 1998) entspricht. Aufgrund der diagnostischen Kriterien, die im Folgenden erläutert werden, ist die Diagnose einer *Bulimie, Purging Typus* (Code: 307.51) in dieser Situation nicht gerechtfertigt. Dieser Kontext wurde gewählt, weil beide Störungen eine große Anzahl an gemeinsamen Symptomen aufweisen, die eine Differentialdiagnostik unter Berücksichtigung von Alternativhypothesen erfordert. Der entscheidende Schritt der von den Probanden durchzuführenden Begutachtung ist daher die differentialdiagnostische Abgrenzung der *Bulimie* von der *Anorexie (Subtypus Binge-Eating/Purging)*, ohne dass dies explizit genannt wird, weil dies als Standard vorausgesetzt werden darf.

Diagnostische Kriterien der Anorexie. Das Störungsbild der *Anorexie* ist nach dem DSM-IV (Saß et al., 1998) durch das Vorliegen von vier diagnostischen Kriterien definiert. Die Patienten zeigen erstens eine Weigerung, das Minimum des für Alter und Körpergröße normalen Körpergewichts zu halten (Kriterium A), zweitens haben sie trotz ihres Untergewichts große Ängste vor einer Gewichtszunahme (Kriterium B) und drittens liegt eine Körperschemastörung bzw. ein übertriebener Einfluss des Gewichts/der Figur auf die Selbstbewertung vor (Kriterium C). Außerdem liegt bei postmenarchalen Frauen eine Amenorrhoe vor (Kriterium D), d.h. das Ausbleiben von mindestens drei aufeinander folgenden Menstruationszyklen.

Der *Anorexie vom Subtypus Binge-Eating/Purging* ist darüber hinaus dadurch charakterisiert, dass während der Episode der *Anorexie* regelmäßig Heißhungeranfälle und so genanntes *Purging-Verhalten* auftreten, d.h. dass selbstinduziert erbrochen wird und/oder Substanzen eingenommen werden, die einer Gewichtszunahme entgegenwirken sollen.

Diagnostische Kriterien der Bulimie. Das Störungsbild der *Bulimie* ist nach dem DSM-IV (Saß et a., 1998) durch das Vorliegen von fünf diagnostischen Kriterien definiert. Die Patienten zeigen erstens wiederholte Episoden von Fressanfällen/Heißhungerattacken, während derer eine erhebliche Menge an Nahrung in einem relativ kurzen Zeitraum zu sich genommen wird und die vom Gefühl des Kontrollverlusts begleitet sind (Kriterium A) und zweitens kommt es zur wiederholten Anwendung von Kompensationsmaßnahmen, die der Gewichtszunahme entgegenwirken (Kriterium B). Das dritte Kriterium besteht darin, dass Fressanfälle und Kompensationsmaßnahmen über einen bestimmten Zeitraum mit einer bestimmten Regelmäßigkeit aufgetreten sind (Kriterium C), außerdem muss die Selbstbewertung übermäßig stark von Figur und Körpergewicht abhängen (Kriterium D). Das letzte Kriterium (E) stellt ein Ausschlusskriterium dar, da vorgegeben wird, dass die Symptome nicht ausschließlich im Verlauf von Episoden einer *Anorexie* auftreten. Deshalb kann keine *Bulimie* diagnostiziert werden, wenn die Kriterien einer *Anorexie* erfüllt sind. Aufgrund dessen ist die differentialdiagnostische Abgrenzung von *Anorexie* und *Bulimie* der zentrale Aspekt dieser Diagnostik.

Gemeinsamkeiten und Unterschiede der Störungsbilder. Dabei sind die Gemeinsamkeiten insbesondere zwischen *Anorexie (Subtypus Binge-Eating/Purging)* und *Bulimie* beträchtlich: die Kernsymptome der *Bulimie* (Fressanfälle und Kompensationen) kommen auch bei anorektischen Patienten dieses Subtypus vor. Beide Gruppen teilen die

Angst vor einer Gewichtszunahme und die übermäßige Abhängigkeit des Selbstwertgefühls von Figur und Körpergewicht. Des Weiteren sind beide Patientengruppen übermäßig mit dem Thema Essen und Körpergewicht beschäftigt und darauf kognitiv fixiert (Fichter & Warschburger, 2002).

In Bezug auf die diagnostischen Kriterien nach dem DSM-IV (Saß et al., 1998) lässt sich eine *Anorexie (Binge-Eating/Purging-Typus)* von einer *Bulimia nervosa* durch das Gewichtskriterium, das Vorliegen der Körperschemastörung und die Amenorrhoe unterscheiden. Die restlichen Symptome sind unspezifisch. Von Wietersheim (2003) nennt weitere Unterschiede, die sich jedoch nicht auf die diagnostischen Kriterien, sondern begleitende Symptome beziehen. So seien bei Bulimie-Patienten mehr Leidensdruck, größere Schuldgefühle sowie eine höhere Therapiebereitschaft bzw. Compliance zu beobachten als bei Anorektikern. Darüber hinaus können bei Letzteren häufiger sexuelle Ängste bzw. Störungen sowie Perfektionismus beobachtet werden (Davison & Neale, 2002). Die Begutachtung und differentialdiagnostische Unterscheidung sollte sich jedoch im Wesentlichen auf die Unterschiede in den diagnostischen Kriterien beziehen. Gemeinsamkeiten (unspezifische Kriterien) und Unterschiede (spezifische Kriterien) sind in diesem Sinne in der Abbildung 6 aufgezeigt.



Abbildung 6. Überschneidung der diagnostischen Kriterien von Bulimie und Anorexie (Subtypus Binge-Eating/Purging) im Sinne des DSM-IV (Saß et al. 1998)

Der Abbildung kann entnommen werden, dass die bulimiekonstituierenden Symptome ebenfalls bei einer Anorexie vorliegen. Nicht in die Veranschaulichung aufgenommen wurden das Zeitkriterium (C, s.o.) sowie des Ausschlusskriteriums E der Bulimie (s.o.), weil diese nicht im Sinne eines qualitativen Symptoms die Krankheit kennzeichnen, sondern äußere Auftretensbedingungen explizieren.

4.1.3.1.2 Konstruktion und Eignungsprüfung der Fallbeschreibung

Für die Konstruktion der Fallbeschreibung wurde auf die eigene Vorarbeit (Dreger, 2006) zurückgegriffen, wobei einige Modifikationen vorzunehmen waren. So wurden bspw. diejenigen Füllinformationen entfernt, die eingeflochten worden waren, um die damals genutzte Coverstory (Evaluation des diagnostischen Curriculum des Lehrstuhls) zu implementieren. Desweiteren wurde der Text sprachlich mit dem Ziel überarbeitet, möglichst noch einfachere und kürzere Formulierungen zu finden, um nur die inhaltlich notwendigen Informationen zu übermitteln. Die Themenblöcke zur Sexualität bzw. dem Ehrgeiz bei Wettkämpfen wurden aus der Fallvignette entfernt, um jegliche diagnostischen Hinweise im ersten Informationsblock zu vermeiden. Die resultierende Fallbeschreibung ist in Anhang A zu finden.

Aufgrund der Änderungen in der Fallbeschreibung aus der eigenen Vorarbeit (Dreger, 2006) war eine neue Überprüfung der Eignung der Zusammenstellung erforderlich. Die Fallbeschreibung wurde daher zusammen mit einer Instruktion sowie den Items zur Eignungsprüfung $N = 13$ Personen (100% weiblich; Alter: $M = 29.31$, $SD = 6.87$) vorgelegt, die entweder Studierende der Psychologie oder bereits diplomierte, praktisch arbeitende Psychologen waren. Hier sollen die Ergebnisse berichtet werden, die vorgelegten Materialien sind in Anhang A dargestellt.

Die Einschätzungen hinsichtlich der *Lesbarkeit* und *Plausibilität* auf einer siebenstufigen Skala von jeweils $0 = \text{sehr schlecht}$ bis $6 = \text{sehr gut}$, waren zufriedenstellend (Lesbarkeit: $M = 5.00$; $SD = 1.41$; Plausibilität: $M = 4.62$; $SD = 0.96$).

Desweiteren interessierte, ob mit der Fallbeschreibung die Prüfhypothese Bulimie ausgelöst werden konnte, die durch den Hinweis des fiktiven Arztes auf seine Verdachtsdiagnose induziert wurde. Insofern war gewollt, dass einerseits das Ausmaß, in dem die Informationen als charakteristisch für eine Bulimie eingeschätzt wurden, höher ist, als dasjenige Maß für eine Anorexie und andererseits die Wahrscheinlichkeit des Vorliegens einer Bulimie höher bewertet wird, als bei einer Anorexie. Die errechneten Mittelwerte (s. Tab. 6) stehen damit im Einklang.

Der Tabelle ist außerdem zu entnehmen, dass beide Störungen in beiden Abfragen im mittleren Skalenbereich eingeschätzt werden. Hieraus ist ableitbar, dass eine hinreichend ambigue bzw. unsichere Situation konstruiert wurde. Die Eignung hinsichtlich der angelegten Kriterien wird dementsprechend als gesichert angesehen.

Tabelle 6. *Einschätzung der Wahrscheinlichkeit des Vorliegens einer Bulimie bzw. Anorexie sowie des Ausmaßes, in dem die Informationen charakteristisch für Bulimie bzw. Anorexie sind*

<i>Merkmal</i>	<i>Störung</i>	<i>M</i>	<i>SD</i>
<i>Wahrscheinlichkeit</i>	Bulimie	3.77	1.79
	Anorexie	2.54	1.27
<i>Wie charakteristisch sind die Informationen für...</i>	Bulimie	3.69	1.18
	Anorexie	2.85	1.14

Anmerkungen. Die Einschätzungen erfolgten jeweils auf Skalen von 0 = *gar nicht...* bis 6 = *sehr...*

4.1.3.1.3 Konstruktion der Zusatzinformationen

Neben der Fallbeschreibung waren weitere 30 Informationen zu konstruieren, die den Probanden im weiteren Verlauf der Studie jeweils einzeln präsentiert wurden, um daraus Rückschlüsse auf ein eventuell verzerrtes Interpretationsmuster ziehen zu können, welches im Rahmen mehrerer AVn erhoben wurde.

Abbildung 7 (siehe folgende Seite) zeigt den mehrstufigen Prozess zur Konstruktion der resultierenden 30 Items, dem zwei unterschiedliche Konstruktionsstränge zugrundeliegen. Denn einerseits wurde zur Generierung von essstörungsrelevanten Items auf den eigenen Fragenpool (Dreger, 2006) zurückgegriffen, um daraus Items zu extrahieren und andererseits wurden essstörungsirrelevante Items neu generiert, da der Fragenpool keine solchen enthielt.

Essstörungsrelevante Items. Beim angesprochenen Fragenpool handelt es sich um eine Zusammenstellung von Fragen an eine fiktive Patientin zur Abklärung einer Essstörungssymptomatik, die entweder aus den diagnostischen Kriterien des DSM-IV (Saß et al., 1998; konkret für Bulimie oder Anorexie Subtypus Binge-Eating/Purging) abgeleitet wurden oder in Anlehnung an die Items des *Eating Disorder Inventory 2* (EDI-2; Paul & Thiel, 2005) bzw. an diejenigen des *Strukturierten Interviews für Anorektische und Bulimische Essstörungen* (SIAB; Fichter & Quadflieg, 1999) generiert wurden. Weitere Informationen waren vor dem Hintergrund der Arbeiten von Pudel und Westenhöfer (1991), Jacobi, Thiel und Paul (2000) sowie Wittek (2003) vorgenommen worden. Außerdem lagen zu diesen Items bereits Einschätzungen von unabhängigen Ratern aus anderen Vorarbeiten vor, die sich auf die Diagnostizität, die Bedeutung für die Diagnose Bulimie sowie Anorexie (Subtypus Binge-Eating/Purging) bezogen.

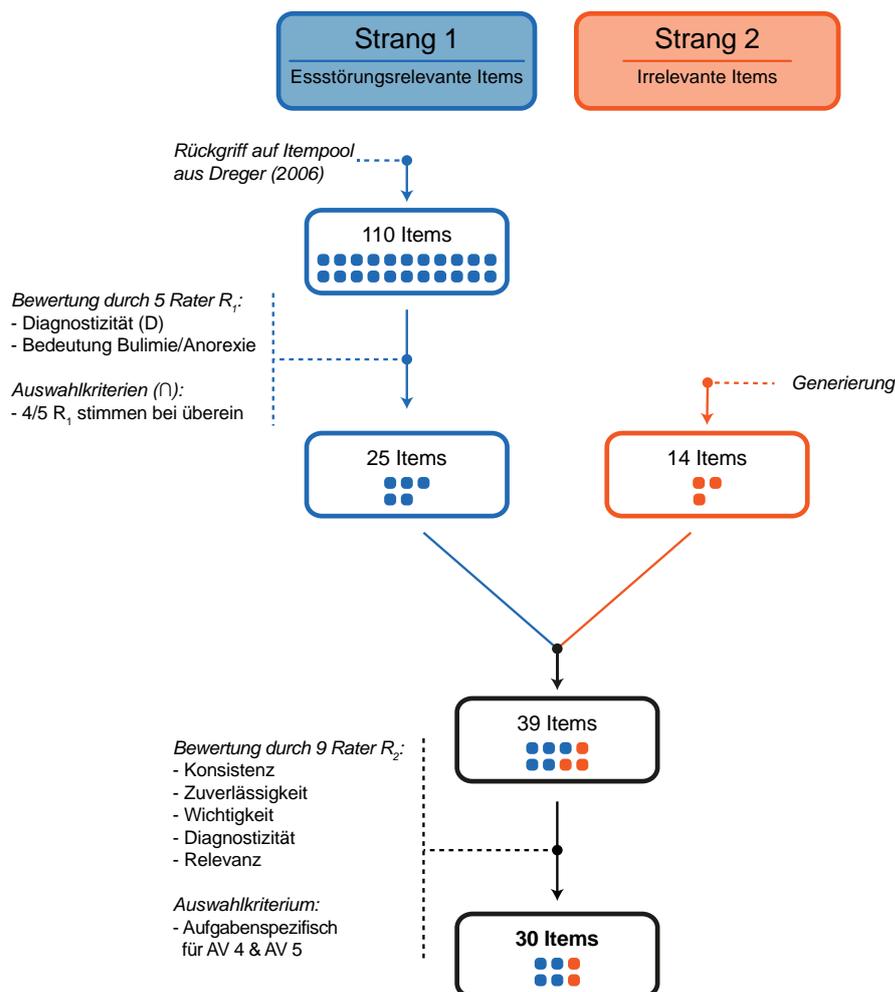


Abbildung 7. Veranschaulichung Konstruktion und Auswahl der Zusatzitems

Als Auswahlkriterium wurde nun gefordert, dass ein Item von mindestens 4 der 5 Rater sowohl hinsichtlich der Bedeutung für die Diagnose Bulimie als auch hinsichtlich der Bedeutung für die Diagnose Anorexie (Subtypus Binge-Eating/Purging) und der Diagnostizität übereinstimmend eingeschätzt worden war. Dieses strenge Auswahlkriterium führte zu einer Reduktion auf 24 Items, die Hälfte davon jeweils zuvor als diagnostisch oder nondiagnostisch eingeschätzt. Ein weiteres Item wurde trotz des nicht erfüllten Kriteriums mit aufgenommen, das es sich unmittelbar auf das Vorliegen einer Körperschemastörung bezieht, die für die Diagnosestellung relevant ist.

Die Fragen wurden im Anschluss zu Aussagen umformuliert, weil nicht der Prozess der Informationssuche im Vordergrund steht, sondern die Bewertung von Informationen über die Patientin, die in Form von Aussagen vorgelegt werden.

Essstörungsirrelevante Items. Da der Itempool von Dreger (2006) so konstruiert war, dass alle Informationen Bezug zur Essstörungsdiagnostik aufweisen, mussten die einzusetzenden irrelevanten Variablen neu generiert werden. Konkret wurden 14 Items formuliert, die bspw. die Angst vor Hunden thematisieren, die die Patientin als Kind nach Angaben der Mutter aufgewiesen habe.

Abschließendes Rating. Die insgesamt 39 Items wurden nun einem gemeinsamen Rating unterzogen. Insgesamt neun unabhängige und hypothesenblinde Rater bewerteten die Items in Bezug auf fünf Merkmale. Zunächst war auf einer Skala von $-5 = \textit{sehr inkonsistent}$ über $0 = \textit{neutral/irrelevant}$ bis $+5 = \textit{sehr konsistent}$ die *Konsistenz der Information mit der Prüfhypothese* Bulimie einzuschätzen, d.h. es sollte berücksichtigt werden, ob sie zu dem typischen Krankheitsbild passen (konsistent) oder eigentlich nicht erwartet werden, wenn die Störung vorliegt (inkonsistent). Wenn die Informationen in dieser Hinsicht neutral sind oder gar nichts mit der Essstörung zu tun haben, waren Einschätzungen im mittleren Skalenbereich angezeigt. Die Einschätzung der Konsistenz ist relevant, weil im Zuge der Erfassung der AV 4 (Bestätigungstendenz B) jeweils eine bestimmte Anzahl konsistenter wie inkonsistenter Items zu präsentieren ist, die die Probanden dann bewerten. Das beschriebene Rating im Vorfeld dient nun dazu, den vorzugebenden Informationen vorab die Eigenschaft *konsistent* oder *inkonsistent* zuschreiben zu können, was durch den Einsatz der neun verschiedenen Rater abgesichert wurde. Ein Item wird dann als konsistent angesehen, wenn der Mittelwert der Ratereneinschätzungen signifikant größer ist als null. Analog wird ein inkonsistentes Items durch einen Mittelwert repräsentiert, der signifikant kleiner ist als null. Items, die keinen von null verschiedenen Mittelwert aufweisen, gelten als bidirektional bzw. irrelevant und werden nicht in die Auswertung von AV 4 mit einbezogen. Diese Vorgehensweise orientiert sich an derjenigen von Rassin et al. (2010).

Um ausschließen zu können, dass die mit Hilfe der AV 4 zu messenden Effekte der Einflussvariablen durch Störeinflüsse beeinträchtigt werden, musste vorab sicher gestellt werden, dass die einzuschätzenden Items sich lediglich hinsichtlich der *Konsistenz* unterschieden, nicht jedoch hinsichtlich ihrer *Wichtigkeit*, ihrer *Zuverlässigkeit* und ihrer *Diagnostizität*. Die Rater schätzten diese drei Merkmale für jedes Item auf jeweils elfstufigen Skalen von $0 = \textit{gar nicht}$ bis $100 = \textit{sehr bzw. vollkommen ein}$. Für die konkrete Auswahl der Items je AV wird dann zu berücksichtigen sein, dass sich keine Mittelwertunterschiede ergeben. Dieser Nachweis wird in Abschnitt 4.2.4 im Zusammenhang mit der Präsentation der ausgewählten Items erbracht.

Da für AV 5 (*Hypothesenkonsistente Interpretation uneindeutiger Evidenz bzw. Bestätigungstendenz I*) die *Bedeutung der Information für die Prüfhypothese* (Bulimie) relevant ist, mussten die Rater die Items vorab daher auch als *eindeutigen Beleg, brauchbaren Hinweis, Widerspruch* oder *irrelevante Information* kennzeichnen: so gibt es Informationen, die vorliegen müssen, weil die diagnostischen Kriterien des obligatorischen Klassifikationssystems dies vorschreiben (fortan als *eindeutiger Beleg* bezeichnet). Zudem gibt es solche Symptome, die zwar typischerweise vorliegen, wenn die Diagnose gerechtfertigt ist, die jedoch aufgrund der Tatsache, dass sie keine Erwähnung in den Kriterien finden, nicht als notwendigen Bedingungen fungieren. Sie zeigen daher eine gewisse Wahrscheinlichkeit der Störung an, ohne diese tatsächlich zu belegen (Bezeichnung als *brauchbarer Hinweis*). Darüber hinaus gibt es Informationen, die gegen das Vorliegen der Störung sprechen (*Widerspruch*) und solche, die in keinen Zusammenhang mit der Störung gebracht werden können, weil es bspw. keine empirischen Hinweise auf Kovariationen des Merkmals mit den störungsspezifischen Symptomen gibt (*irrelevant*).

Ein Merkmal wurde dann als Repräsentant einer Merkmalskategorie zugelassen, wenn mindestens 5 der 9 Rater übereinstimmend geurteilt haben. So wurde die Qualität der Items vorab festgelegt, um später einen Vergleich mit den Einschätzungen der eigentlichen Probanden vorzunehmen, bei dem v.a. interessieren wird, ob die Probanden im konkreten Studienkontext eine abweichende Aussagekraft der Items angeben. Dies wäre bspw. der Fall, wenn ein Proband ein Item als *brauchbaren Hinweis* kategorisiert, das seitens der Rater vorab mehrheitlich als *irrelevant* klassifiziert worden war.

Im Anhang A sind für alle Items die Einschätzungen hinsichtlich aller hier beschriebenen Merkmale (und die Items selbst) gelistet.

4.1.3.2 *Fallmaterialien in der Bedingung Fachwissen irrelevant (Brandstiftung)*

Wie in Abschnitt 4.1.3 erwähnt, wurde die Bedingung *Fachwissen irrelevant (Brandstiftung)* aus methodischen Gesichtspunkten gewählt, um die psychologisch geschulten Probanden in einem Kontext urteilen lassen zu können, über den sie nur wenig und wenn überhaupt kein aus ihrer beruflichen Rolle entspringendes Fachwissen aufweisen sollten. Deswegen wurde auf eine von der Arbeitsgruppe entwickelte Fallbeschreibung zurückgegriffen, die auch in anderen Arbeiten zum Einsatz kam (vgl. Konitzky, 2009; vom Schemm, 2008b).

4.1.3.2.1 Konstruktion und Eignungsprüfung der Fallbeschreibung

Die Fallbeschreibung greift den Fall von Kenneth Richey gegen den Staat Ohio von 1987 auf, der wegen Mordes, Brandstiftung und Einbruchdiebstahl zum Tode verurteilt wurde. Für eine mediale Aufarbeitung s. Cramb (2007). Ihm wurde vorgeworfen, in ein Gewächshaus eingebrochen zu sein, um Brandbeschleuniger zu beschaffen. Damit habe er aus Eifersuchtsgründen ein Feuer in der Wohnung verursacht, die über derjenigen seiner Ex-Freundin und ihrem neuen Partner gelegen habe. Im Zuge dieses Feuers sei ein zweijähriges Mädchen (in der Brandwohnung) umgekommen, für die der Angeklagte die *Babysitter-Rolle* übernommen habe.

Richey habe im gesamten Prozessgeschehen stets seine Unschuld beteuert, einem ersten *plea bargain* (im Sinne einer im amerikanischen Rechtssystem üblichen Absprache zwischen Staatsanwaltschaft und Angeklagtem) nicht zugestimmt, das immerhin eine Strafreduktion auf 11 Jahre bedeutet habe und habe wiederholt Berufungsanträge gestellt. Als 2007 Zweifel an der Aussagekraft der Indizienbeweise aufgekommen seien – diese betrafen v.a. die wissenschaftliche Korrektheit der seinerzeit durchgeführten Brandschuttanalysen – sei der Fall nach nunmehr 20 jährigem Aufenthalt in der Todeszelle neu aufgerollt worden, was u. a. von *Amnesty International* gefordert wurde (vgl. Amnesty International, 2007). Im Zuge der Vorbereitung auf die Verhandlung habe Richey dann überraschender Weise einem *plea bargain* zugestimmt, wobei er sich nicht schuldig bekannte, sondern lediglich zu Protokoll gab, keinen Einspruch gegen die Vorwürfe fahrlässige Tötung, Gefährdung eines Kindes sowie Einbruchdiebstahl zu erheben. Brandstiftung und Mord wurden fallen gelassen. Richey wurde zu der bereits verbüßten Strafe verurteilt und umgehend freigelassen.

Amnesty International wird dahingehend zitiert (Cramb, 2007), dass sie in diesem *cause celebre* einen der eindeutigsten Fälle von Unschuld sehen, den Menschenrechtler je erlebt hätten. Aufgrund der Tatsache, dass Richey in den Handel eingewilligt hat und es daher keine richterliche Entscheidung im Sinne eines Freispruchs aus erwiesener Unschuld gab, so muss zumindest im Sinne des *in dubio pro reo* Grundsatzes von seiner Unschuld ausgegangen werden, da im konkreten Fall sowohl eine größere Anzahl schuldkonsistenter als auch schuldinkonsistenter Informationen vorliegt. Diese Konstellation macht den Fall auch so geeignet für die vorliegende Studie. Es sei aber nochmal daraufhin gewiesen, dass im Rahmen der vorliegenden Arbeit keine Partei für oder gegen Kenneth Richey ergriffen werden soll. Für den Einsatz mussten die Informationen allerdings aufbereitet werden. Änderungen, die Konitzky vorgenommen hat und in ihrer Arbeit von 2009 beschreibt, betreffen die Namen der Beteiligten und den Tatort, der nach Deutschland verlagert wurde. Außerdem wurde

eine zeitlich engere Relation der Abläufe konstruiert und die Zusammenstellung danach ausgerichtet, dass in der Fallvignette zu Beginn der Studie vor allem schuldkonsistente (neben einigen irrelevanten) Informationen gegeben wurden, die die Prüfhypothese der Schuld salient machen, in ihrer Gesamtheit aber nicht bestätigen konnten. Die resultierende Fallbeschreibung ist Anhang B nachzulesen und wurde einer analogen Eignungsprüfung unterzogen wie die Fallvignette der essgestörten Patientin, die zu akzeptablen Ergebnissen führte, was nunmehr zu belegen ist.

Die Fallbeschreibung wurde $N = 7$ Personen (86% weiblich; Alter: $M = 29.85$; $SD = 6.67$; entweder Studierende oder bereits Diplomierte) vorgelegt. Die Einschätzungen hinsichtlich der *Lesbarkeit* und *Plausibilität* auf einer siebenstufigen Skala von jeweils $0 = \text{sehr schlecht}$ bis $6 = \text{sehr gut}$, waren geringer als bei der Fallvignette zur Essstörungsdiagnostik, da sie aber mindestens dem Skalenmittelpunkt entsprachen dennoch ausreichend (Lesbarkeit: $M = 3.57$; $SD = 1.27$ sowie Plausibilität: $M = 3.14$; $SD = 1.57$). Im Unterschied zur klinischen Vignette enthält die Beschreibung des Brandstiftungsfalls deutlich mehr Detailinformationen wie die Namen sehr vieler beteiligter Personen, konkrete Wohnorte und Uhrzeitangaben, die zwar nicht urteilsrelevant sind, aber die Lesbarkeit natürlich beeinträchtigen und unter Umständen die Nachvollziehbarkeit erschweren und sich damit indirekt auch auf die Plausibilität auswirken könnten.

Desweiteren interessierte, ob die Fallbeschreibung zuverlässig eine Schuld-Hypothese auslösen konnte. Wie der Tabelle 7 zu entnehmen ist, wurde die *Wahrscheinlichkeit eines Schuldurteils* und die *Charakterisierung der Schuld* durch die Items geringer eingeschätzt als jeweils bzgl. eines Freispruchs aus Mangel an Beweisen. Da andererseits die Einstellung wegen Geringfügigkeit bzw. der Freispruch wegen erwiesener Unschuld als sehr unwahrscheinlich beurteilt wurden, kann man davon ausgehen, dass zumindest eine hohe Ambiguität bzgl. einer schwerwiegenden Tat wahrgenommen wird. In dieser Situation der Ambiguität ist es im Kontext einer juristischen Entscheidung über die Schuld bzw. Unschuld ja auch gerade in der Berücksichtigung des in *dubio pro reo* nur statthaft, dem Freispruch aus Mangel an Beweisen den Vorzug zu geben. Insofern handelten die Rater juristisch korrekt, da die Fallvignette ja gerade keine solche Menge an schuldkonsistenten Informationen liefern sollte, dass eine Verurteilung gerechtfertigt sein sollte. Die im Sinne der identischen Abschätzung des Fallmaterials aus dem Essstörungsfall übertragenen Items konnten daher unter Umständen aufgrund dieses Umstands kein aussagekräftiges Abbild liefern. Die scheinbar erwartungswidrigen Ergebnisse sollen daher weniger gewichtet werden. Der Hinweis auf den ambigen Charakter des Materials muss ausreichen.

Tabelle 7. *Einschätzung der Wahrscheinlichkeit der möglichen Urteilsoptionen sowie des Ausmaßes, in dem die Informationen charakteristisch für die jeweiligen Optionen sind*

<i>Merkmal</i>	<i>Urteilsoption</i>	<i>M</i>	<i>SD</i>
<i>Wahrscheinlichkeit</i>	Einstellung wegen Geringfügigkeit	0.14	0.38
	Schuld-Urteil	2.71	1.11
	Freispruch aus Mangel an Beweisen	3.71	0.76
	Freispruch wegen erwiesener Unschuld	1.71	0.95
<i>Wie charakteristisch sind die Informationen für...</i>	Erwiesene Schuld	2.71	1.50
	Freispruch aus Mangel an Beweisen	3.29	1.50
	Freispruch wegen erwiesener Unschuld	1.86	0.90

Anmerkungen. Die Einschätzungen erfolgten jeweils auf Skalen von 0 = *gar nicht...* bis 6 = *sehr...*

4.1.3.2.2 *Konstruktion der Zusatzitems*

Desweiteren wurden aus den recherchierten Fallinformationen 42 Items generiert, die als Ausgangspool für die Konstruktion der Zusatzitems dienten und von denselben Ratern hinsichtlich der gleichen Merkmale eingeschätzt wurden wie diejenigen des Essstörungskontextes. Alle Definitionen und Kriterien gelten analog.

4.2 *Abhängige Variablen*

Es liegen insgesamt sechs bedingte Variablen vor, die teilweise in mehrere Maße differenziert werden. Dies soll in der Folge dargestellt werden.

4.2.1 *AV1: Schlussfolgerung*

Um zu überprüfen, ob es zu einer Selbstbestätigung der Prüfhypothese kommen kann, werden die Probanden am Ende der Studie um ein abschließendes Urteil in offenem Antwortformat gebeten (s. Anhang C).

Probanden der Bedingung *Fachwissen relevant (Essstörung)* weisen dann einen *confirmation bias* auf, wenn sie die objektiv falsche Diagnose *Bulimie* (unabhängig ob mit oder ohne entsprechenden Subtypus) stellen, die als Verdachtsdiagnose seitens des Hausarztes mitgeteilt wurde. Die Probanden der Bedingung *Fachwissen irrelevant (Brandstiftung)* weisen analog einen *confirmation bias* auf, wenn sie den Angeklagten im Sinne der Anklage für *schuldig* befinden. Dies ergibt sich aus der Forderung, einen Angeklagten im Sinne des Grundsatzes *in dubio pro reo* freizusprechen, wenn Zweifel an seiner Schuld nicht hinreichend ausgeräumt werden können. Aufgrund der Konstruktion des Falles liegen sowohl stark schuldkonsistente als auch schuldinkonsistente Informationen vor, die weder einen Freispruch wegen erwiesener Unschuld noch eine Verurteilung wegen erwiesener Schuld rechtfertigen. Formal korrekt ist daher der Freispruch aus Mangel an Beweisen.

4.2.2 AV 2: Subjektive Sicherheit

Zusätzlich zum Urteil geben alle Probanden auf einer siebenstufigen Skala von 0=*sehr unsicher* bis 6=*sehr sicher* an, wie sicher sie sich bei ihrer Entscheidung hinsichtlich des Schlussurteils sind (zum genauen Wortlaut s. auch Anhang C).

4.2.3 AV 3: Güte der Evidenzregeln

Da die Evidenzregeln im Sinne von Trope und Liberman (1996) den zweiten Schritt nach der Hypothesengenerierung darstellen, werden die Probanden unmittelbar nach dem Lesen der ersten Fallinformationen gebeten, insgesamt sechs Evidenzregeln selbst zu generieren. Die in Abbildung 8 wiedergegebene Instruktion veranschaulicht den Probanden ihre Aufgabe.

Da die Daten der Probanden als Freitextantworten vorlagen, mussten sie nach Abschluss der Datenerhebung zunächst in einem Ratingverfahren quantifiziert werden. Das gesamte Ratingmanual samt Schulungsbeispielen findet sich in Anhang D. Zunächst kodierten zwei unabhängige und hypothesenblinde Rater die *Instruktionskonformität* jeder einzelnen Evidenzregel, denn nur wenn tatsächlich beachtet wurde, die angenommene Störung bzw. die Annahme von Schuld/Unschuld in die *Wenn-* oder *Dann-Passage* mit aufzunehmen, konnten die weiteren Auswertungsschritte vorgenommen werden.

In der ersten Phase der Urteilsbildung ist es wichtig, sich Gedanken über mögliche Symptomkonstellationen bzw. mögliche Informationsmuster zu machen. Wir möchten nun an dieser Stelle erfahren, welche Bezüge Sie herstellen und bitten Sie, dazu diese erste Aufgabe zu bearbeiten.

Modelle der Informationsverarbeitung - z.B. von Trope und Liberman (1996) - gehen davon aus, dass Beurteiler eingangs sogenannte *Evidenzregeln* bilden. Damit ist gemeint, dass Personen bewusst oder unbewusst *Wenn-Dann*-Regeln aufstellen, die Vorhersagen darüber beinhalten, wann eine mögliche Erklärung (in diesem Fall eine Diagnose) bestätigt oder widerlegt wird. Eine solche Evidenzregel beinhaltet demnach die Forderung, dass bestimmte Gegebenheiten vorliegen müssen oder aber gerade nicht gegeben sein dürfen, damit eine vorgegebene Vermutung zutrifft.

Zur Veranschaulichung hier ein alltagspsychologisches Beispiel: Student A möchte herausfinden, ob Student B feindselig ist. Mögliche Evidenzregeln lauten wie folgt:

- Wenn B feindselig ist, dann müsste er schon auf kleine Provokationen sehr ärgerlich und unangemessen reagieren.
- Wenn B feindselig ist, dann dürfte er es nicht kommentarlos hinnehmen, dass jemand anderes ihn versehentlich anrempelt.

Ihre Aufgabe wird es nun sein, 6 Evidenzregeln aufzustellen, die Sie für die Beurteilung des Falles hilfreich finden. Überlegen Sie sich bitte, welche Symptomkonstellation Sie erwarten bzw. nicht erwarten würden. Geben Sie dazu immer an, welches Störungsbild Sie in Betracht ziehen (Wenn-Passage) und welche Datenlage (Dann-Passage) Sie erwarten.

z.B.	Wenn die Person feindselig ist (Wenn-Passage),	dann sollte sie schon bei Kleinigkeiten aggressiv reagieren (Dann-Passage).
1.		
2.		
3.		
4.		

Abbildung 8. Bildschirmansicht der Instruktion zur Generierung der sechs Evidenzregeln sowie der Eingabefelder (Fachwissen relevant)

Anmerkungen. Aus Platzgründen sind hier nur die ersten 4 Eingabefelder dargestellt.

AV 3a – Präferenz für Evidenzregeln mit positiver Teststrategie. Für das erste Merkmal für die Güte der Evidenzregeln wird festgehalten, ob es sich um eine Evidenzregel im Sinne einer *positiven* oder *negativen Teststrategie* handelt. Eine positive Teststrategie zeigen die Probanden dann, wenn die Evidenzregeln so formuliert werden, dass durch sie das Vorliegen von Merkmalen abgefragt wird, die konsistent mit der eigenen Vermutung sind. In diesem Sinne wird nur auf Informationen geschaut, die das Vorliegen belegen könnten, während Merkmale, die das Vorliegen widerlegen könnten, nicht betrachtet werden. Keine positive, sondern eine negative Teststrategie liegt vor, wenn ein Merkmal in Betracht gezogen wird, dass das Vorliegen der Prüfhypothese widerlegen kann. Die *Dann*-Passage gibt zu verstehen, dass das angesprochene Merkmal gerade nicht vorliegen darf, wenn die Annahme gilt. Tabelle 8 gibt jeweils ein Beispiel der positiven und negativen Teststrategie wieder.

Tabelle 8. Beispiele zur Veranschaulichung der positiven vs. negativen Teststrategie bei den Evidenzregeln

<i>Teststrategie</i>	<i>Beispiel einer Evidenzregel</i>
<i>positiv</i>	Wenn eine Person feindselig ist, dann reagiert sie schon bei Kleinigkeiten aggressiv.
<i>negativ</i>	Wenn eine Person feindselig ist, dann wird nicht zu beobachten sein, dass sie bei ausgesprochenen Provokationen die Ruhe bewahrt.

Auswertungsrelevant ist die Differenz zwischen der Anzahl positiver und negativer Evidenzregeln, wobei Werte größer Null eine Präferenz der positiven Teststrategie nahelegen, der Wert von Null ein ausgewogenes Verhältnis impliziert und negative Werte eine Mehrzahl negativer Evidenzregeln anzeigt.

AV 3b – Präferenz für Evidenzregeln mit nicht diagnostischen Inhalten. Die Diagnostizität im Sinne einer differenziellen Auftretenswahrscheinlichkeit zwischen den Alternativhypothesen, d.h. für Bulimie vs. Anorexie sowie Schuld vs. Unschuld, wird von den Ratern in den beiden Ausprägungen *diagnostisch* und *nondiagnostisch* kodiert. Tabelle 9 nennt je ein Beispiel für die beiden Optionen, die im Anschluss begründet werden sollen.

Tabelle 9. Beispiele zur Veranschaulichung von diagnostischen und nicht diagnostischen Inhalten von Evidenzregeln

<i>Teststrategie</i>	<i>Beispiel einer Evidenzregel</i>
<i>diagnostisch</i>	Wenn eine Person feindselig ist, dann reagiert sie schon bei Kleinigkeiten aggressiv.
<i>nondiagnostisch</i>	Wenn eine Person feindselig ist, dann reagiert sie bei ausgesprochenen Provokationen aggressiv.

Es ist plausibel, dass feindselige Personen auch bei kleinen Provokationen aggressiv reagieren, wohingegen freundliche Personen (Alternativhypothese) dies mit großer Wahrscheinlichkeit nicht tun, weshalb die Reaktion auf eine milde Provokation diagnostischen Wert hat. Dies gilt nicht für die Reaktion auf eine starke Provokation, da man annehmen muss, dass freundliche Menschen hierauf ebenfalls aggressiv reagieren. Die unterschiedliche Auftretenswahrscheinlichkeit ist nicht mehr gegeben und daher auch

keine Diagnostizität. Auswertungsrelevant ist analog zu AV 3a die Differenz zwischen Evidenzregeln mit nicht diagnostischen und diagnostischen Merkmalen. Positive Werte zeigen die zu untersuchende Präferenz an.

AV 3c – Anzahl berücksichtigter Alternativhypothesen. An dieser Stelle soll analysiert werden, ob die Probanden sich auf die Bulimie-Diagnose bzw. Schuldvermutung fokussiert haben und Alternativen außer Acht gelassen wurden. Im Bulimie-Fall meint dies konkret, dass die Probanden nicht ausschließlich Merkmale bzgl. eines Störungsbildes ansprechen, sondern mehrere Störungsbilder in Betracht ziehen. Im Brandstiftungs-Fall meint dies konkret, dass nicht nur die Schuld sondern auch die Unschuld thematisiert wird.

Zunächst wird hierfür ausgezählt, wie viele unterschiedliche Alternativhypothesen in den Evidenzregeln enthalten sind und anschließend wird pro Alternative registriert, wie viele Evidenzregeln hierfür aufgestellt wurden.

4.2.4 AV 4: Bestätigungstendenz B

Um zu untersuchen, ob Personen hypothesenkonsistenten Informationen mehr Gewicht beimessen als inkonsistenten Informationen, was einer Bestätigungstendenz B sensu Hager und Weißmann (1991) entspricht, werden ihnen im Anschluss an das Generieren der Evidenzregeln 18 Items im Sinne von weiteren Fallinformationen präsentiert, die von den Probanden hinsichtlich zweier Merkmale eingeschätzt werden müssen. Zunächst geben sie auf einer siebenstufigen Skala von 0 = *gar nicht wichtig* bis 6 = *sehr wichtig* an, wie wichtig die Information für den Fall aus ihrer Sicht ist. Danach geben sie eine Einschätzung ab, ob die entsprechende Information die Prüfhypothese abschwächt oder bekräftigt. Konkret geben sie dieses Urteil auf einer Skala von -3 = *sehr starke Abschwächung der Prüfhypothese* über 0 = *keinen Einfluss* bis +3 = *sehr starke Bekräftigung der Prüfhypothese* an.

Ordnen die Personen den konsistenten Informationen einerseits eine größere Wichtigkeit als den inkonsistenten Daten zu und schätzen andererseits ihre Bekräftigung der Hypothese größer ein als die Abschwächung seitens der inkonsistenten Angaben, so liegt eine Bestätigungstendenz B vor. Die Bildschirmansicht der Aufgabe ist in Anhang C veranschaulicht.

Bedingung Fachwissen relevant (Essstörung). Die 18 eingesetzten Items können hinsichtlich ihres Wortlauts und der Reihenfolge dem Anhang A entnommen werden. Angestrebt war die Aufnahme von jeweils sechs zuvor als konsistent, inkonsistent

oder irrelevant eingeschätzten Items. Aufgrund der Notwendigkeit, zusätzlich und gleichzeitig die Diagnostizität, Wichtigkeit und Zuverlässigkeit zwischen den konsistenten und inkonsistenten Items auszubalancieren, war jedoch nur die Verwendung von je vier konsistenten und inkonsistenten Items möglich. Die restlichen 10 Items waren Füllitems, die jedoch größtenteils aus sachlogischen Gründen ebenfalls zu präsentieren waren, um im Gesamtkontext der Studie eine hinreichende Datenbasis für eine Entscheidung über das Zutreffen der Prüfhypothese fällen zu können.

Tabelle 10. Begründung der Einteilung in konsistente vs. inkonsistente Items (Essstörung)

Ein- teilung	Item	<i>M</i>	<i>SD</i>	<i>t</i> ^a	<i>p</i>	<i>d</i>
<i>konsistent</i>	Im SIAB danach befragt, gab die Patientin zu, ab und zu unter ihrem Essverhalten zu leiden.	3.11	1.27	7.35	.00	2.45
	Im SIAB gab Marisa an, dass ihr Selbstvertrauen von ihrer Figur abhängt.	4.22	0.97	13.03	.00	4.35
	Die Patientin war nach Ansicht des Bezugstherapeuten durchaus zu einer Mitarbeit in der Therapie bereit.	2.56	1.24	6.20	.00	2.06
	Die Patientin sagte, dass ihr Selbstwertgefühl von ihrem Körpergewicht abhängig sei.	3.89	1.05	11.07	.00	3.70
<i>Inkonsistent</i>	Mit dem subjective body dimensions apparatus wird die Einschätzung des Ausmaßes einzelner Körperpartien erfasst. Die Patientin schätzte alle zur Beurteilung vorgegebenen Körperteile (Bauch, Arme und Beine) weitaus größer ein, als sie objektiv waren.	-2.78	2.54	-3.28	.01	1.09
	Das Ergebnis des SIAB machte deutlich, dass die Patientin keine Krankheitseinsicht hat.	-1.78	2.68	-1.99	.04	0.66
	Die Antworten im EDI ergaben, dass die Patientin keine Schuldgefühle hat, wenn sie eine große Menge Nahrung zu sich genommen hat.	-3.67	1.23	-8.98	.00	2.98
	Marisas Zahnarzt konnte keine Veränderungen am Zahnschmelz feststellen.	-2.00	1.87	-3.21	.01	1.07

Anmerkungen. Untersucht wurde jeweils die Abweichung vom Skalenmittel 0, welches eine neutrale bzw. irrelevante Einschätzung auf der bipolaren Skala von -5 = inkonsistent bis +5 = konsistent repräsentiert

^a Die zugehörigen Freiheitsgrade betragen jeweils $df = 8$

Tabelle 10 sind einerseits die auswertungsrelevanten Items und andererseits die errechneten Kennwerte zu entnehmen, die die Einteilung als konsistente und inkonsistente Items rechtfertigen, da in allen Fällen die H_0 zu verwerfen war und mit einer einzigen Ausnahme *große* Effekte vorlagen.

Tabelle 11 kann darüber hinaus entnommen werden, dass die Mittelwerte der konsistenten Items sich hinsichtlich Diagnostizität, Wichtigkeit und Zuverlässigkeit nicht von denen der inkonsistenten Items unterscheiden. Die angestrebte Ausbalancierung ist daher als erfolgreich zu bewerten, weil auf allen drei Maßen die H_0 beizubehalten ist und die Effekte bestenfalls als klein zu bezeichnen sind.

Tabelle 11. Untersuchung der Ausbalancierung der konsistenten vs. inkonsistenten Items (Essstörung)

<i>Merkmal</i>	<i>Konsistenz</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>df</i>	<i>p</i>	<i>d</i>
<i>Wichtigkeit</i>	<i>konsistent</i>	68.29	30.24	0.59	69	.56	0.14
	<i>inkonsistent</i>	64.17	28.62				
<i>Zuverlässigkeit</i>	<i>konsistent</i>	80.83	19.48	0.19	70	.85	0.05
	<i>inkonsistent</i>	81.67	17.32				
<i>Diagnostizität</i>	<i>konsistent</i>	33.06	33.79	-1.12	70	.27	0.26
	<i>inkonsistent</i>	42.22	35.79				

Bedingung Fachwissen irrelevant (Brandstiftung). Das Auswahlprozedere war analog zum Essstörungsfall, d.h. es wurde nach jeweils vier konsistenten und inkonsistenten Items gesucht, die sich gleichzeitig nicht hinsichtlich ihrer *Wichtigkeit*, *Zuverlässigkeit* und *Diagnostizität* unterschieden. Wie Tabelle 12 entnommen werden kann, wurden alle entsprechenden Nullhypothesen bei bestenfalls geringen Effekten wie erwartet beibehalten.

Tabelle 12. Untersuchung der Ausbalancierung der konsistenten vs. inkonsistenten Items (Brandstiftung)

<i>Merkmal</i>	<i>Konsistenz</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>df</i>	<i>p</i>	<i>d</i>
<i>Wichtigkeit</i>	<i>konsistent</i>	53.06	25.39	-0.92	70	.36	0.22
	<i>inkonsistent</i>	59.72	35.41				
<i>Zuverlässigkeit</i>	<i>konsistent</i>	84.72	18.28	1.16	70	.25	0.27
	<i>inkonsistent</i>	78.89	24.00				
<i>Diagnostizität</i>	<i>konsistent</i>	33.89	28.01	-0.79	70	.43	0.19
	<i>inkonsistent</i>	39.72	34.52				

Tabelle 13. Begründung der Einteilung in konsistente vs. inkonsistente Items (Brandstiftung)

Ein- teilung	Item	M	SD	t ^a	p	d
konsistent	Eine Probe des Teppichs aus der verbrannten Wohnung enthielt Benzin, eine andere Teppichprobe und Holzsplitter aus der Wohnung enthielten Verdünnungsmittel. Zwei unabhängige Spezialisten bestätigten dies.	2.44	1.51	4.86	.00	1.62
	Nachdem Herr Wagner am 24. Juni erfahren hatte, dass Frau Schröder eine sexuelle Beziehung mit Herrn Braun hatte, konfrontierte er Herrn Lange und bedrohte ihn mit einem Messer.	2.44	1.24	5.93	.00	1.97
	Ein Rauchmelder war nachweislich von der Decke entfernt worden.	1.89	1.54	3.69	.01	1.23
	Mehrere Psychologen bezeugten, dass Herr Wagner an einer Antisozialen und Borderline-Persönlichkeitsstörung litt. Er beschäftige sich übermäßig mit Tod, Blut und Gewalt sowie mit Selbstverletzung und habe bereits in der Vergangenheit versucht, sich das Leben zu nehmen. Der Psychiater Dr. Kröger gab an, er schätze Herrn Wagner als extrem impulsiv, selbstzentriert und hedonistisch ein.	2.11	1.36	4.64	.00	1.55
inkonsistent	Frau Bauer gab an, Herr Wagner habe regelmäßig auf Katharina Acht gegeben.	-1.33	1.58	-2.53	.02	0.84
	Weder auf Herrn Wagners Kleidung noch auf seinen Stiefeln oder dem Verband, den er an diesem Abend getragen hatte, konnten Spuren von leicht entflammaren Substanzen wie z.B. Benzin gefunden werden.	-4.00	1.12	-10.73	.00	3.57
	In den Wochen vor dem Brand wurde der Feuerwehrhauptmann aufgrund von Rauchbildung aus unbekannter Quelle drei Mal zur Wohnung gerufen. Dies war bereits einmal geschehen, bevor Herr Wagner eine sexuelle Beziehung mit Frau Schröder einging.	-3.00	0.87	-1.39	.00	3.45
	Zeugen sagten aus, dass Herr Wagner, nachdem er von der neuen Beziehung seiner Ex-Freundin erfahren hatte, nach einer Diskussion mit dieser und Herrn Lange die Beziehung akzeptierte. Bei der Beschlagnahme seines Tagebuches fand die Polizei einen Eintrag, der dies bestätigte.	-3.44	1.33	-7.75	.00	2.59

Anmerkungen. Untersucht wurde jeweils die Abweichung vom Skalenmittel 0, welches eine neutrale bzw. irrelevante Einschätzung auf der bipolaren Skala von -5 = inkonsistent bis +5 = konsistent repräsentiert

^a Die zugehörigen Freiheitsgrade betragen jeweils $df = 8$

Die Items selbst sind in Tabelle 13 veranschaulicht und hinsichtlich ihrer Einteilung in konsistente vs. inkonsistente Items begründet, indem alle Einschätzungen signifikant und mit ausnahmslos großen Effekten vom neutralen Skalenmittelpunkt abweichen.

Auswertungsrelevante Parameter. In Bezug auf die *Wichtigkeit* wird für jeden Probanden zunächst die Summe der angegebenen Wichtigkeiten für die konsistenten einerseits und inkonsistenten Items andererseits berechnet. Der Wichtigkeitsindex wird anschließend als Differenz der Wichtigkeit der konsistenten minus der inkonsistenten Items gebildet. Positive Werte zeigen die Bestätigungstendenz B an, ein negativer Wert die höhere Gewichtung der inkonsistenten Informationen.

Für den *Einfluss auf die Bewährung der Prüfhypothese* werden die Einschätzungen der Probanden von den konsistenten und inkonsistenten Items addiert. Durch die bipolare Skala sollte sich im unverzerrten Falle die Summe null ergeben. Werte größer null zeigen hingegen die Bestätigungstendenz B an.

Die übrigen Items, die vorab als irrelevant für die Diagnosestellung eingeschätzt worden sind und im Wesentlichen die Funktion von Füllitems zur Verschleierung der Messintention erfüllen, bleiben bei der Auswertung unberücksichtigt.

4.2.5 AV 5: *Hypothesenkonsistente Interpretation uneindeutiger Evidenz und Bestätigungstendenz I*

Zur Erfassung einer verzerrten Informationsbewertung bei uneindeutigen Items wurden die Probanden nach der Bewertung der oben dargestellten 18 Items zur Bestätigungstendenz B vor die Aufgabe gestellt, insgesamt 12 weitere Fallinformationen hinsichtlich ihres Wertes zu beurteilen. Hierfür wurde ihnen immer eine einzelne Information eingeblendet, die sie einer von vier Kategorien in Bezug auf ihre Bedeutung für die Prüfhypothese zuordnen mussten. Eine Information konnte als *eindeutiger Beleg*, als *brauchbarer Hinweis*, als *Widerspruch* oder als *irrelevante Information* eingeordnet werden.

Ergänzend ist anzumerken, dass auch hier alle Probanden die Informationen in Bezug auf Bulimie beurteilten. Um zu verschleiern, dass das Forschungsinteresse ausschließlich der Abhängigkeit ihres Vorgehens von dieser Prüfhypothese galt, wurden die Probanden im Zuge der Instruktion jedoch darüber getäuscht, dass die sich anschließende Bewertung von verschiedenen Teilnehmern aus unterschiedlichen Blickwinkeln vorgenommen werde und das Computerprogramm nun zufällig eine Perspektive für sie auswählen würde. Allen Teilnehmer wurde nach einem Wartebildschirm hingegen vorgegeben, den Blickwinkel *Bulimie* bei ihrer Bewertung einzunehmen.

Die Items wurden dem oben bereits beschriebenen Itempool entnommen. Für die Auswahl war zunächst darauf zu achten, dass einerseits *uneindeutige* Informationen enthalten sind, um die Analyse der Variablen *hypotheseinkonsistente Interpretation uneindeutiger Evidenz* zu ermöglichen. Andererseits mussten auch *prüfhypotheseninkonsistente* Informationen ausgewählt werden, um die *Bestätigungstendenz I* zu untersuchen, die sich dadurch auszeichnet, dass Widersprüche nicht als solche erkannt bzw. bewertet werden.

Bedingung Fachwissen relevant (Essstörung). Die 8 *uneindeutigen* Items im Essstörungskontext sind zur Hälfte solche Informationen, die von den Ratern vorab mehrheitlich als *brauchbare Hinweise* bzw. *irrelevant* eingeschätzt wurden. Bei den *irrelevanten* Items ist dies unmittelbar einleuchtend, denn ohne einen Zusammenhang mit der Prüfhypothese kann auch kein Beitrag zur Bewertung derselben erbracht werden. Aber auch die *brauchbaren Hinweise* sind aus sachlogischen Gründen *uneindeutig*, weil sie keine notwendige Bedingung für das Zutreffen der Prüfhypothese darstellen, sondern nur typischerweise bzw. mit bestimmter Wahrscheinlichkeit kovariieren. Ein valider Rückschluss auf die Prüfhypothese ist dementsprechend nicht möglich, was als Uneindeutigkeit einzustufen ist.

Die vier prüfhypotheseninkonsistenten Items sind genau solche, die von den Ratern mehrheitlich als *Widerspruch* eingestuft wurden.

Die Korrektheit bzw. Verzerrung der Urteile seitens der Probanden wurde nun durch die Analyse von Abweichungen zu den vorab von den Ratern erhobenen Klassifikationen bestimmt. Eine korrekte Einschätzung lag demnach vor, wenn Rater und Proband bei einem Item übereinstimmten. Eine Verzerrung im Sinne der AV 5a (*hypotheseinkonsistente Interpretation uneindeutiger Evidenz*) lag dann vor, wenn ein Proband aus der Gruppe der acht *uneindeutigen* Items entweder eine *irrelevante* Information als *brauchbaren Hinweis* (ggf. sogar *eindeutigen Beleg*) wertete oder einen *brauchbaren Hinweis* als *eindeutigen Beleg*.

Eine Verzerrung im Sinne der AV 5b (*Bestätigungstendenz I*) liegt vor, wenn ein Proband einem seitens der Rater als *Widerspruch* klassifizierten Merkmal eine der anderen drei Kategorien zuweist. Die entsprechende Anzahl vorgefundener Aufwertungen wird im Anschluss in Abhängigkeit von den Einflussvariablen analysiert.

Die Items sowie deren Klassifikationen sind Tabelle 14 zu entnehmen.

Tabelle 14. Übersicht über die (Reihenfolge der) verwendeten Items für AV 5 (Essstörung)

Item	Relevanz	Übereinstimmung
Die Patientin hatte kein Normalgewicht, denn bei einer Körpergröße von 1,75m und einem Gewicht von 50 kg lag ihr BMI bei 16,3.	Widerspruch	9/9
Im CFT-20-R zeigte sich eine überdurchschnittliche Intelligenz.	Irrelevant	9/9
Laut Mutter tritt bei der Patientin Angst auf, weiter zuzunehmen, wenn sie ein Pfund zugenommen habe.	Brauchbarer Hinweis	5/9
Die Patientin wurde sehr ungehalten, als sie befürchtete, ihr Tagebuch auf der Station verloren zu haben.	Irrelevant	9/9
Gefragt danach, über welches eigene Verhalten sie sich am meisten ärgere, berichtete sie, dass sie ein Angebot über einen Auslandsaufenthalt abgelehnt habe, um bei ihrem Freund zu bleiben.	Irrelevant	5/9
Mit der Patientin wurde ein Computer Body Image Test durchgeführt. Bei diesem Verfahren können die Patienten am Computer bei einem Foto von sich selbst die Körperdimensionen so verändern, dass es ihrem eigenen Bild von sich selbst entspricht. Die Patientin schätzte ihre Körperdimensionen weitaus größer ein, als sie sind. Das heißt, dass sich eine bedeutsame Differenz zwischen den wahrgenommenen und den tatsächlichen Proportionen ergibt.	Widerspruch	6/9
Die Patientin war unzufrieden mit ihrem Essverhalten.	Brauchbarer Hinweis	8/9
Die Patientin war nach Angabe der Kunsttherapeutin nicht bereit, in der Therapie mitzuarbeiten.	Widerspruch	6/9
Die Patientin glaubte nicht, ihrem Körper durch ihr Essverhalten zu schaden.	Widerspruch	5/9
Über die Geschwisterbeziehung sagte die Mutter, dass die beiden die üblichen Streitigkeiten untereinander hätten, aber sich in jugendtypischer Weise gegen die Eltern verbündeten.	Irrelevant	9/9
Große Mengen an Nahrung nahm die Patientin nur dann zu sich, wenn sie alleine war.	Brauchbarer Hinweis	6/9
Die Patientin war gedanklich häufig mit Diäten beschäftigt.	Brauchbarer Hinweis	7/9

Anmerkungen. Die Übereinstimmung der Rater ist in der Form X übereinstimmende / alle 9 Rater angegeben.

Bedingung Fachwissen irrelevant (Brandstiftung). Aus Gründen der methodischen Konsistenz muss sich die Bedingung *Fachwissen irrelevant (Brandstiftung)* derjenigen der Bedingung *Fachwissen relevant (Essstörung)* maximal anpassen. Weil für den Bereich klinisch-psychologischer Diagnostik die Bedeutsamkeit der Unterscheidung des Wertes von Informationen in *eindeutige Belege, brauchbare Hinweise, Widersprüche* und *irrelevante Informationen* an anderer Stelle bereits erbracht wurde (s. Abschnitt 4.1.3.1.3), musste dieser in den Brandstiftungsfall übertragen werden, wenngleich die inhaltliche Begründung entfällt, da eine analoge Unterscheidung zwischen *eindeutigen Belegen* und *brauchbaren Hinweisen* nach Ansicht der Verfasserin nicht valide möglich ist, weil kein Analogon zu Klassifikationssystemen wie dem DSM-IV (Saß et al., 1998) vorliegt. Selbst wenn die Rater also eine Information im Brandstiftungsfall übereinstimmend als *brauchbaren Hinweis* wahrnahmen und beurteilten, so entbehrt dies doch einer sachlogischen Grundlage. Aus Sicht der Verfasserin verbietet es sich daher, *brauchbare Hinweise* im Brandstiftungskontext als *uneindeutig* zu klassifizieren bzw. Probanden, die eine solche Information als *eindeutigen Beleg* klassifizieren ein aufwertendes Vorgehen zu unterstellen.

Statt der vier brauchbaren Hinweise, die im Essstörungsfall zum Einsatz kamen, werden weitere vier *irrelevante* Informationen gesucht, um analog zum Bulimiekontext acht Möglichkeiten einer Aufwertung zu präsentieren. Die konkreten Items sind in Tabelle 15 veranschaulicht.

Analog zur Bedingung *Fachwissen relevant (Essstörung)* liegt eine Verzerrung im Sinne der AV 5a (*hypotheseinkonsistente Interpretation uneindeutiger Evidenz*) dann vor, wenn ein Proband aus der Gruppe der acht *uneindeutigen* Items eine irrelevante Information als *brauchbaren Hinweis* (ggf. sogar *eindeutigen Beleg*) wertete. Eine Verzerrung im Sinne der AV 5b (*Bestätigungstendenz I*) liegt vor, wenn ein Proband einem seitens der Rater als *Widerspruch* klassifizierten Merkmal eine der anderen drei Kategorien zuweist.

Tabelle 15. Übersicht über die (Reihenfolge der) verwendeten Items für AV 5 (Brandstiftung)

<i>Item</i>	<i>Relevanz</i>	<i>Übereinstimmung</i>
Als Frau Bauer Nachricht von dem Feuer und dem Tod ihrer Tochter bekam, machte sie keine Angaben bezüglich eines Babysitters, auch fragte sie nicht bei der Polizei nach dessen Verbleib. Erst später gab sie an, Herr Wagner hätte eingewilligt, auf Katharina in dieser Nacht aufzupassen.	Irrelevant	5/9
Frau Schröder wurde von mehreren Zeugen als promiskuitiv beschrieben.	Irrelevant	9/9
Während Herr Wagner in Untersuchungshaft saß, schrieb er einem Freund in Schottland: „Erinnerst du dich an den Tag an dem ich Greg mit deiner Waffe in den Kopf geschossen habe? Ich habe mich fast tot gelacht. Wenn die Polizei [in Schottland] auch nur die Hälfte von den Sachen herausfindet, die wir getan haben, führen sie für uns die Todesstrafe wieder ein.“	Irrelevant	6/9
Herr Wagner nahm am Trauergottesdienst für Katharina teil.	Irrelevant	6/9
Frau Bauer war von dem Verlust ihrer Tochter Katharina tief erschüttert und nahm eine Woche später psychologische Betreuung in Anspruch.	Irrelevant	9/9
Herr Wagner gab in einem polizeilichen Interview an, dass er an dem Abend des Brandes in die Wohnung seines Vaters gegangen sei und dort verblieben wäre, bis er zum Einsatzort gekommen sei.	Widerspruch	8/9
Der Eigentümer des Gewächshauses konnte nicht sagen, ob etwas von seinen Vorräten an Benzin und Verdünnungsmittel fehlte.	Widerspruch	5/9
Frau Schröder und Herr Lange gaben beide an, sie seien leichte Schläfer, hätten in den Stunden vor dem Brand aber keinerlei Geräusche gehört. Das Schlafzimmerfenster war offen und weniger als zwei Meter von dem Elektrizitätskasten entfernt, an dem Herr Wagner hochgeklert sein soll.	Widerspruch	8/9
2007 wiesen führende Wissenschaftler der USA darauf hin, dass die ursprünglichen Brandschuttanalysen völlig unzuverlässig waren. Von ihnen durchgeführte Untersuchungen führten zu der Erklärung, dass man keine Aussage darüber treffen könne, ob die Proben Spuren von brennbaren Substanzen enthielten oder nicht.	Irrelevant	5/9
Einer von Herr Wagners früheren Lehrern sagte aus, dass er ein aufmerksamer und fleißiger Schüler gewesen sei.	Irrelevant	9/9
Zwei Experten führten die Untersuchungen der Brandschuttproben durch. Dabei benutzte Herr Roth eine Untersuchungsmethode, die wissenschaftlich nicht anerkannt war und von einer internationalen Autorität in forensischen Fragen kritisiert wurde. Der zweite hinzugezogene Experte war Herr Vogel. Dieser hatte zu diesem Zeitpunkt ein viertägiges Training in der forensischen Untersuchung von Brandstiftung absolviert und wurde angewiesen seine Untersuchung auf zehn Stunden zu begrenzen, um Geld zu sparen. Er wandte keine eigenen Testverfahren an und bestätigte das Urteil von seinem Kollegen Herrn Roth, das für Brandstiftung sprach.	Irrelevant	5/9
Als Herr Wagner an den Einsatzort ankam, versuchte er wiederholt in das brennende Gebäude einzudringen - nach eigener Aussage, um Katharinas Leben zu retten. Die Polizei hielt ihn jedoch davon ab.	Widerspruch	5/9

Anmerkungen. Die Übereinstimmung der Rater ist in der Form X übereinstimmende / alle 9 Rater angegeben.

4.2.6 AV 6: Verhaltenskonsequenzen

Für die Bedingung *Fachwissen relevant (Essstörung)* wird zudem in Anlehnung an Haverkamp (1993) eine AV erhoben, die (potentielle) Verhaltenskonsequenzen der Diagnosestellung erfassen soll. Konkret werden die Probanden aufgefordert anzugeben, welche *Themen* sie mit der Patientin in der ersten Therapiesitzung ansprechen würden und welche *Hausaufgabe* sie ihr auftragen würden. In der Bedingung *Fachwissen irrelevant (Brandstiftung)* gibt es kein Äquivalent, weil es keine der logischen Natur der Arbeit eines Schöffen folgende Aufgabe ist, Vorschläge für den Umgang mit dem Verurteilten vorzulegen.

Da die Probanden aufgefordert waren, Freitextantworten zu generieren, mussten diese vor der quantitativen Analyse zunächst kodiert werden. Zwei unabhängige und hypothesenblinde Rater bewerteten die Angaben der Probanden dahingehend, ob das angesprochene Thema bzw. die angesprochene Hausaufgabe *eher für anorektische Patienten* oder *eher für bulimische Patienten* angewendet wird oder *unspezifisch für beide Störungen* ist. Sollten eher Maßnahmen für bulimische Patienten eingesetzt werden, entspricht dies einer Falschbehandlung, die im Ernstfall bei einer Anorexie sogar tödliche Folgen haben könnte, da die Mortalität hier stark erhöht ist.

4.3 Störvariablen

Im Folgenden wird die Umsetzung der Ausschaltung bzw. Kontrolle der inhaltlichen Störvariablen beschrieben.

Bewusstsein hinsichtlich confirmatorischer Prozesse. Diese Störvariable ist auszuschalten, was durch ein mehrfach gestuftes Vorgehen gewährleistet wird. Erstens wird das Studienziel allgemein als *Optimierung psychologischer Urteilsfindungsstrategien* im Vergleich von erfahrenen und unerfahrenen Psychologen angegeben, ein direkter Bezug zum *confirmation bias* wird nicht hergestellt. Darüber hinaus wird zweitens ein *suspicion check* durchgeführt, d.h. zum Abschluss der Studie wurden alle Probanden aufgefordert „*Bitte geben Sie zum Schluss noch einmal in eigenen Worten an, was das Ziel dieser Studie ist!*“ Auffällige Antworten führen zum Ausschluss des Probanden von der Auswertung, was von zwei unabhängigen Ratern bewertet wird. Die Instruktionen sind in Anhang C einsehbar.

Der diagnostische Prozess als soziale Interaktion. Dieser Störvariablen kann nur durch Ausschaltung begegnet werden, weswegen das vorliegende Experiment als Online-Studie konzipiert wurde. Eine genauere Begründung dieser Wahl als Alternative zu einer postalischen Befragung als paper-pencil Erhebung wird im folgenden Abschnitt vorgenommen, wo gleichermaßen die mit der Wahl der Onlinebefragung verbundenen methodischen Störvariablen abzuhandeln sind.

4.4 Begründung der Wahl einer Online-Studie

Die Wahl der Umsetzung als Onlinestudie ist hauptsächlich dadurch motiviert, dass auf diese Weise unter vertretbarem Einsatz von Ressourcen die Stichprobe maximiert werden kann.

Aufgrund der infrastrukturellen Situation schien es unrealistisch, in der näheren Umgebung der Universität Kiel aus den niedergelassenen Psychotherapeuten 80 Personen für eine Teilnahme zu gewinnen, um diese in einem realen Setting, d.h. an ihrem Arbeitsplatz, an der Studie teilnehmen zu lassen, wie Strohmer und Shivy (1994) dies umsetzten – wobei sie auch lediglich $n = 20$ Probanden aufsuchten. Da außerdem dieser Studie zu entnehmen war, dass der *confirmation bias* trotz dieser vergrößerten Realitätsnähe auftrat und andererseits im Rahmen der vorliegenden Arbeit argumentiert wurde, dass die soziale Interaktion mit dem Patienten in ihrer Gesamtheit als Störvariable aufzufassen ist (s. Abschnitt 3.3) wurde der Einsatz eines *mock patient* grundsätzlich ausgeschlossen. Die Probanden an ihrem Arbeitsplatz schriftliche Fallmaterialien bearbeiten zu lassen, schien außerdem kein anzustrebendes Ziel im Sinne des Erkenntnisgewinns, so dass das Fehlen eines Versuchsleiters, welches mit der Realisierung einer postalischen oder webbasierten Befragung verbunden ist, in Kauf genommen wurde, zumal es in der Untersuchung von Expertengruppen nicht unüblich ist, postalische (und mit der Entwicklung des Internets auch webbasierte) Befragungen unter der Verwendung von schriftlichem Fallmaterial durchzuführen. Eine auffällige Abweichung vom üblichen Vorgehen kann daher nicht festgestellt werden.

Ein zweiter Aspekt spielte eine Rolle: es ist das erklärte Ziel der vorliegenden Arbeit, einen vollständigen Prozess der Hypothesenprüfung von der Generierung, über das Aufstellen von Evidenzregeln, das Suchen und Bewerten von Informationen bis hin zur Entscheidung über das Zutreffen der Hypothese zu untersuchen. Dies hat notwendigerweise einen komplexen Ablauf zur Folge, der sehr zeitintensiv ist.

Es wird eine Bearbeitungsdauer von ca. einer Stunde angenommen. V.a. innerhalb der Teilstichprobe der Psychotherapeuten ist damit zu rechnen, dass dieser Zeitaufwand eine abschreckende Wirkung hat, da die Ressource Zeit gewöhnlich knapp ist und daher eine eher geringe Teilnahmebereitschaft erwartet wird bzw. dass eine hohe Anzahl von Abbrüchen aus zeitlichen Gründen angenommen werden muss. Die Umsetzung als Onlinestudie bot hierbei die Möglichkeit, ohne größere finanzielle Ressourcen eine große Anzahl von Probanden per Email anzuwerben und v.a. um die Weiterleitung der Einladungen an interessierte Kollegen zu bitten, was im Falle der Zusendung schriftlicher Unterlagen auf den Postweg unmöglich ist und durch Kopier- und Portokosten nicht unerhebliche finanzielle Mittel verschlungen hätte. Außerdem wäre dieses Prozedere nur für die Psychotherapeuten denkbar gewesen. Adressen der Psychologiestudierenden hätten nicht ermittelt und eine andere Übermittlungsweise gefunden werden müssen. Methodische Unterschiede zwischen den Erfahrenen und Unerfahrenen wären dann jedoch nicht auszuschließen gewesen.

Die Wahl fiel daher auf eine Onlinerealisierung, die keinen dieser Nachteile aufwies und es zudem ermöglichte, die Bearbeitungszeit der Probanden zu erfassen. Außerdem war so weitestgehend das Auftreten von fehlenden Werten vermeidbar, da die Probanden nur dann zur nächsten Aufgabe weiterklicken konnten, wenn alle Felder ausgefüllt waren. Ein Vor- und Zurückblättern war zudem unmöglich, so dass keine Interferenzen denkbar sind.

Neben den typischen Selbstselektionseffekten sind bei der Interpretation allerdings auch Besonderheiten hinsichtlich des Internetnutzungsverhaltens zu beachten, weil durch diese Realisierung u.U. bestimmte Teilgruppen der Gesamtstichprobe unterschiedlich stark angesprochen werden.

4.5 Studienablauf

An dieser Stelle soll erstens die technische Umsetzung der Online-Studie und zweitens die Rekrutierung beschrieben werden. In einem dritten Unterpunkt wird der Ablauf der Studie aus der Sicht der Probanden ausführlich erläutert.

Technische Umsetzung der Online-Studie. Die technische Umsetzung des Studienkonzepts erfolgte durch einen Informatiker. Bei dem verwendeten Programm handelt es sich um das von ihm entwickelte *SciPoll_FX* in der Version 0.13 (Dreger, 2008). Die technische Basis der Programmanwendung stellt die Skriptsprache Actionscript 3.0

dar. Mit Hilfe des Adobe Flash Player Plugins für den Webbrowser wurde der Programmcode ausgeführt und dem Anwender auf dem Bildschirm präsentiert. Durch die Verwendung eines Produktes auf Basis der Adobe Flash Technologie konnte sichergestellt werden, dass ein möglichst großer Anteil der eingeladenen Probanden über die technischen Voraussetzungen zur Durchführung verfügte, da aus den Herstellerangaben (Adobe Systems Incorporated, 2012) abzuleiten war, dass das Flash-Plugin auf mindestens 98% der PCs im Zielgebiet und zum Zeitpunkt der Studiendurchführung in einer geeigneten Softwareversion (bei Studiendurchführung mindestens Version 9) installiert ist.

Alle Eingaben wurden unmittelbar in einer Datenbank gespeichert und konnten später direkt in das Statistikprogramm importiert werden.

Für den Anwender liefen alle technischen Schritte unsichtbar im Hintergrund ab. Ihm bot sich lediglich ein typisches Browser-Fenster, in dem er durch Anklicken oder Eintippen kurzer Sätze agierte. Andere Anforderungen wurden nicht gestellt, so dass angenommen werden kann, dass ein Internetnutzer dies ohne Probleme bewältigen kann.

Rekrutierung. Die Rekrutierung der Studienteilnehmer erfolgte i.d.R. via Email und einem angehängten Einladungsschreiben (s. Anhang C). Sie erhielten Informationen über das Forschungsprojekt, die ungefähre Dauer der internetbasierten Befragung, den Hyperlink zum Starten der Anwendung sowie ein Passwort, welches auf der Startseite einzugeben war, damit keine Personen, die rein theoretisch zufällig die Internetseite aufrufen könnten, an der Befragung teilnehmen konnten. Desweiteren wurde ihnen mitgeteilt, dass die Verfasserin für jeden vollständig ausgefüllten Fragebogen 2,50 € an eine gemeinnützige Stiftung (konkret: Stiftung Kinderherz) spendet.

Damit den Studierenden vs. Psychotherapeuten völlig unterschiedliche Probandengruppen anzusprechen waren, mussten zwei unterschiedliche Rekrutierungsstrategien eingesetzt werden. Um unter den *Psychologiestudierenden* in Deutschland für die Studie zu werben, wurden alle Institute für Psychologie im Inland angeschrieben. Die Bitte um Unterstützung ging einerseits an den jeweiligen Lehrstuhlinhaber für Psychologische Diagnostik bzw. ersatzweise Klinische Psychologie. Andererseits wurden die Fachschaften angeschrieben. Die Angesprochenen erhielten das Anschreiben mit der Bitte um Verbreitung in Form von Aushängen, Forenbeiträgen oder der mündlichen Ankündigung in Vorlesungen. Außerdem wurde ein Beitrag zur Studie in den Newsletter der Bundesvereinigung der Psychologiestudierenden im BDP aufgenommen. Zur Rekrutierung der *Berufserfahrenen* wurden einerseits die Psychotherapeutenkammern angeschrieben und andererseits gingen die Einladungsschreiben direkt an diejenigen in Deutschland tätigen Psychotherapeuten, die ihre Email-Adresse in Psychotherapeutendatenbanken gelistet haben.

Um eine potentielle Überlastung des Servers auszuschließen, wurden die Einladungen sukzessive im Zeitraum von Oktober 2008 bis Mitte Januar 2009 verschickt. Ende Januar 2009 war die Datenerhebung abgeschlossen.

Studienablauf aus Sicht der Probanden. Aus Sicht des Probanden begann die Studie mit dem Aufrufen des in der Einladung enthaltenen Hyperlinks, wo ihm zunächst für sein Interesse an der Untersuchung gedankt und das Akronym OPUS – Optimierung Psychologischer Urteilsfindungs-Strategien - vorgestellt wurde. Außerdem mussten die Probanden sich auf der Startseite einer der vorgegebenen Kategorien entsprechend ihres beruflichen Status zuordnen (z.B. Psychologiestudierender vs. Psychologischer Psychotherapeut). Nach Informationen zum Studieninhalt, Ablauf, Kosten und Nutzen sowie der Vertraulichkeit der Daten wurden die Teilnehmer um die Einwilligung oder die Ablehnung der Teilnahme gebeten. Teilnahmebereite Personen erhielten auf der Folgeseite technische Hinweise zum Bearbeiten des Fragebogens.

Es folgte die bedingungsabhängige Präsentation von Hintergrundinformationen zur Sicherstellung des beurteilungsrelevanten Wissens und anschließend die Vorgabe der Fallbeschreibung. Beide Informationstexte konnten zu jedem Zeitpunkt der Studie erneut und so lange wie gewünscht geöffnet werden. Unmittelbar nach dem Lesen der Fallbeschreibung wurden die Probanden gebeten, ihre subjektive Überzeugtheit von der jeweiligen Prüfhypothese anzugeben, bevor die Erfassung der AV 3, d.h. die Generierung der sechs Evidenzregeln erfolgte.

Daran schloss sich die Präsentation und Bewertung von weiteren konsistenten wie inkonsistenten Fallinformationen hinsichtlich ihrer Wichtigkeit und Bekräftigung vs. Abschwächung der Prüfhypothese an (AV 4).

Hiernach wurden sie gebeten, sich von ihrem individuellen Blickwinkel zu lösen und weitere zwölf Informationen durch eine vom PC scheinbar zufällig ausgewählte Perspektive zu betrachten. Ein grafisches Element, das gängigen Wartesymbolen bei Ladevorgängen im Internet nachempfunden war, zeigte den scheinbaren Arbeitsprozess des PC bei der Auswahl an, bevor ein Dialogfenster den Probanden (ausnahmslos) mitteilte, dass sie den Blickwinkel der Prüfhypothese einnehmen sollten. Für die AV 5 erfolgte nun für jede Information isoliert die Kategorisierung ihrer Aussagekraft als *eindeutiger Beleg, brauchbarer Hinweis, Widerspruch* oder *irrelevante Information*.

Die zwölf Informationen machten die Falldarstellung komplett, so dass im Anschluss das endgültige Urteil (AV 1) und die entsprechende subjektive Sicherheit (AV 2) abgefragt werden konnten. Die Probanden der Bedingung *Fachwissen relevant (Essstörung)* bearbeiteten nun auch noch die Items zu dem Thema der ersten Therapiesitzung und den

Hausaufgaben (AV 6), während die Probanden der Bedingung *Fachwissen irrelevant (Brandstiftung)* gleich zur Bearbeitung der biographischen Fragen übergingen. Da nicht anzunehmen ist, dass die unterschiedliche Bearbeitungsdauer einen Effekt auf die Beantwortung der Fragen zum persönlichen und beruflichen Kontext hat, wurde diese Abweichung zwischen den Bedingungen an dieser späten Stelle im Versuchsablauf in Kauf genommen.

Die Fragebögen zu den biografischen Daten waren entsprechend der Selbstzuordnung zu den Ausbildungskategorien Psychologischer Psychotherapeut, Student etc. angepasst. Sie sind in Anhang C aufgeführt.

Abschließend wurde den Teilnehmern für ihren Einsatz gedankt und auf die Erhöhung des Spendenzählers aufgrund ihrer vollständigen Bearbeitung hingewiesen. In einem letzten Schritt wurden weitere Informationen über die Studienhintergründe, d.h. in Bezug auf ein konfirmatorisch verzerrtes Urteilen gegeben.

5 Ergebnisdarstellung

Im Folgenden werden die Auswertung und die resultierenden Ergebnisse beschrieben. Nach einigen Vorbemerkungen zum allgemeinen Vorgehen bei der Auswertung, werden zunächst notwendige Voranalysen (Abschnitt 5.2) berichtet, bevor die Dokumentation der hypothesengeleiteten Auswertung erfolgt (Abschnitt 5.3).

5.1 Vorbemerkungen

Die Vorbemerkungen sollen Transparenz hinsichtlich getroffener Entscheidungen und verwendeter Beurteilungsmaßstäbe schaffen.

5.1.1 Eingesetzte Auswertungsverfahren

Die Wahl der Auswertungsverfahren für die Prüfung der psychologischen Hypothesen orientiert sich in erster Linie an dem Skalenniveau der bedingten Variablen. Sollen Hypothesen über die Abhängigkeit einer nominalskalierten (konkret dichotomen) AV geprüft werden, kommt die logistische Regression zum Einsatz, die die gleichzeitige Berücksichtigung von kategorialen und metrischen Vorhersagevariablen ermöglicht. Liegt hingegen eine metrische AV vor, wird zur simultanen Untersuchung der drei bedingenden Variablen die Kovarianzanalyse verwendet.

Logistische Regression. Der logistischen Regression wird dem Einsatz mehrerer unverbundener χ^2 -Tests auf Gleichverteilung Vorzug gegeben, weil diese die simultane Überprüfung mehrerer Einflussvariablen ermöglicht und der χ^2 -Test zudem keine Analyse der metrischen Einflussvariable *Überzeugtheit von der Prüfhypothese* erlauben würde. Backhaus, Erichson, Plinke und Weiber (2008) nennen als Voraussetzungen die *Unkorreliertheit der Residuen* und eine *weitestgehende Abwesenheit der Multikollinearität* der Prädiktoren. Darüber hinaus wird pro Kategorie des Kriteriums eine Fallzahl größer 25 verlangt, die im vorliegenden Fall erfüllt ist (s. Abschnitt 5.2.2), ebenso wie die *Unabhängigkeit der Beobachtungen*, die aufgrund versuchsplanerischen Handelns als gegeben angesehen wird.

Kovarianzanalyse. Die Kovarianzanalyse ermöglicht es, einerseits den Zusammenhangshypothesen bzgl. der metrischen Variable *Überzeugtheit von der Prüfhypothese* und andererseits den Unterschiedshypothesen bzgl. der Einflussvariablen

Berufserfahrung und *Fachwissen* gerecht zu werden, da nach einer Regression der jeweiligen AV auf die *Überzeugtheit von der Prüfhypothese* eine zweifaktorielle Varianzanalyse der Regressionsresiduen vorgenommen wird. Nach Bortz und Schuster (2010) sowie Backhaus et al. (2008) handelt es sich bei der Kovarianzanalyse um ein sehr robustes Verfahren, das bei hinreichend großen Stichproben, genannt wird u.a. der Richtwert $n = 40$ und gleicher Zellbesetzung sowohl Verletzungen hinsichtlich der *Normalverteilung* und *Varianzhomogenität* als auch der *Homogenität der Steigungen* in den Innerhalb-Regressionen ohne nennenswerte Einflüsse auf den Fehler 1. Art und die Teststärke toleriert. Backhaus et al. (2008) führen weiter aus, dass die regressionsbedingten Voraussetzungen an *unkorrelierte Residuen* nur im Falle von *Zeitreihenanalysen* relevant sind, die in der vorliegenden Arbeit nicht vorliegen und auch eine eventuelle Nicht-Normalverteilung der Störgrößen bei Stichproben ab 40 Beobachtungen tolerabel ist. Um die Robustheit für die vorliegende Arbeit weitgehend geltend zu machen, wird daher auf eine Gleichverteilung von je 40 Probanden auf eine der vier Bedingungskombinationen geachtet. Aufgrund dieser Tatsache kann die Kovarianzanalyse jeweils vorbehaltlos angewendet werden, eine Prüfung der Voraussetzungen ist aufgrund der zu unterstellenden Robustheit entbehrlich.

Bortz und Schuster (2010) sowie Backhaus et al. (2008) fordern allerdings übereinstimmend, dass eine *signifikante Korrelation der Kovariaten und der bedingten Variablen* besteht, damit die Bereinigung der Fehlervarianz nicht auf Zufallseffekte zurück zu führen ist. Deshalb wird stets die Korrelation der *Überzeugtheit von der Prüfhypothese* mit der AV auf der Ebene der Gesamtstichprobe berechnet. Im Falle eines signifikanten Zusammenhangs, schließt sich die Kovarianzanalyse an. Liegt hingegen keine signifikante Korrelation vor, wird ohne Berücksichtigung der Kovariaten eine zweifaktorielle Varianzanalyse bzgl. *Berufserfahrung* und *Fachwissen* berichtet.

5.1.2 Wahl des Signifikanzniveaus

Wird die in den inhaltlichen psychologischen Hypothesen formulierte Annahme durch die statistische Alternativhypothese H_1 repräsentiert, so wird zur Minimierung des Fehlers erster Art das Signifikanzniveau $\alpha = .05$ gewählt. Ist analog eine Beibehaltung der H_0 der hypostasierte Zustand, so soll zur indirekten Minimierung des Fehlers zweiter Art das Signifikanzniveau $\alpha = .20$ gesetzt werden. Sofern nicht anders angegeben, ist immer die Annahme der Alternativhypothese intendiert. Es wird keine Alpha-Adjustierung vorgenommen, was bei der Interpretation zu berücksichtigen ist.

5.1.3 Bericht von Effektgrößen und p-Werten

Die Verfasserin schließt sich der Auffassung der DGPs (2007) an, dass der Bewährungsgrad der psychologischen Hypothesen durch das Ausmaß repräsentiert werden soll, indem die in der AV gefundene Gesamtvariation von den Einflussvariablen abhängt.

Daher werden die jeweils verfahrensabhängigen Effektgrößen angegeben. In Tabelle 16 ist für die verwendeten Maß des Effekts wiedergegeben, wie sie bestimmt wurden und welche Werte konventionell verbal als *kleiner*, *mittlerer* und *großer* Effekt umschrieben werden. Die Maße des Effekts wurden mit *IBM SPSS Statistics* (Version 19) oder *G-Power* (Version 3.1.3; Faul, Erdfelder, Lang & Buchner, 2007) berechnet.

Tabelle 16. Übersicht über die verwendeten Maße des Effekts und deren verbale Klassifizierung

Maß für die Effektgröße	Findet Anwendung im Zusammenhang mit	Formel	Kleiner Effekt	Mittlerer Effekt	Großer Effekt
f^2	(Ko-)Varianzanalyse	$\frac{\eta^2}{1 - \eta^2}$.01 ¹	.06	.16
d	t-Test, one sample	$\frac{M - c}{s}$.20 ¹	.50	.80
	t-Test, independent samples	$\frac{M_1 - M_2}{s}$.20 ¹	.50	.80
r	Korrelationen	r	.10 ¹	.30	.50
Nagelkerke-R ²	Logistische Regression	$\frac{1 - \left[\frac{L_0}{L_V} \right]^{\frac{2}{K}}}{1 - L_0^K}$.20 ²	.40	.50
w^2	X ² -Tests	$\sum_{i=1}^k \frac{(\pi_{1i} - \pi_{0i})^2}{\pi_{0i}}$.01 ¹	.09	.25

¹ Empfehlung von Cohen (1988)

² Formel und Empfehlungen von Backhaus et al. (2008). L_0 als *Likelihood des Nullmodells* gegenüber dem L_V als *Likelihood des vollständigen Modells*. K = Stichprobenumfang

Der p -Wert wird ebenfalls in Anlehnung an die DGPs als deskriptiver Parameter verstanden, der im Falle des Unterschreitens (bzw. auch des Entsprechens) des Signifikanzniveaus α die Annahme der Alternativhypothese anzeigt, während andernfalls

die Beibehaltung der Nullhypothese angezeigt wird. Um sprachlich umständliche Konstruktionen zu vermeiden, wird daher im Sinne der hier dargestellten Definition davon gesprochen, ob ein Test signifikant wird oder nicht. Wie in englischsprachigen Publikationen üblich wird der p -Wert auch angeführt.

5.1.4 Beurteilungsmaßstab für die Übereinstimmung von Raterurteilen

Wann immer die Übereinstimmung von mehreren Beurteilern hinsichtlich eines Urteilsobjekts zu berichten ist, wird neben der Angabe der prozentualen Übereinstimmung als zufallskorrigiertes Maß Cohen's Kappa (κ) dargestellt. Die Beurteilung der Höhe orientiert sich dabei an den Vorschlägen von Wirtz und Caspar (2002). Sie sprechen bei einem $\kappa < 0.4$ von einer *schlechten*, ab einem $\kappa = 0.4$ von einer *akzeptablen*, ab $\kappa = 0.6$ von einer *guten* und ab $\kappa = 0.75$ von einer *sehr guten* Übereinstimmung. Außerdem wird der Ausgang des Hypothesentests mit der zu verwerfenden Nullhypothese, die Übereinstimmungsrate der Beurteilung liege im Zufallsbereich, berichtet. Für die inhaltliche Interpretation maßgeblich ist aber die Höhe des κ -Koeffizienten.

Um der Problematik der Unterschätzung der Beurteilerübereinstimmung durch κ im Falle von asymmetrischen Randverteilungen entgegenzuwirken, wird in konkreten Fällen als alternative Schätzung der Übereinstimmung zusätzlich der *random error (RE)* von Maxwell (1977) berichtet, der anders als κ nicht die Randhäufigkeiten der Merkmale zur Schätzung der zufälligen Übereinstimmungen heranzieht, sondern eine tatsächlich zufällige, d.h. gleichwahrscheinliche, Zuordnung der Merkmalsobjekte zu den alternativen Kategorien unterstellt. In diesem Sinne ist *RE* interpretierbar als der Überschuss der Übereinstimmung über die Nicht-Übereinstimmung und variiert von -1 bis +1, wobei positive Werte den gemeinten Überschuss anzeigen und für die Gültigkeit der Beurteilungen sprechen.

5.2 Voranalysen

Die Voranalysen umfassen die Analyse der Teilnahmebereitschaft, das Vorgehen zum Ausschluss von Probanden sowie die Beschreibung der resultierenden Stichprobe.

5.2.1 Rücklaufquote bzw. Teilnahmebereitschaft

Aufgrund des notwendigen Rekrutierungsvorgehens (Argumentation s. Abschnitt 4.5) kann die Anzahl der Personen, die eine Einladung zur Studie erhalten hat, nicht abgeschätzt werden. Es kann allerdings berichtet werden, dass insgesamt 976 Einträge in der Datenbank verzeichnet wurden. Ein Eintrag kam dann zustande, wenn auf dem ersten Bildschirm der Weiter-Button nach dem Ausfüllen aller erforderlichen Daten angeklickt wurde. Dementsprechend haben 22,4% der Anklickenden die Studie vollständig absolviert ($N = 219$).

Analyse der Abbrüche im Zusammenhang mit Berufserfahrung und Fachwissen. Während sich die Häufigkeit eines Abbruchs bei den *Unerfahrenen* zwischen den Bedingungen *Fachwissen relevant (Essstörung)* und *irrelevant (Brandstiftung)* nicht unterschied (Abbrecher: relevant: 49 % zu irrelevant: 51%), ist bei den Psychotherapeuten, die die Kerngruppe der *Erfahrenen* darstellen, ein häufigerer Abbruch in der Bedingung *Fachwissen irrelevant (Brandstiftung)* zu verzeichnen (65% irrelevant). Bei den Psychotherapeuten, die die Studie beendeten, war die Verteilung hingegen annähernd ausgeglichen (48,1% relevant zu 51,9% irrelevant). Die anderen Berufsgruppen sind zahlenmäßig so gering vertreten, dass ein Vergleich nicht sinnvoll erscheint. Analysen über das Alter und Geschlecht in Abhängigkeit vom Abbruch oder Vollenden der Studie sind nicht möglich, da diese Daten zu einem so späten Untersuchungszeitpunkt erhoben wurden, dass sie von den Abbrechern nicht vorliegen.

Analyse der Abbruchzeitpunkte. 20% der 976 Probanden, die den Studienlink insgesamt angeklickt haben, entschieden sich bereits nach dem Lesen der einleitenden Informationen zum Abbruch, ohne überhaupt eine der Fallgeschichten vorgelegt bekommen zu haben. Bei der Erfassung der *Überzeugtheit der Prüfhypothese* liegt die Ausfallrate bereits bei 53% und unmittelbar danach, d.h. beim Formulieren der Evidenzregeln haben 75% der eingangs anklickenden Teilnehmer die Studie abgebrochen. Von den 243 Probanden, die sich der Aufgabe der Formulierung von Evidenzregeln angenommen haben, führten 219 Personen die Studie dann auch bis zu Ende durch. Die Abbildung 9 veranschaulicht die Verteilung der Abbrüche auf die frühen Studienzeitpunkte grafisch.

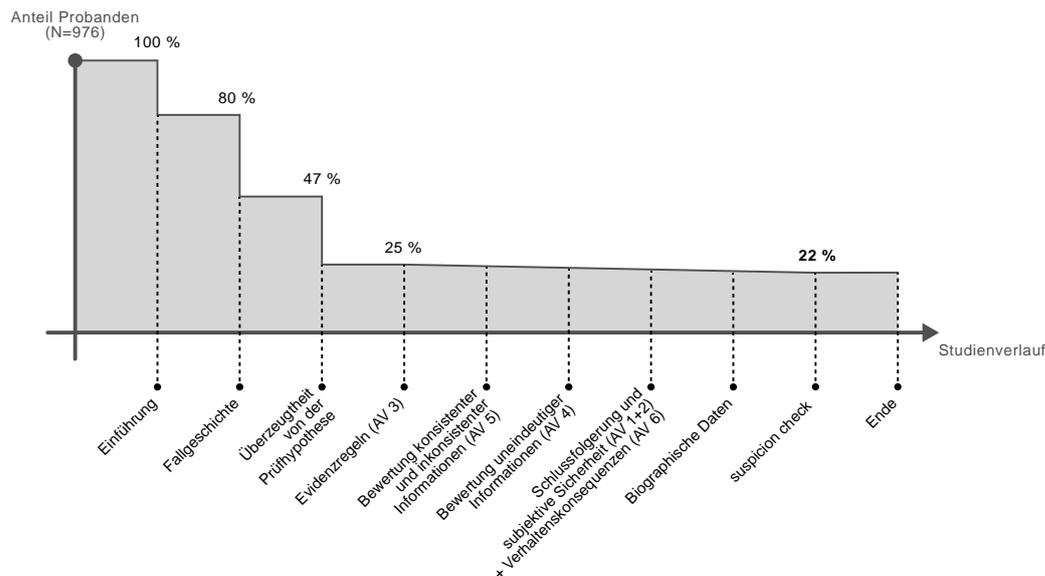


Abbildung 9. Verteilung der Studienabbrüche über den Studienverlauf

5.2.2 Ausschluss von Probanden

Insgesamt liegen von 219 Personen vollständig bearbeitete Datensätze vor. Für diese wurde zunächst geprüft, ob sie eines von drei Ausschlusskriterien erfüllen.

Auffälliger suspicion check. Einerseits wurden 11 Personen (5%) aufgrund ihrer auffälligen Antworten im *suspicion check* (SV *Bewusstsein hinsichtlich konfirmatorischer Prozesse*) von der weiteren Datenanalyse ausgeschlossen. Eine Angabe war dann auffällig, wenn die Freitextantwort erkennen ließ, dass die Probanden sich seitens der Studie durch die Vorinformationen gezielt beeinflusst bzw. manipuliert fühlten, die jeweilige Prüfhypothese zu übernehmen **und** diese Voreinstellung/Erwartung dann „blind“ zu bestätigen. Das Vorgehen war dementsprechend konservativ, weil nicht nur bei der direkten Verbalisierung der Begrifflichkeiten im Zusammenhang mit dem *confirmation bias* ein Ausschluss erfolgte, sondern schon bei jeder Umschreibung. Außerdem wurde eine Antwort als auffällig gewertet, wenn keine instruktionsgemäße Beantwortung vorlag. Z.B. teilte ein Proband Folgendes mit: „Eine originelle Frage, die zu beantworten ich nicht gewillt bin!“

Die Einschätzung als *unauffällig* oder *auffällig* wurde nach den in Anhang D dargestellten Instruktionen von zwei unabhängigen und hypothesenblinden Ratern durchgeführt, die in 95% der Fälle übereinstimmend urteilten. Die Höhe der Übereinstimmung ist aufgrund des $\kappa = 0.60$ ($p = .00$) als *gut* zu bewerten. Ein dritter Rater

bewertete die diskrepanten Items anhand derselben Instruktionen. Ein Proband wurde ausgeschlossen, wenn entweder die ersten beiden Rater übereinstimmend eine auffällige Antwort feststellten oder der dritte Rater eine Auffälligkeit kodierte.

Berufserfahrung. Probanden, die nicht eindeutig den beiden Bedingungen der Variable *Berufserfahrung (erfahren vs. unerfahren)* zugeordnet werden konnten, wurden von der weiteren Datenanalyse ausgeschlossen. Von den 208 nach der Auslese aufgrund des *suspicion checks* verbleibenden Probanden wurden weitere 14 Personen ausgeschlossen, die angaben, sich derzeit in der Psychotherapieausbildung zu finden. Fünf Teilnehmer gaben an, sich keinem der genannten Status zuordnen zu können. Eine genauere Analyse ihrer biographischen Angaben ergab, dass es sich um eine Doktorandin, eine Schulpsychologin, einen Gutachter im Bereich der Fahreignung und einen wirtschaftspsychologischen Coach handelte. Eine weitere Person macht keine Angaben. Da bei keinem von ihnen mit hinreichender Sicherheit ein hinreichendes oder gänzlich fehlendes Fachwissen im Hinblick auf klinisch relevante Störungsbilder unterstellt werden kann, wurden vorsorglich alle ausgeschlossen. Zu guter Letzt waren zwei Personen auszuschließen, die angaben, ohne Therapieausbildung im klinischen Bereich zu arbeiten, da sie konkretisierten, ihre Studienabschlüssen 2004 und 2006 erhalten zu haben. Die Berufserfahrung zum Studienzeitpunkt reicht demnach nicht für eine Zuordnung zu den Berufserfahrenen aus, weswegen die beiden Personen auszuschließen waren.

Zu geringes prüfhypothesenbezogenes Wissen. Wie im Abschnitt 4.1.2 argumentiert, sollte durch den permanenten Zugriff auf Hintergrundinformationen, z.B. zu den diagnostischen Kriterien der Bulimie und Anorexie, ein Mindestmaß an Hintergrundwissen gewährleistet sein. Es muss jedoch in Frage gestellt werden, inwieweit ein Studierender im ersten Fachsemester, der aufgrund des Zeitpunkts der Studiendurchführung möglicherweise erst ein paar Wochen an der Hochschule ist, hiervon aber tatsächlich profitieren kann. Um jegliche Störeinflüsse zu unterbinden, wurden daher auch sämtliche Erstsemester ($n = 21$) von der weiteren Analyse ausgeschlossen.

Aufgrund dieses Vorgehens resultiert eine Stichprobe von 166 Probanden, davon je 40 *Erfahrene* und *Unerfahrene* in der Bedingung *Fachwissen relevant (Essstörung)* sowie 41 *Unerfahrene* bzw. 45 *Erfahrene* in der *Bedingung Fachwissen irrelevant (Brandstiftung)*. Im Hinblick auf die einzusetzenden Auswertungsverfahren, die in der Mehrzahl eine gleichgroße Zellenbesetzung erfordern bzw. sich nur dann robust gegen die Verletzung von Verteilungsvoraussetzungen gezeigt haben, wenn die Zellbesetzung identisch ist, wurden aus der Bedingung *Fachwissen irrelevant* weitere 6 Personen (eine unerfahrene

und 5 erfahrene) zufällig ausgewählt und von der Analyse ausgeschlossen. Die endgültige Analysestichprobe enthält daher mit $N = 160$ Personen je $n = 40$ Probanden in den vier Bedingungskombinationen von *Berufserfahrung* und *Fachwissen*.

5.2.3 Analyse der Beurteilerübereinstimmungen bzgl. der AVn

Im Zusammenhang mit AV 3 (Güte der Evidenzregeln) und AV 6 (Verhaltenskonsequenzen) liegen Freitextantworten vor, die zunächst von unabhängigen und hypothesenblinden Ratern beurteilt wurden. Urteilsobjekte, die unterschiedlich eingeschätzt wurden, sind den Ratern in einem *forced consent* Verfahren erneut vorgelegt worden, so dass letztendlich Einigkeit über die Kodierung aller Evidenzregeln erzielt wurde, bevor die hypothesengeleitete Auswertung vorgenommen wurde.

5.2.3.1 AV 3 Güte der Evidenzregeln

Die $N=160$ Probanden haben insgesamt 960 Evidenzregeln produziert. Da sich jeweils 480 davon auf den Bulimiefall (*Fachwissen relevant*) bzw. den Brandstiftungsfall (*Fachwissen irrelevant*) beziehen, werden die Ermittlungen der Übereinstimmungen sowohl getrennt für die Fälle als auch für die Gesamtstichprobe vorgenommen. Tabelle 17 (siehe folgende Seite) zeigt die Beurteilerübereinstimmungen der drei Merkmale, die für jede der sechs Evidenzregeln einzeln ermittelt wurden.

Die Beurteilerübereinstimmung für die *Instruktionskonformität* ist als *gut* bis *sehr gut* zu bewerten (κ 0.64 bis 0.77). Die Kongruenz für das Merkmal *Teststrategie* liegt hingegen durchgängig im *akzeptablen* Bereich (κ von 0.41 bis 0.47). Ein ähnliches Bild ergibt sich für die Diagnostizität, wobei diese in dem Kontext *Brandstiftung* den *akzeptablen* Bereich verfehlt ($\kappa = 0.34$). Sowohl hinsichtlich der *Teststrategie* als auch bzgl. der *Diagnostizität* ist allerdings ergänzend anzumerken, dass die Randverteilungen jeweils sehr asymmetrisch waren, d.h. dass bspw. die Kategorien *positive vs. negative Teststrategie* mit sehr unterschiedlicher Häufigkeit gewählt wurden. Typischerweise unterschätzt κ in dieser Situation die Übereinstimmung, weshalb auch das alternative Maß *RE* (Maxwell, 1977) betrachtet werden soll.

Für die *Teststrategie* ergibt sich auf Gesamtebene ein $RE = 0.68$ (vs. $\kappa = 0.44$), für die Essstörungsbedingung ein $RE = .74$ (vs. $\kappa = 0.47$) und im Brandstiftungsfall ein $RE = 0.64$ (vs. $\kappa = 0.41$). Der Überschuss der Übereinstimmungen beträgt somit 64% bis 74%, was als Hinweis auf die Gültigkeit der Einschätzungen gewertet werden kann.

Tabelle 17. Beurteilerübereinstimmung für Konformität, Teststrategie & Diagnostizität pro Evidenzregel

<i>Merkmal</i>	<i>Stichprobe</i>	<i>Anzahl beurteilter Evidenzregeln</i>	<i>Prozentuale Über- einstimmung</i>	<i>Cohens κ</i>	<i>p</i>
Instruktionskonformität	Gesamt	960	85	0.72	.00
	Essstörung	480	85	0.77	.00
	Brandstiftung	480	84	0.64	.00
Teststrategie	Gesamt	669	84	0.44	.00
	Essstörung	349	87	0.47	.00
	Brandstiftung	320	82	0.41	.00
Diagnostizität	Gesamt	669	70	0.42	.00
	Essstörung	349	75	0.51	.00
	Brandstiftung	320	65	0.34	.00

Anmerkungen. Die Anzahl der beurteilten Evidenzregeln ist abhängig von den als instruktionskonform gewerteten Regeln, da nur diese hinsichtlich der nachfolgenden Kriterien bewertbar sind.

Die *RE*-Koeffizienten im Hinblick auf die *Diagnostizität* unterscheiden sich hingegen kaum von den *k*-Koeffizienten (Gesamt: $RE = 0.40$ vs. $\kappa = 0.42$; Essstörung: $RE = 0.50$ vs. $\kappa = 0.51$; Brandstiftung: $RE = 0.30$ vs. $\kappa = 0.34$), so dass die allenfalls *akzeptablen* Beurteilerübereinstimmungen nicht auf die asymmetrische Randverteilung attribuiert werden können. Aus Sicht der Verfasserin konnte daher kein Hinweis auf die Gültigkeit der Einschätzungen erbracht werden.

Tabelle 18 zeigt die Beurteilerübereinstimmung derjenigen Merkmale, die über alle sechs Evidenzregeln des Probanden hinweg bestimmt wurden, indem bspw. ausgezählt wurde, wie viele unterschiedliche Hypothesen in den Evidenzregeln berücksichtigt wurden.

Die Beurteilerübereinstimmung für die *Anzahl berücksichtigter Hypothesen* ist als *gut* bis *sehr gut* zu bewerten (κ 0.62 bis 0.84). Die Kongruenz für das Merkmal *Anzahl von Evidenzregeln für die Prüfhypothese* liegt durchgängig im *sehr guten* Bereich (κ von 0.83 bis 0.86). Dasselbe Bild ergibt sich für die *Anzahl von Evidenzregeln für die Alternativhypothese*. Die zugehörigen Koeffizienten liegen ebenfalls ausnahmslos im *sehr guten* Bereich (κ von 0.79 bis 0.95). Die Gültigkeit der Einschätzungen wird daher als gesichert bewertet.

Tabelle 18. Beurteilerübereinstimmung für die Anzahl berücksichtigter Hypothesen bzw. Anzahl Evidenzregeln für Prüf- und Alternativhypothese

<i>Merkmal</i>	<i>Stichprobe</i>	<i>Prozentuale Übereinstimmung</i>	<i>Cohens κ</i>	<i>p</i>
Anzahl berücksichtigter Hypothesen	Gesamt	82	0.72	.00
	Essstörung	73	0.62	.00
	Brandstiftung	91	0.84	.00
Anzahl Evidenzregeln für Prüfungshypothese	Gesamt	88	0.85	.00
	Essstörung	89	0.86	.00
	Brandstiftung	86	0.83	.00
Anzahl Evidenzregeln für Alternativhypothese	Gesamt	94	0.91	.00
	Essstörung	91	0.79	.00
	Brandstiftung	96	0.95	.00

5.2.3.2 AV 6 Verhaltenskonsequenzen

Diese AV wurde ausschließlich bei den $n = 80$ Probanden der Bedingung *Fachwissen relevant (Essstörung)* erhoben, so dass sich die in Tabelle 19 dargestellte Analyse auf diese Teilstichprobe beschränkt.

Da sowohl die Übereinstimmung hinsichtlich der Konsistenz der vorgeschlagenen Themen bei $\kappa = 0.60$ als auch diejenige der Konsistenz der Hausaufgaben mit $\kappa = 0.69$ als gut zu bewerten ist, wird die Gültigkeit der Einschätzungen als gegeben angesehen.

Tabelle 19. Beurteilerübereinstimmung für die Einschätzung der Konsistenz der vorgeschlagenen Themen bzw. Hausaufgaben im Hinblick auf eine typische Behandlung von Bulimikern bzw. Anorektikern

<i>Bezug</i>	<i>Prozentuale Übereinstimmung</i>	<i>Cohens κ</i>	<i>p</i>
Konsistenz vorgeschlagener Themen	75	0.60	.00
Konsistenz Hausaufgaben	80	0.69	.00

5.2.4 Beschreibung der Stichprobe

Die Analytestichprobe umfasst $N = 160$ Probanden, die im Mittel 36.38 Jahre alt sind ($SD = 13.58$; Minimum: 19 Jahre; Maximum 76 Jahre) und zu 75% weiblich sind. Die Probanden stammen aus 14 der 16 deutschen Bundesländer (Mecklenburg-Vorpommern

und Sachsen-Anhalt sind nicht repräsentiert). Eine Person lebte zum Studienzeitpunkt im Ausland. Die Stichprobe wird im Folgenden getrennt für die drei Einflussvariablen beschrieben und hinsichtlich relevanter Störvariablen analysiert.

5.2.4.1 Beschreibung der Stichprobe: Überzeugtheit von der Prüfhypothese

In diesem Zusammenhang interessiert, ob die Induktion der Prüfhypothese geglückt ist. Die Überzeugtheit von der Prüfhypothese liegt in der Gesamtstichprobe bei einem Mittelwert von 4.00 ($SD = 1.27$) und weicht mit $t(159) = 9.93$ ($p = .00$; $d = 0.79$) signifikant vom Skalenmittel (Wert 3) ab. Die Probanden betrachten die Wahrscheinlichkeit des Zutreffens der Prüfhypothese höher als den Skalenmittelwert, der eine mittlere Wahrscheinlichkeit repräsentiert. Aufgrund der Maße für die Variabilität (z.B. $SD = 1.27$, Spannweite 5 bei Minimum = 1 und Maximum = 6) ist jedoch gleichzeitig eine hinreichende Unterschiedlichkeit der Einschätzungen zu erkennen.

5.2.4.2 Beschreibung der Stichprobe: Berufserfahrung

Die Einflussvariable *Berufserfahrung* kann nicht experimentell manipuliert werden. Vielmehr wird die natürliche Variation der Berufserfahrung genutzt, um eine Einteilung in *Erfahrene* vs. *Unerfahrene* vorzunehmen, die sich in definitionsrelevanten Merkmalen daher unterscheiden. Anders als bei den randomisiert zugewiesenen Experimentalbedingungen des *Fachwissens* ist in dieser Konstellation v.a. die Repräsentativität der gewonnenen Stichprobe hinsichtlich der zugrundeliegenden Population von Interesse, da Abweichungen die externe Validität im Sinne der Generalisierbarkeit einschränken würde. Es sollen die beiden Gruppen der *Erfahrenen* und *Unerfahrenen* nun hinsichtlich ihrer charakteristischen Merkmale beschrieben und in Bezug auf ihre Repräsentativität für die zugrundeliegenden Populationen überprüft werden.

5.2.4.2.1 Beschreibung der Erfahrenen

Die $n = 80$ *Erfahrenen* sind im Mittel 47.32 Jahre alt ($SD = 9.77$), 50 davon sind weiblich (63%). Sie arbeiten im Mittel 38.20 Stunden pro Woche ($SD = 14.05$), und davon im Durchschnitt 6.42 Stunden diagnostisch ($SD = 8.65$).

Die $n = 70$ Psychotherapeuten unter den *Erfahrenen* sind im Mittel seit 8.04 Jahren approbiert ($SD = 4.46$). Diese Art der Schätzung der Berufserfahrung ist insofern

konservativ, als dass die im Rahmen der Psychotherapieausbildung gemachten Erfahrungen hierbei nicht berücksichtigt sind. Betrachtet man den Zeitraum nach der Diplomierung als alternative Schätzung für die Berufserfahrung, so sind im Mittel 17.49 Jahre ($SD = 9.41$) seit dem Erhalt des Diploms vergangen. Der eine Psychologe ohne Therapieausbildung gab an, seit 1996, d.h. zum Studienzeitpunkt seit 16 Jahren im klinischen Bereich tätig zu sein und fügt sich damit gut in die Gruppe ein. Eine Abschätzung der Berufserfahrung der $n = 8$ beteiligten Fachärzte anhand der seit der Approbation vergangenen Zeit wurde nicht vorgenommen, da ihre Approbation nicht an die Absolvierung einer Facharztausbildung geknüpft ist. Als Annäherung sei berichtet, dass die Fachärzte ein mittleres Alter von 49.67 Jahren ($SD = 12.88$) aufweisen, was eine vergleichbare Erfahrung zumindest nicht ausschließt.

5.2.4.2.2 Vergleich mit der Population deutscher Psychotherapeuten

Um sich der Frage der Repräsentativität der Stichprobe zu nähern, werden Geschlecht, Alter und Therapieschule betrachtet.

Geschlecht. Der Gesundheitsberichtserstattung des Bundes (Statistisches Bundesamt, 2011) auf Basis der Daten der Bundespsychotherapeutenkammer ist zu entnehmen, dass im Jahr 2008, d.h. zum Zeitpunkt der Studiendurchführung, 33 077 Psychotherapeuten (gemeint sind sowohl Psychologische Psychotherapeuten als auch Kinder- und Jugendlichenpsychotherapeuten) in Deutschland beschäftigt waren. Davon waren 21 735 Personen weiblich, d.h. der Frauenanteil betrug 66%. Dies entspricht weitgehend der Geschlechtsverteilung innerhalb der Stichprobe der Erfahrenen mit 63% Frauen.

Alter. Abbildung 10 veranschaulicht die Altersverteilung der Psychotherapeuten in Deutschland und vergleicht diese mit der in der Untersuchungsstichprobe vorgefundenen. Die gezogene Stichprobe der Erfahrenen ist insgesamt jünger als die Gesamtgruppe der Psychotherapeuten in Deutschland.

Fazit. Die Gruppe der Erfahrenen kann hinsichtlich der Geschlechtsverteilung als repräsentativ für die Gruppe der Psychotherapeuten in Deutschland betrachtet werden. In Bezug auf das Alter weicht die Stichprobe von der Gesamtgruppe jedoch insofern ab, als dass eher jüngere Personen an der (Online-) Studie teilgenommen haben. Letzteres ist bei der Interpretation der Ergebnisse zu berücksichtigen, da eine statistische Kontrolle des Effekts nicht möglich ist.

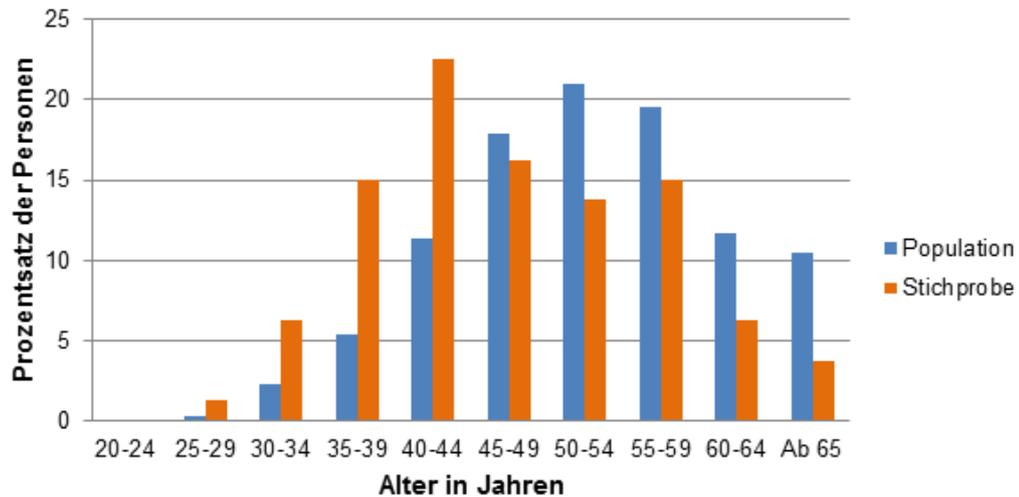


Abbildung 10. Altersverteilung von Studienteilnehmern (Stichprobe) und Psychotherapeuten in Deutschland (Population) zum Studienzeitpunkt

5.2.4.2.3 Beschreibung der Unerfahrenen

Die $n = 80$ Unerfahrenen sind im Mittel 25.44 Jahre alt ($SD = 5.76$), 70 von ihnen sind weiblich (88%). Sie belegen im Mittel das Fachsemester 5.81 ($SD = 3.25$). 78% der Unerfahrenen haben bereits Veranstaltungen der Klinischen Psychologie belegt, 69% in der Diagnostik.

5.2.4.2.4 Vergleich mit der Population Psychologiestudierender in Deutschland

Im Hinblick auf den Vergleich mit der Population Psychologiestudierender in Deutschland sind v.a. das Geschlecht und das Alter interessierende Variablen, da i.d.R. unter den Studierenden, noch keine nennenswerten Differenzierungen vorliegen, die im Zusammenhang mit der Studie relevant sein könnten.

Geschlecht. Im WS 2008/2009 (Zeitpunkt der Studiendurchführung) waren laut Statistischem Bundesamt (2009) 35 153 Studierende für das Fach Psychologie eingeschrieben, davon waren 77% Frauen (resp. 23% Männer). Der Frauenanteil in der Stichprobe der Unerfahrenen ist dementsprechend höher als in der Gesamtgruppe Psychologiestudierender. Es ist kann allerdings festgehalten werden, dass der bereits in der Population vorliegende Frauenüberschuss in der vorliegenden Studie nur etwas erhöht wurde, was weniger stark zu gewichten ist, als wenn eine Umkehrung des Geschlechterverhältnisses aufgetreten wäre.

Alter. Fachspezifische Angaben zum Alter der Studierenden werden im Rahmen der zitierten Statistik nicht gemacht. Ihr ist allerdings zu entnehmen, dass das Durchschnittsalter aller Studierenden in Deutschland (im WS 08/09) 25.4 Jahre betrug, was praktisch identisch zum dem erhobenen Gruppenmittelwert der Unerfahrenen ist (s.o.). Als weitere Annäherung kann auch auf die Arbeit von Schneller und Schneider (2004) zurückgegriffen werden, die in der bundesweiten Absolventenbefragung (Abschluss in Psychologie im Jahre 2003) berichten, dass die Absolventen im Durchschnitt 29.93 Jahre alt waren ($SD = 5.61$). Dass die Stichprobe der *Unerfahrenen* unter dem Alter der Absolventen liegt, spricht in Ermangelung stichhaltigerer Vergleichszahlen für deren Repräsentativität.

Fazit. Die vorliegende Stichprobe der Unerfahrenen weist gegenüber der Population der Psychologiestudierenden einen größeren Frauenanteil auf. Hinweise gegen eine Repräsentativität hinsichtlich des Alters konnten nicht gefunden werden.

5.2.5 Beschreibung der Stichprobe: Fachwissen

Die Variable *Fachwissen* wird experimentell manipuliert und die Zuweisung der Probanden auf die beiden Stufen erfolgt randomisiert, um eine Gleichverteilung von eventuell störenden Variablen zu erreichen. Der Erfolg der Randomisierung ist nun durch einen Vergleich der beiden Experimentalgruppen zu prüfen, wobei die Erwartung getestet wird, dass die Gruppen sich nicht unterscheiden.

Alter und Geschlecht. Die entsprechenden deskriptiven Kennwerte sind in Tabelle 20 veranschaulicht. Die Gleichverteilung durch die randomisierte Zuweisung ist sowohl hinsichtlich des Alters ($t(158) = -0.82$; $p = .41$; $d = 0.13$) als auch des Geschlechts ($\chi^2(.20; 1, N = 160) = 0.53$; $p = .58$; $w^2 = 0.00$) geglückt, da in beiden Fällen die H_0 bei $\alpha = .20$ beibehalten werden konnte und beide Effektgrößen bestenfalls als klein zu beschreiben sind.

Fazit. Es sind keine störenden Einflüsse des Alters und Geschlechts im Zusammenhang mit der Variable *Fachwissen* anzunehmen.

Tabelle 20. Deskriptive Kennwerte des Alters (in Jahren) und des Geschlechts in Abhängigkeit vom Fachwissen

		Alter (in Jahren)		Geschlecht (Anzahl VPn)	
		M	SD	weiblich	männlich
Fachwissen	relevant	35.44	12.84	62	18
	irrelevant	37.20	14.19	58	22

5.3 Hypothesengeleitete Darstellung der Ergebnisse

Die Darstellung der Auswertung erfolgt entsprechend der psychologischen Hypothesenkomplexe, wobei mit den Analysen zum verzerrten Urteil (PH-A_i) begonnen wird (Abschnitt 5.3.1). Die Ergebnisse bzgl. der Variablen zur verzerrten Informationsverarbeitung (PH-B_i sowie PH-C) in Anlehnung an Schulz-Hardt und Köhnken (2000) schließen sich an, bevor die Bestätigungstendenzen I (PH-D) und B (PH-E_i) in den Abschnitten 5.3.4 sowie 5.3.5 thematisiert werden. Den Abschluss bildet die Analyse zu den Verhaltenskonsequenzen (PH-F).

5.3.1 PH-A_i: Verzerrtes Urteil

Der Hypothesenkomplex A_i bezieht sich einerseits auf die AV 1 *Schlussfolgerung* und andererseits auf die AV 2 *Subjektive Sicherheit*.

5.3.1.1 PH-A_i: Schlussfolgerung (AV 1)

Die korrekte Schlussfolgerung wurde insgesamt von 99 Personen (62%) gezogen. 40 Probanden (50%) in der Bedingung *Fachwissen relevant (Essstörung)* stellten die richtige Diagnosen einer Anorexie und 59 Probanden (74%) in der Bedingung *Fachwissen irrelevant (Brandstiftung)* vergaben Freisprüche.

In der Bedingung *Fachwissen relevant (Essstörung)* fielen 33 Personen (41%) das falsche, selbstbestätigende Urteil Bulimie. Zwei Personen (3%) gaben, zwischen den beiden Störungsbildern zu schwanken und weitere zwei (3%) verweigerten eine Diagnosestellung. Die verbleibenden drei Probanden (4% - aufgrund von Rundungen ergeben sich 101%) diagnostizierten eine andere Störung (z.B. Depression).

In der Bedingung *Fachwissen irrelevant (Brandstiftung)* zogen 14 Probanden (18%) die falsche, selbstbestätigende Schlussfolgerung, indem sie den Angeklagten für schuldig befanden. Die restlichen sieben Personen (8%) gaben an, keine Entscheidung fällen zu können. Die folgende Analyse bezieht sich auf die jeweils $n = 73$ Probanden pro Bedingung der Variablen *Fachwissen*, die entweder die korrekte Schlussfolgerung oder die selbstbestätigende zogen.

Überprüfung des Zusammenhangs von Prädiktor und Kriterium. In der Gesamtstichprobe liegt mit $r = 0.28$ ($p = .00$) ein als *klein bis mittelgroß* zu bezeichnender positiver Zusammenhang zwischen der *Überzeugtheit von der Prüfhypothese* und der *Schlussfolgerung* vor. Abbildung 11 kann eine Veranschaulichung der Verteilung der *Überzeugtheit von der Prüfhypothese* über die Kategorien der *Schlussfolgerung* in Form eines Boxplots entnommen werden.

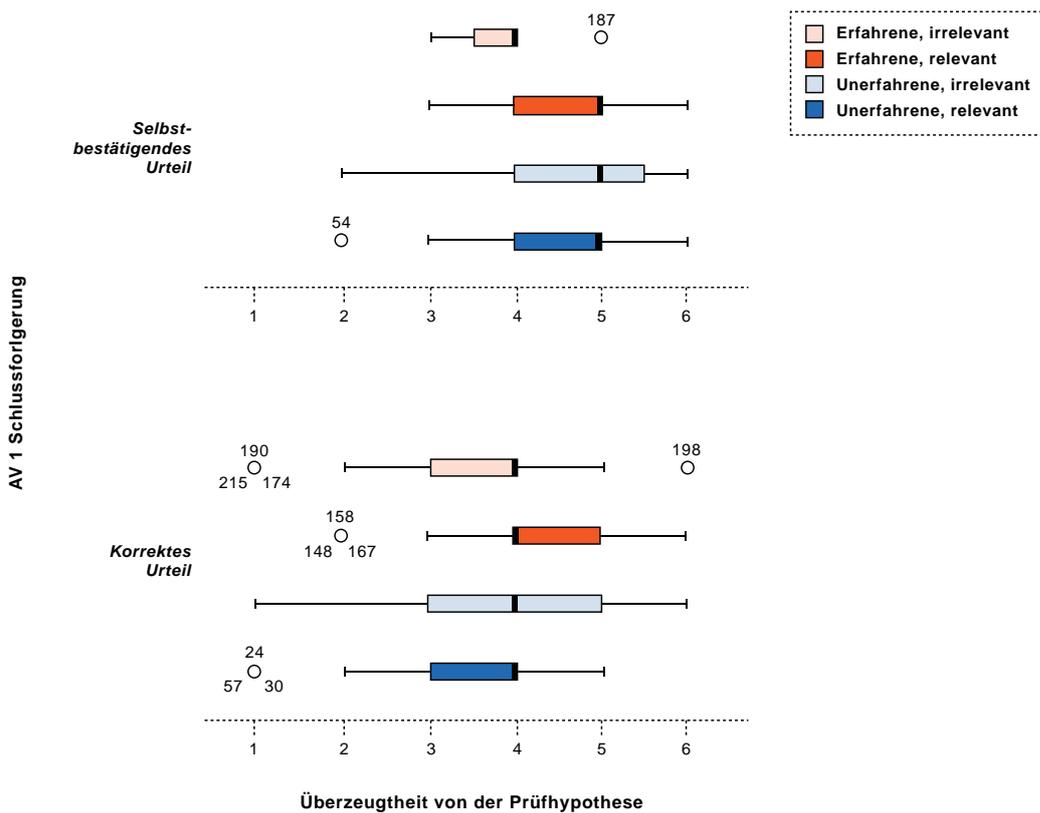


Abbildung 11. Boxplot zur Veranschaulichung des Zusammenhangs der Überzeugtheit von der Prüfhypothese und der Schlussfolgerung in den Bedingungskombinationen von Fachwissen und Berufserfahrung

Probanden, die ein *korrektes Urteil* gefällt haben, wiesen zuvor den Median 4 hinsichtlich der *Überzeugtheit der Hypothese* (ablesbar als senkrechter Strich in der Box) auf, während die Probanden mit *selbstbestätigendem Urteil* mehrheitlich den Median 5 aufwiesen. Lediglich die *Erfahrenen* der Bedingung mit *irrelevantem Fachwissen* wie auch den Median von 4 auf. Diese deskriptive Beschreibung steht im Einklang mit der signifikanten positiven Korrelation.

Außerdem sind der Abbildung Hinweise auf die unterschiedliche Variabilität zu entnehmen, denn während bei Personen mit *selbstbestätigendem Urteil* linkssteile Verteilungen dominieren (angezeigt durch das Zusammenfallen von Median und oberem Ende der Box, die dem oberen Quartil entspricht), gibt es bei Personen mit korrektem Urteil ein heterogeneres Bild. Außerdem weisen die Personen mit korrektem Urteil untereinander eine größere Variabilität auf, da die Antennen, d.h. die sich an die Box anschließenden Querbalken, tendenziell größere Wertebereiche abdecken und mehr Ausreißer zu beobachten, die von den Kreisen repräsentiert werden.

Logistische Regression. In der logistischen Regressionsanalyse erwies sich das Modell mit den drei Prädiktoren *Überzeugtheit von der Prüfhypothese*, *Berufserfahrung* und *Fachwissen* als geeignet zur Vorhersage der *Wahrscheinlichkeit einer selbstbestätigenden Diagnose* ($-2\ln L = 161.83$; $\chi^2(.05; 3, N = 146) = 21,63$; $p = .00$). Der Anteil der Varianzaufklärung durch die Prädiktoren ist allerdings als maximal akzeptable einzuschätzen, da das Nagelkerke- $R^2 = 0.19$ nach den Empfehlungen von Backhaus et al. (2008) den *akzeptablen* Bereich (Werte ab 0.20) knapp unterschreitet. Im Einklang mit den bisherigen Ausführungen ist auch das Klassifikationsergebnis zu betrachten, denn mit einem Gesamtprozentsatz von 72.6 richtigen Klassifikationen liegt die Rate einerseits über der aufgrund des Zufalls zu erwartenden Rate von 50%, ist jedoch bei einer Rate von gut einem Viertel Falschklassifikationen keineswegs als gut zu bezeichnen.

Tabelle 21 (siehe folgende Seite) präsentiert die Größen zur Einschätzung des Einflusses der *drei Prädiktoren*. Für die *Überzeugtheit von der Prüfhypothese* wurde ein signifikanter Regressionskoeffizient von 0.513 ($p = .00$) ermittelt, womit die Variable aufgrund des positiven Vorzeichens des Koeffizienten die *Wahrscheinlichkeit eines selbstbestätigenden Urteils* in erwarteter Weise ($PH-A_{1a}$) bei Erhöhung der Variable *Überzeugtheit von der Prüfhypothese* erhöht. Interpretatorisch ist dem Effektkoeffizienten e^b zu entnehmen, dass die Erhöhung der *Überzeugtheit von der Prüfhypothese* um eine Einheit das Chancenverhältnis zu Gunsten des selbstbestätigenden Urteils um den Faktor 1.670 erhöht.

Tabelle 21. Zusammenfassung der logistischen Regressionsanalyse zur Vorhersage der Wahrscheinlichkeit selbstbestätigenden Urteils in Abhängigkeit von der Überzeugtheit der Prüfhypothese, der Berufserfahrung und dem Fachwissen ($n = 146$)

Variable	Regressions- koeffizient B	Standard- fehler	Wald	df	p	e^b	95%-Konfidenz- intervall e^b	
							Untere Grenze	Obere Grenze
Überzeugtheit von der Prüfhypothese	0.513	0.17	8.72	1	.00	1.670	1.188	2.347
Berufserfahrung	0.068	0.38	0.03	1	.86	1.071	0.505	2.268
Fachwissen	-1.164	0.39	8.83	1	.00	3.203	1.486	6.903

Anmerkungen. Der Effektkoeffizient e^b ist äquivalent zu den *odds ratio* (OR) zu betrachten.

Für die *Berufserfahrung* wurde ein Regressionskoeffizient von 0.068 ($p = .00$) gefunden, der die Wahrscheinlichkeit eines selbstbestätigenden Urteils von *Unerfahrenen* gegenüber *Erfahrenen* nicht signifikant beeinflusst, was im Zusammenhang mit der PH-A_{1b} erwartet worden war. Dem Effektkoeffizienten e^b ist zu entnehmen, dass das Chancenverhältnis des selbstbestätigenden zum korrekten Urteil zu Gunsten des selbstbestätigenden Urteils aufgrund des Faktors 1.071 kaum ändert, weswegen der Variable kein Einfluss zugesprochen werden kann.

Für das *Fachwissen* wurde ein Regressionskoeffizient von -1.164 ($p = .00$) gefunden. Das negative Vorzeichen zeigt an, dass die Wahrscheinlichkeit für ein selbstbestätigendes Urteil in der Bedingung *Fachwissen irrelevant (Brandstiftung)* gegenüber der Referenzkategorie *Fachwissen relevant (Essstörung)* signifikant abnimmt. Interpretatorisch ist dem Effektkoeffizienten e^b zu entnehmen, dass das Chancenverhältnis zu Gunsten des selbstbestätigenden Urteils um den Faktor 3.203 reduziert ist, wenn das Fachwissen irrelevant ist. Dies widerspricht der vorab formulierten Annahme (PH-A_{1c}), die eine Zunahme der Wahrscheinlichkeit bei irrelevantem Fachwissen unterstellt hat.

5.3.1.2 PH-A₂: Subjektive Sicherheit (AV 2)

Die subjektive Sicherheit bzgl. der letztendlich gestellten Diagnose repräsentiert den Grad des Vertrauens in die Entscheidung bzw. die damit verbundene Bewährung der Hypothese. Bei den Analysen werden wiederum nur diejenigen Probanden berücksichtigt, die entweder die korrekte oder die selbstbestätigende Schlussfolgerung gezogen haben. Die mittlere Sicherheit über alle Personen mit eindeutigen Urteil beträgt 3.90 ($SD = 1.38$). Tabelle 22 gibt die bedingungsspezifischen Mittelwerte wieder.

Tabelle 22. Mittelwerte und Streuungen der subjektiven Sicherheit in den Bedingungskombinationen von Berufserfahrung und Fachwissen

		Berufserfahrung	
		erfahren	unerfahren
		M (SD)	M (SD)
Fachwissen	relevant	4.29 (1.36)	3.87 (1.21)
	irrelevant	3.46 (1.52)	4.00 (1.35)

Anmerkungen. Skala von 0 = sehr unsicher bis 6 = sehr sicher

Überprüfung des Zusammenhangs von Prädiktor und Kriterium. Die Korrelation der *Überzeugtheit von der Prüfhypothese* mit der *subjektiven Sicherheit* erweist sich mit $r = -0.10$ ($p = .13$) entgegen der Erwartung als nicht signifikant (PH-A_{2a}). Es konnte demnach kein Hinweis dafür gefunden werden, dass eine höhere *Überzeugtheit von der Prüfhypothese* mit einer höheren *subjektiven Urteilssicherheit* einhergeht. Deswegen ist die Voraussetzung zum sinnvollen Einschluss der Variablen *Überzeugtheit von der Prüfhypothese* in die Kovarianzanalyse nicht gegeben. Der Einfluss von *Berufserfahrung* und *Fachwissen* wird daher im Rahmen einer zweifaktoriellen Varianzanalyse ermittelt.

Zweifaktorielle Varianzanalyse. Tabelle 23 zeigt die Ergebnisse der zweifaktorielle Varianzanalyse. Dass hierbei kein Haupteffekt für die *Berufserfahrung* beobachtet werden konnte (PH-A_{2b}; $F < 1$) stimmt mit der Erwartung über ein, keine Unterschiede zwischen den *Erfahrenen* und den *Unerfahrenen* zu finden. Erwartungskonträr ist jedoch die Tatsache, dass auch kein Haupteffekt für das *Fachwissen* festzustellen ist (PH-A_{2c}; $F(1,156) = 2.42$; $p = .12$; $f^2 = .00$). Hinweise auf eine Interaktion der beiden Variablen konnten im Rahmen der explorativen Analyse nicht gefunden werden ($F(1,156) = 3.19$; $p = .08$; $f^2 = .02$).

Tabelle 23. Tafel der zweifaktoriellen Varianzanalyse zur Überprüfung der Abhängigkeit der subjektiven Sicherheit von der Berufserfahrung und dem Fachwissen

Quelle der Variation	SS	df	MS	F	p	$\frac{A-H_0^a}{A-H_1}$	f^2
Berufserfahrung	0.06	1	0.06	0.03	.86	$A-H_0$.00
Fachwissen	4.56	1	4.56	2.42	.12	$A-H_0$.01
Berufserfahrung*Fachwissen	6.01	1	6.01	3.19	.08	$A-H_0$.02
Fehler	293.33	156	1.88				
Total	0.06	1	0.06	0.03	.86		

Anmerkungen. Für die Prüfung der Wirkung der *Berufserfahrung* gilt $\alpha = .20$

^a $A-H_0/A-H_1$: Beibehalten der Nullhypothese ($A-H_0$) bzw. Annahme der Alternativhypothese ($A-H_1$)

5.3.2 PH-B_i: Asymmetrische Evidenzregeln

Der Hypothesenkomplex PH-B_i bezieht sich auf die Güte der Evidenzregeln, die mithilfe von drei Maßen im Rahmen der AV 3 *Evidenzregeln* eingeschätzt wird.

5.3.2.1 PH-B₁: Überwiegen der positiven Teststrategie (AV 3a)

Über alle Probanden hinweg liegt der Mittelwert der Evidenzregeln, die eine positive Teststrategie beinhalten, bei 3.67 ($SD = 2.25$), derjenigen der negativ formulierten Evidenzregeln bei 0.21 ($SD = 0.30$). Da eine Einschätzung allerdings nur innerhalb der instruktionskonformen Evidenzregeln sinnvoll ist, erfolgt eine Relativierung an der individuellen Anzahl dieser. Konkret wurde die Anzahl positiver und negativer Evidenzregeln jeweils an der Anzahl instruktionskonformer Regeln relativiert und anschließend die Differenz gebildet. Die resultierende Variable nimmt Werte zwischen -1 und +1 an, negative Werte zeigen ein Überwiegen negativer Evidenzregeln an, Werte um Null ein ausgeglichenes Verhältnis und bei positiven Werten überwiegt die positive Teststrategie. Der Mittelwert 0.84 ($SD = 0.30$) deutet somit eine starke Präferenz für die positive Teststrategie in der Gesamtstichprobe an. Die bedingungsspezifischen deskriptiven Kennwerte können Tabelle 24 entnommen werden.

Tabelle 24. Mittelwerte und Streuungen der Variable „Überwiegen der Evidenzregeln im Sinne einer positiven Teststrategie“ in den Bedingungskombinationen von Berufserfahrung und Fachwissen

		Berufserfahrung	
		<i>erfahren</i>	<i>unerfahren</i>
		<i>M (SD)</i>	<i>M (SD)</i>
<i>Fachwissen</i>	<i>relevant</i>	0.92 (0.17)	0.84 (0.28)
	<i>irrelevant</i>	0.85 (0.29)	0.76 (0.38)

Anmerkungen. Mögliche Werte von -1 bis +1, wobei positive Werte ein Überwiegen der positiven Evidenzregeln anzeigen. Je höher der Betrag, desto ausgeprägter die jeweilige Präferenz.

Überprüfung des Zusammenhangs von Prädiktor und Kriterium. In der Gesamtstichprobe liegt ein *kleiner* Zusammenhang von der *Überzeugtheit der Prüfhypothese* und dem *Überwiegen von Evidenzregeln mit positiver Teststrategie* vor ($r = 0.27$; $p = .00$). Deswegen erfolgt eine kovarianzanalytische Auswertung aller drei bedingenden Variablen.

Kovarianzanalyse. Die Ergebnisse der Kovarianzanalyse sind in Tabelle 25 wiedergegeben. In Übereinstimmung mit der positiven Korrelation zwischen der *Überzeugtheit von der Prüfhypothese* und dem *Überwiegen positiver Evidenzregeln* erweist sich die Kovariate mit einem *mittleren* Effekt als bedeutsam (PH-B_{1a}: $F(1,130) = 8.23$; $p = .01$; $f^2 = 0.06$). Die Nullhypothese der Gleichheit von *Erfahrenen* und *Unerfahrenen* hinsichtlich des Überwiegens positiver Evidenzregeln muss hingegen entgegen der Erwartung bei $\alpha = .20$ und einen kleinen Effekt verworfen werden (PH-B_{1b}: $F(1,130) = 1.92$; $p = .17$; $f^2 = .02$). Der erwartete Unterschied für das *Fachwissen* blieb allerdings aus (PH-B_{1c}: $F(1,130) = 1.60$; $p = .21$; $f^2 = .01$). Desweiteren wurde keine Hinweise auf eine Interaktion der Einflussvariablen gefunden ($F < 1$).

Tabelle 25. Tafel der Kovarianzanalyse zur statistischen Prüfung der PH-B₁ zum Überwiegen positiver Evidenzregeln

Quelle der Variation	SS	df	MS	F	p	$\frac{A-H_0^a}{A-H_1}$	f^2
Überzeugtheit	0.67	1	0.67	8.23	.01	A-H ₁	.06
Berufserfahrung	0.16	1	0.16	1.92	.17	A-H ₁	.02
Fachwissen	0.13	1	0.13	1.59	.21	A-H ₀	.01
Berufserfahrung*Fachwissen	0.01	1	0.01	0.06	.80	A-H ₀	.00
Fehler	10.59	130	0.08				
Total	108.09	135					

Anmerkungen. Für die Prüfung der Wirkung der *Berufserfahrung* gilt $\alpha = .20$

^a A-H₀/A-H₁: Beibehalten der Nullhypothese (A-H₀) bzw. Annahme der Alternativhypothese (A-H₁)

5.3.2.2 PH-B₂: Überwiegen nicht diagnostischer Merkmale (AV 3b)

Geplant war an dieser Stelle die Auswertung der Variablen *Überwiegen nicht diagnostischer Merkmale* in den Evidenzregeln, die sich aus der Differenz der Evidenzregeln mit nicht diagnostischen und diagnostischen Evidenzregeln ergibt, wenn die entsprechenden Anzahlen zuvor an der Zahl instruktionskonformer Evidenzregeln relativiert wurden.

Hierfür ist jedoch vorauszusetzen, dass die Qualität eines Merkmals als *diagnostisch* oder *nicht diagnostisch* zuverlässig bestimmbar ist. Aufgrund der allenfalls akzeptablen

Beurteilerübereinstimmungen hinsichtlich der *Diagnostizität*, die sowohl seitens κ als auch *RE* angezeigt wurden (s. Abschnitt 5.2.3.1), muss geschlussfolgert werden, dass nicht mit hinreichender Sicherheit festgestellt werden kann, welche Qualität eine Evidenzregel im Hinblick auf die *Diagnostizität* aufweist. Diese Feststellung verbietet die geplante Hypothesenprüfung, um Aussagen zu vermeiden, die auf unreliaiblen und invaliden Einschätzungen beruhen und damit zufällig sein könnten.

5.3.2.3 *PH-B₃: Positive Evidenzregeln mit nicht diagnostischen Merkmalen*

Da die Variable *positive Evidenzregeln mit nicht diagnostischen Merkmalen* ebenfalls auf Grundlage der Einschätzungen der *Diagnostizität* der verwendeten Merkmale gebildet wird, muss analog zur Argumentation bzgl. der *PH-B₂* auf eine Auswertung verzichtet werden.

5.3.2.4 *PH-B₄ und PH-B₅: Berücksichtigung alternativer Hypothesen (AV 3c)*

Bei der Berücksichtigung alternativer Hypothesen werden zwei Maße unterschieden, deren Auswertung getrennt erfolgen soll: die Anzahl insgesamt berücksichtigter Hypothesen sowie das Überwiegen der Evidenzregeln zur Prüfhypothese in Relation zur Alternativhypothese.

5.3.2.4.1 *PH-B₄: Anzahl berücksichtigter Hypothesen*

Die Berücksichtigung alternativen Hypothesen innerhalb der sechs zu generierenden Evidenzregeln wird aus der Anzahl unterschiedlicher Hypothesen erschlossen. Theoretisch sind Werte von null bis sechs denkbar, wobei null die Situation repräsentiert, dass keine einzige Hypothese genannt, was genau dann der Fall wäre, wenn keine einzige der sechs Evidenzregeln instruktionskonform ist. Der Wert sechs repräsentiert die Situation, in der jede Evidenzregel eine andere Hypothese anspricht. Dies ist jedoch sowohl für die Fallbeschreibung im Brandstiftungskontext als auch im Essstörungskontext unwahrscheinlich. Plausibel ist jedoch die Nutzung von zwei unterschiedlichen Hypothesen: im Brandstiftungsfall Schuld vs. Unschuld sowie Bulimie und Anorexie im Essstörungskontext.

Tabelle 26. Mittelwerte und Streuungen der Anzahl berücksichtigter Hypothesen in den Bedingungskombinationen der Variablen Berufserfahrung und Fachwissen

		Berufserfahrung	
		erfahren	unerfahren
		M (SD)	M (SD)
Fachwissen	relevant	1.62 (1.19)	1.38 (0.81)
	irrelevant	1.55 (0.82)	1.60 (0.67)

Anmerkungen. Mögliche Werte theoretisch von 0 bis 6, aus Plausibilitätsgründen jedoch von 0 bis 2

Überprüfung des Zusammenhangs von Prädiktor und Kriterium. Der unterstellte Zusammenhang von *Überzeugtheit von der Prüfhypothese* und der *Anzahl berücksichtigter Hypothesen* lässt sich als *kleiner*, signifikanter Effekt bestätigen ($r = -0.13$; $p = .05$), indem eine höhere *Überzeugtheit von der Prüfhypothese* mit einer geringeren *Anzahl berücksichtigter Hypothesen* einhergeht. Deswegen erfolgt eine kovarianzanalytische Auswertung aller drei bedingenden Variablen.

Kovarianzanalyse. Die Ergebnisse der Kovarianzanalyse sind in Tabelle 27 wiedergegeben. Trotz des in der Korrelationsanalyse gefundenen *kleinen* Effekts für den negativen Zusammenhang der *Überzeugtheit von der Prüfhypothese* und der *Anzahl berücksichtigter Hypothesen* erweist sich die Kovariate als unbedeutsam (PH-B_{4a}: $F(1,155) = 2.95$; $p = .09$; $f^2 = .02$), da die Nullhypothese trotz eines *kleinen* Effekts beibehalten werden muss. In Abhängigkeit der Variablen *Berufserfahrung* ergibt sich wiederum erwartungsgemäß kein Effekt (PH-B_{4b}: $F < 1$). Anders als erwartet liegt zudem kein Haupteffekt für das *Fachwissen* (PH-B₄: $F < 1$) vor. Ein Interaktionseffekt zeigte sich ebenfalls nicht ($F(1,155) = 1.53$; $p = .22$; $f^2 = .01$).

Tabelle 27. Tafel der Kovarianzanalyse zur statistischen Prüfung der PH-B₄ zur Anzahl berücksichtigter Hypothesen

Quelle der Variation	SS	df	MS	F	p	$\frac{A-H_0^a}{A-H_1}$	f ²
Überzeugtheit	2.32	1	2.32	2.95	.09	A-H ₀	.02
Berufserfahrung	0.52	1	0.52	0.66	.42	A-H ₀	.00
Fachwissen	0.07	1	0.07	0.09	.77	A-H ₀	.00
Berufserfahrung*Fachwissen	1.20	1	1.20	1.53	.22	A-H ₀	.01
Fehler	121.93	155	0.79				
Total	504.00	160					

Anmerkungen. Für die Prüfung der Wirkung der *Berufserfahrung* gilt $\alpha = .20$

^a A-H₀/A-H₁: Beibehalten der Nullhypothese (A-H₀) bzw. Annahme der Alternativhypothese (A-H₁)

5.3.2.4.2 PH-B₅: Überwiegen der Evidenzregeln zur Prüfhypothese

Das *Überwiegen der Evidenzregeln zur Prüfhypothese* wird durch die Differenz der Evidenzregeln zur Prüfhypothese und zur Alternativhypothese bestimmt, wobei theoretisch Werte von -6 bis +6 denkbar sind. Negative Werte zeigen an, dass mehr Evidenzregeln zur Alternativhypothese aufgestellt wurden. Bei einer Differenz von null ist die Anzahl ausgeglichen und positive Werte deuten auf das Überwiegen der Evidenzregeln zur Prüfhypothese hin.

In der Gesamtstichprobe ergibt sich für das *Überwiegen* ein Mittelwert von 1.35 ($SD = 2.33$), was auf eine Übervorteilung der Prüfhypothese hindeutet. Die bedingungsspezifischen deskriptiven Kennwerte können Tabelle 28 entnommen werden.

Tabelle 28. Mittelwerte und Streuungen der Variablen *Überwiegen der Evidenzregeln zur Prüfhypothese* in den Bedingungskombinationen der Variablen *Berufserfahrung* und *Fachwissen*

		<i>Berufserfahrung</i>	
		<i>erfahren</i>	<i>unerfahren</i>
		<i>M (SD)</i>	<i>M (SD)</i>
<i>Fachwissen</i>	<i>relevant</i>	1.88 (2.45)	1.58 (2.68)
	<i>irrelevant</i>	0.98 (1.73)	0.98 (2.29)

Anmerkungen. Mögliche Werte von -6 bis +6, positive Werte zeigen das Überwiegen der Evidenzregeln zur Prüfhypothese an. Je höher der Betrag, desto ausgeprägter die Präferenz.

Überprüfung des Zusammenhangs von Prädiktor und Kriterium. Der unterstellte Zusammenhang von *Überzeugtheit von der Prüfhypothese* und dem *Überwiegen der Evidenzregeln zur Prüfhypothese* lässt sich als *kleiner*, signifikanter Effekt bestätigen ($r = 0.20$; $p = .01$), indem eine höhere Überzeugtheit mit einem stärkeren Überwiegen von prüfhypothesenspezifischen Evidenzregeln einhergeht. Deswegen erfolgt eine kovarianzanalytische Auswertung aller drei bedingenden Variablen.

Kovarianzanalyse. Die Ergebnisse der Kovarianzanalyse sind in Tabelle 29 wiedergegeben. In Übereinstimmung mit der *kleinen* positiven Korrelation der *Überzeugtheit von der Prüfhypothese* mit dem *Überwiegen der Evidenzregeln zur Prüfhypothese* erweist sich die Kovariate mit als bedeutsam (PH-B_{5a}: $F(1,155) = 5.04$; $p = .03$; $f^2 = .03$), da die Nullhypothese bei Vorhandensein eines *kleinen* Effekts verworfen werden muss. Haupteffekte liegen allerdings ebenso wenig hinsichtlich der *Berufserfahrung* vor (PH-B_{5b}: $F < 1$) wie in Bezug auf das *Fachwissen* (PH-B_{5c}: $F(1,155) = 3.04$; $p = .08$; $f^2 = .02$) vor, was in Bezug auf die Berufserfahrung erwartet auch erwartet

wurde. Im Zusammenhang mit dem *Fachwissen* war jedoch der auch deskriptiv sichtbare Vorteil für die Personen mit relevantem Fachwissen angenommen worden. Trotz eines kleinen Effekts muss die Nullhypothese jedoch im Sinne eines konservativen Vorgehens beibehalten werden. Desweiteren ist festzuhalten, dass keine Interaktion beobachtet wurde ($F < 1$).

Tabelle 29. *Tafel der Kovarianzanalyse zur statistischen Prüfung der PH-B5 zum Überwiegen von Evidenzregeln zur Prüfhypothese*

<i>Quelle der Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	$\frac{A-H_0^a}{A-H_1}$	<i>f</i> ²
Überzeugtheit	26.34	1	26.34	5.04	.03	$A-H_1$.03
Berufserfahrung	0.41	1	0.41	0.08	.78	$A-H_0$.00
Fachwissen	15.89	1	15.89	3.04	.08	$A-H_0$.02
Berufserfahrung*Fachwissen	0.19	1	0.19	0.04	.85	$A-H_0$.00
Fehler	809.76	155	5.22				
Total	1152.00	160					

Anmerkungen. Für die Prüfung der Wirkung der *Berufserfahrung* gilt $\alpha = .20$

^a $A-H_0/A-H_1$: Beibehalten der Nullhypothese ($A-H_0$) bzw. Annahme der Alternativhypothese ($A-H_1$)

5.3.3 PH-C: Hypothesenkonsistente Interpretation uneindeutiger Evidenz

Hypothesenkomplex PH-C bezieht sich auf die hypothesenkonsistente Interpretation uneindeutiger Evidenz sensu Schulz-Hardt und Köhnken (2000). Im Folgenden wird analysiert, wie die *Anzahl der Aufwertungen* von den drei bedingenden Variablen abhängt. Pro Person sind null bis acht Aufwertungen möglich, da insgesamt acht uneindeutige Items vorgelegt wurden.

In der Gesamtstichprobe ergibt sich für die *Anzahl der Aufwertungen* ein Mittelwert von 2.33 ($SD = 1.73$). Die bedingungsspezifischen deskriptiven Kennwerte können Tabelle 30 (siehe folgende Seite) entnommen werden.

Überprüfung des Zusammenhangs von Prädiktor und Kriterium. Der unterstellte Zusammenhang von *Überzeugtheit* von der *Prüfhypothese* und der *Anzahl von Aufwertungen* lässt sich erwartungskonträr nicht bestätigen (PH-C_a), da trotz eines kleinen Effekts bei einer Korrelation von $r = 0.12$ ($p = .07$) die Nullhypothese beibehalten

werden muss. Deswegen ist die Voraussetzung zum sinnvollen Einschluss der Variablen *Überzeugtheit von der Prüfhypothese* in die Kovarianzanalyse nicht gegeben. Der Einfluss von *Berufserfahrung* und *Fachwissen* wird daher im Rahmen einer zweifaktoriellen Varianzanalyse ermittelt.

Tabelle 30. Mittelwerte und Streuungen der Variablen Anzahl der Aufwertungen in den Bedingungskombinationen der Variablen *Berufserfahrung* und *Fachwissen*

		<i>Berufserfahrung</i>	
		<i>erfahren</i>	<i>unerfahren</i>
		<i>M (SD)</i>	<i>M (SD)</i>
<i>Fachwissen</i>	<i>relevant</i>	1.80 (1.60)	2.13 (1.90)
	<i>irrelevant</i>	2.78 (1.67)	2.63 (1.63)

Anmerkungen. Mögliche Werte von 0 bis 8

Zweifaktorielle Varianzanalyse. Tabelle 31 zeigt die Ergebnisse der zweifaktoriellen Varianzanalyse. Dass hierbei kein Haupteffekt für die *Berufserfahrung* beobachtet werden konnte (PH-C_b: $F < 1$) stimmt mit der Erwartung über ein, keine Unterschiede zwischen den *Erfahrenen* und den *Unerfahrenen* zu finden. Erwartungskonform ist ebenfalls der signifikante Haupteffekt für das *Fachwissen* (PH-C_c: $F(1,156) = 7.49$; $p = .01$; $f^2 = .05$), der einem *kleinen bis mittleren* Effekt entspricht und sich darin ausdrückt, dass Personen mit *relevantem Fachwissen* weniger Aufwertungen vornehmen als Personen mit *irrelevantem Fachwissen*. Hinweise auf eine Interaktion der Variablen *Berufserfahrung* und *Fachwissen* konnten im Rahmen der explorativen Analyse nicht gefunden werden ($F < 1$).

Tabelle 31. Tafel der zweifaktoriellen Varianzanalyse zur statistischen Prüfung der PH-C zur Anzahl von Aufwertungen

<i>Quelle der Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	$\frac{A-H_0^a}{A-H_1}$	<i>f</i> ²
<i>Berufserfahrung</i>	0.31	1	0.31	0.11	.75	$A-H_0$.00
<i>Fachwissen</i>	21.76	1	21.76	7.49	.01	$A-H_1$.05
<i>Berufserfahrung*Fachwissen</i>	2.26	1	2.26	0.78	.38	$A-H_0$.00
<i>Fehler</i>	453.13	156	2.91				
<i>Total</i>	1347.00	160					

Anmerkungen. Für die Prüfung der Wirkung der *Berufserfahrung* gilt $\alpha = .20$

^a $A-H_0/A-H_1$: Beibehalten der Nullhypothese ($A-H_0$) bzw. Annahme der Alternativhypothese ($A-H_1$)

5.3.4 PH-D: Bestätigungstendenz I (AV 5b)

Der Hypothesenkomplex PH-D bezieht sich auf die Abhängigkeit der Erkennensleistung von Widersprüchen sensu Hager und Weißmann (1991) von den drei bedingenden Variablen. Im Folgenden wird die Analyse der *Anzahl der nicht erkannten Widersprüche* präsentiert. Da insgesamt vier Informationen vorgelegt wurden, die mehrheitlich als Widerspruch bewertet worden waren, liegen die möglichen Werte der Variablen zwischen null und vier.

In der Gesamtstichprobe ergibt sich für die *Anzahl nicht erkannter Widersprüche* ein Mittelwert von 2.20 ($SD = 1.31$). Die bedingungsspezifischen deskriptiven Kennwerte können Tabelle 32 entnommen werden.

Tabelle 32. Mittelwerte und Streuungen der Variablen *Anzahl der nicht erkannten Widersprüche* in den Bedingungskombinationen der Variablen *Berufserfahrung* und *Fachwissen*

		<i>Berufserfahrung</i>	
		<i>erfahren</i>	<i>unerfahren</i>
		<i>M (SD)</i>	<i>M (SD)</i>
<i>Fachwissen</i>	<i>relevant</i>	2.93 (1.23)	2.23 (1.21)
	<i>irrelevant</i>	2.05 (1.20)	1.60 (1.28)

Anmerkungen. Mögliche Werte von 0 bis 4

Überprüfung des Zusammenhangs von Prädiktor und Kriterium. Der unterstellte Zusammenhang von *Überzeugtheit von der Prüfhypothese* und der *Anzahl nicht erkannter Widersprüche* lässt sich erwartungskonträr nicht bestätigen (PH-D_a), da trotz eines *kleinen* Effekts bei einer Korrelation von $r = 0.11$ ($p = .09$) die Nullhypothese beibehalten werden muss. Deswegen ist die Voraussetzung zum sinnvollen Einschluss der Variablen *Überzeugtheit von der Prüfhypothese* in die Kovarianzanalyse nicht gegeben. Der Einfluss von *Berufserfahrung* und *Fachwissen* wird daher im Rahmen einer zweifaktoriellen Varianzanalyse ermittelt.

Zweifaktorielle Varianzanalyse. Tabelle 33 zeigt die Ergebnisse der zweifaktoriellen Varianzanalyse. Entgegen der Erwartung, keinen Unterschied zwischen *Erfahrenen* und *Unerfahrenen* zu finden (PH-D_b), zeigt sich signifikanter Haupteffekt für die *Berufserfahrung*, der als *mittelgroß* zu bewerten ist ($F(1,156) = 8.77$; $p = .00$; $f^2 = .06$). Auf Seiten der *Erfahrenen* lässt sich eine höhere Anzahl nicht erkannter Widersprüche feststellen als bei den *Unerfahrenen*.

Tabelle 33. *Tafel der zweifaktoriellen Varianzanalyse zur statistischen Prüfung der PH-D zur Anzahl der nicht erkannten Widersprüche*

<i>Quelle der Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	$\frac{A-H_0^a}{A-H_1}$	<i>f</i> ²
Berufserfahrung	13.23	1	13.23	8.77	.00	A-H ₁	.06
Fachwissen	22.50	1	22.50	14.92	.00	A-H ₁	.10
Berufserfahrung*Fachwissen	0.63	1	0.63	0.41	.52	A-H ₀	.00
Fehler	235.25	156	1.51				
Total	1046.00	160					

Anmerkungen. Für die Prüfung der Wirkung der *Berufserfahrung* gilt $\alpha = .20$

^a A-H₀/A-H₁: Beibehalten der Nullhypothese (A-H₀) bzw. Annahme der Alternativhypothese (A-H₁)

Ebenfalls signifikant ist der Haupteffekt für das *Fachwissen* ($F(1,156) = 14.92$; $p = .00$; $f^2 = .10$), der einem *mittleren* Effekt entspricht und sich darin ausdrückt, dass Personen mit *relevantem Fachwissen* mehr Widersprüche nicht erkennen als Personen mit *irrelevantem Fachwissen*. Dieser signifikante Befund ist daher aufgrund der unerwarteten Richtung dennoch konträr zur Hypothese (PH-D).

Hinweise auf eine Interaktion der Variablen *Berufserfahrung* und *Fachwissen* konnten im Rahmen der explorativen Analyse nicht gefunden werden ($F < 1$).

5.3.5 PH-E_i: Bestätigungstendenz B (AV 4)

Die Bestätigungstendenz B, die vom Hypothesenkomplex E_i angesprochen wird, kann sich aufgrund der gewählten Operationalisierung einerseits darin ausdrücken, dass konsistenten Informationen eine größere Wichtigkeit zugesprochen wird als inkonsistenten (*Wichtigkeit*) und andererseits durch einen größeren Einfluss im Sinne einer größeren Bekräftigung der Prüfhypothese durch konsistente denn als Entkräftung durch inkonsistente Informationen (*Einfluss auf die Bewährung der Prüfhypothese*).

5.3.5.1 PH-E_i: Wichtigkeit

In der Gesamtstichprobe wurde den (jeweils bedingungsspezifischen) vier konsistenten Items im Mittel eine aggregierte Wichtigkeit von 19.63 ($SD = 3.51$) auf einer Skala von

null bis sechs zugeordnet, den inkonsistenten eine mittlere (aggregierte) Wichtigkeit von 19.78 ($SD = 3.71$). Die mittlere Differenz liegt damit bei -0.15 ($SD = 3.71$) und damit nicht unweit des Nullpunktes, der eine ausgewogene Wichtigkeit anzeigt. Die bedingungsspezifischen deskriptiven Kennwerte können Tabelle 34 entnommen werden.

Tabelle 34. Mittelwerte und Streuungen des Indexwerts für die größere Wichtigkeit konsistenter Informationen in den Bedingungskombinationen der Variablen Berufserfahrung und Fachwissen

		<i>Berufserfahrung</i>	
		<i>erfahren</i>	<i>unerfahren</i>
		<i>M (SD)</i>	<i>M (SD)</i>
<i>Fachwissen</i>	<i>relevant</i>	0.13 (3,34)	-0.03 (2.77)
	<i>irrelevant</i>	-1.25 (4.87)	0.55 (3.45)

Anmerkungen. Mögliche Werte von -24 bis $+24$, da je 4 Informationen auf einer Skala von $0 =$ sehr unwichtig bis $6 =$ sehr wichtig eingeschätzt wurden, was getrennt für konsistente und inkonsistente Items addiert und dann voneinander abgezogen wurde. Positive Werte zeigen die Bestätigungstendenz B an, ein negativer Wert die höhere Gewichtung der inkonsistenten Informationen.

Überprüfung des Zusammenhangs von Prädiktor und Kriterium. Der unterstellte Zusammenhang von *Überzeugtheit von der Prüfhypothese* und der *größeren Wichtigkeit konsistenter Informationen* lässt sich als *kleiner*, signifikanter Effekt bestätigen ($r = 0.17$; $p = .02$), indem eine höhere *Überzeugtheit der Prüfhypothese* mit einer stärkeren Wichtigkeitszuschreibung für konsistente gegenüber inkonsistenten Informationen verbunden ist. Deswegen erfolgt eine kovarianzanalytische Auswertung aller drei bedingenden Variablen.

Kovarianzanalyse. Die Ergebnisse der Kovarianzanalyse sind in Tabelle 35 (siehe folgende Seite) wiedergegeben. In Übereinstimmung mit der *kleinen* positiven Korrelation der *Überzeugtheit von der Prüfhypothese* mit der *größeren Wichtigkeit bei der Beurteilung konsistenter Items* erweist sich die Kovariate als bedeutsam (PH-E_{1a} : $F(1,155) = 4.15$; $p = .04$; $f^2 = .03$), da die Nullhypothese bei Vorhandensein eines *kleinen* Effekts verworfen werden muss.

Bei $\alpha = .20$ muss allerdings ebenfalls die intendierte Nullhypothese bzgl. der *Berufserfahrung* verworfen werden (PH-E_{1b} : $F(1,155) = 2.40$; $p = .12$; $f^2 = .02$), wobei der Effekt als *klein* zu bezeichnen ist. Ein ebenfalls erwartungskonträres Ergebnis stellt die Beibehaltung der Nullhypothese hinsichtlich des *Fachwissens* dar (PH-E_{1c} : < 1) vor. Ein Interaktionseffekt liegt außerdem nicht vor ($F(1,155) = 2.18$; $p = .14$; $f^2 = .01$).

Tabelle 35. *Tafel der Kovarianzanalyse zur statistischen Prüfung der PH-E₁ zur Beurteilung der Wichtigkeit von konsistenten und inkonsistenten Informationen*

<i>Quelle der Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	$\frac{A-H_0^{\alpha}}{A-H_1}$	<i>f²</i>
Überzeugtheit	55.39	1	55.39	4.15	.04	A-H ₁	.03
Berufserfahrung	31.97	1	31.97	2.40	.12	A-H ₁	.02
Fachwissen	2.17	1	2.17	0.16	.69	A-H ₀	.00
Berufserfahrung*Fachwissen	29.10	1	29.10	2.18	.14	A-H ₀	.01
Fehler	2067.36	155	13.34				
Total	2198.00	160					

Anmerkungen. Für die Prüfung der Wirkung der Berufserfahrung gilt $\alpha = .20$

^a A-H₀/A-H₁: Beibehalten der Nullhypothese (A-H₀) bzw. Annahme der Alternativhypothese (A-H₁)

5.3.5.2 PH-E₂: Einfluss auf die Bewährung der Prüfhypothese

In der Gesamtstichprobe wurde den (jeweils bedingungsspezifischen) vier konsistenten Items im Mittel ein aggregierter Einfluss von 4.94 (*SD* = 3.03) auf einer Skala von 3 = *sehr starke Abschwächung der Prüfhypothese* über 0 = *keinen Einfluss* bis +3 = *sehr starke Bekräftigung der Prüfhypothese* zugeordnet, den inkonsistenten ein mittlerer aggregierter Einfluss von -4.38 (*SD* = 4.41). Die mittlere Summe, die die Bestätigungstendenz B anzeigt, liegt damit bei 0.56 (*SD* = 5.85). Die bedingungsspezifischen deskriptiven Kennwerte können Tabelle 36 entnommen werden.

Tabelle 36. *Mittelwerte und Streuungen der Summe des Einflusses von konsistenten und inkonsistenten Items in den Bedingungskombinationen der Variablen Berufserfahrung und Fachwissen*

		<i>Berufserfahrung</i>	
		<i>erfahren</i>	<i>unerfahren</i>
		<i>M (SD)</i>	<i>M (SD)</i>
<i>Fachwissen</i>	<i>relevant</i>	3.35 (6.86)	1.83 (6.28)
	<i>irrelevant</i>	-2.68 (3.77)	-0.25 (4.24)

Anmerkungen. Mögliche Werte von -12 bis +12. Positive Werte zeigen die Bestätigungstendenz B an, ein negativer Wert den größeren Einfluss der inkonsistenten im Sinne einer Entkräftung der Prüfhypothese.

Überprüfung des Zusammenhangs von Prädiktor und Kriterium. Der unterstellte Zusammenhang von *Überzeugtheit von der Prüfhypothese* und des *größeren Einflusses konsistenter Informationen* lässt sich als *mittlerer* signifikanter Effekt bestätigen ($r = 0.38$; $p = .00$), indem eine höhere Überzeugtheit von der Prüfhypothese mit einer stärkeren Einflusszuschreibung für konsistente gegenüber inkonsistenten verbunden ist. Deswegen erfolgt eine kovarianzanalytische Auswertung aller drei bedingenden Variablen.

Kovarianzanalyse. Die Ergebnisse der Kovarianzanalyse sind in Tabelle 37 wiedergegeben. In Übereinstimmung mit der *mittleren* positiven Korrelation der *Überzeugtheit von der Prüfhypothese* mit dem *größeren Einfluss konsistenter Informationen auf die Bewährung der Prüfhypothese* erweist sich die Kovariate als bedeutsam (PH-E_{2a}: $F(1,155) = 20.98$; $p = .00$; $f^2 = .14$), da die Nullhypothese bei Vorhandensein eines *mittleren bis großen* Effekts verworfen werden muss.

Tabelle 37. Tafel der Kovarianzanalyse zur statistischen Prüfung der PH-E₂ zur Beurteilung des Einflusses von konsistenten und inkonsistenten Informationen

Quelle der Variation	SS	df	MS	F	p	$\frac{A-H_0^a}{A-H_1}$	f ²
Überzeugtheit	551.97	1	551.97	20.98	.00	A-H ₁	.14
Berufserfahrung	18.05	1	18.05	0.69	.41	A-H ₀	.00
Fachwissen	488.74	1	488.74	18.58	.00	A-H ₁	.12
Berufserfahrung*Fachwissen	101.80	1	101.80	3.87	.05	A-H ₁	.02
Fehler	4077.18	155	26.30				
Total	5500.00	160					

Anmerkungen. Für die Prüfung der Wirkung der Berufserfahrung gilt $\alpha = .20$

^a A-H₀/A-H₁: Beibehalten der Nullhypothese (A-H₀) bzw. Annahme der Alternativhypothese (A-H₁)

Während erwartungskonform kein Haupteffekt für die *Berufserfahrung* vorliegt (PH-E_{2b}: $F < 1$), ist in Bezug auf das *Fachwissen* (PH-E_{2c}: $F(1,155) = 18.58$; $p = .00$; $f^2 = .12$) ein solcher zu beobachten, der dahingehend zu interpretieren ist, dass Probanden mit *relevantem Fachwissen* konsistenten Informationen einen noch größeren Einfluss auf die Bewährung der Prüfhypothese zuschreiben als Probanden der *Bedingung Fachwissen irrelevant (Brandstiftung)*.

Außerdem liegt ein Interaktionseffekt vor ($F(1,155) = 3.87; p = 0.05; f^2 = .02$), der in Abbildung 12 veranschaulicht wird. Während bei den *Erfahrenen* die Bestätigungstendenz B vom *relevanten Fachwissen* hin zum *irrelevanten Fachwissen* ins genaue Gegenteil umschlägt, neutralisiert sich die etwas geringer ausgeprägte Bestätigungstendenz B der *Unerfahrenen* in der Bedingung *Fachwissen relevant* in der Bedingung *Fachwissen irrelevant*.

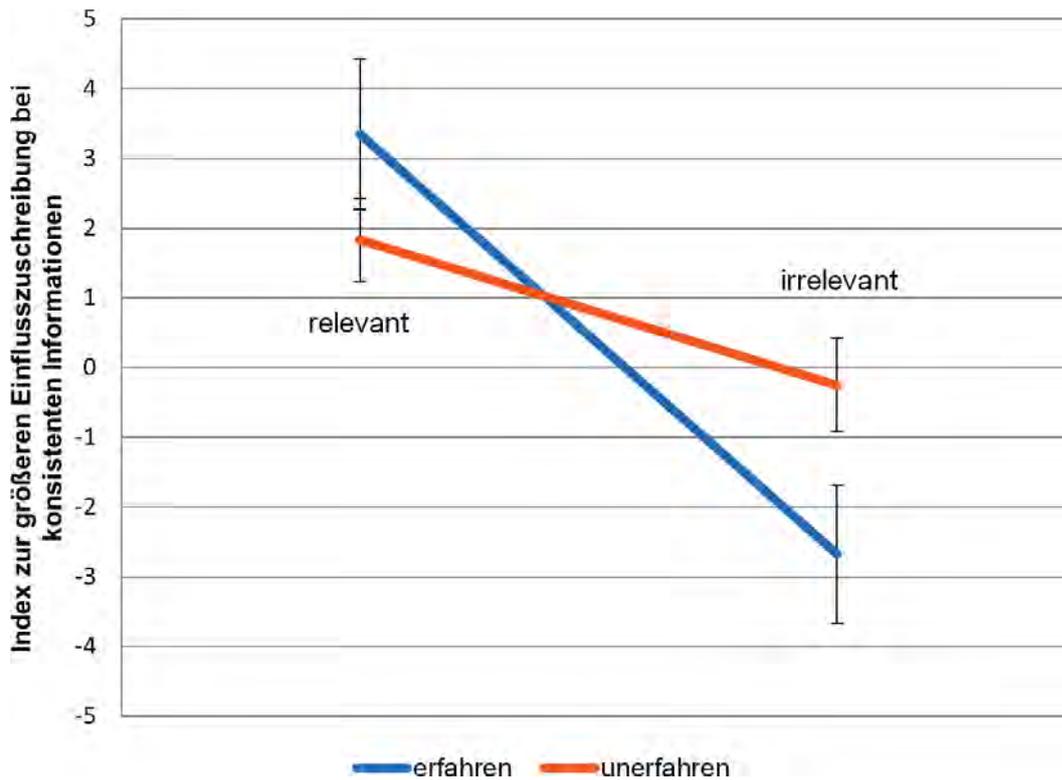


Abbildung 12. Veranschaulichung des Interaktionseffekts zwischen Berufserfahrung und Fachwissen. Positive Werte zeigen eine größere Bekräftigung der Prüfhypothese durch konsistente Informationen als eine analoge Entkräftung durch inkonsistente an.

5.3.6 PH-F_i: Verhaltenskonsequenzen (AV 6)

Die Variable *Verhaltenskonsequenzen (AV 6)* wurde ausschließlich in der Bedingung *Fachwissen relevant (Essstörung)* erhoben (für die inhaltliche Argumentation s. Abschnitt 4.2.6). Die Analysen beziehen sich dementsprechend auf die Angaben von $n = 80$ Probanden in Abhängigkeit ihrer *Berufserfahrung* und *Überzeugtheit von der Prüfhypothese*, die bzgl. der Themen in der ersten Therapiesitzung bzw. Hausaufgaben formuliert wurden.

5.3.6.1 PH-F₁: Themen der ersten Therapiesitzung

Tabelle 38 enthält die Häufigkeiten, mit denen die *Unerfahrenen* bzw. *Erfahrenen* Themen genannt haben, die einer von sieben Kategorien im Zusammenhang mit der Konsistenz des Themas bzgl. Essstörungen zugeordnet wurden.

Tabelle 38. Absolute Häufigkeiten bzgl. der Konsistenz von Themenvorschlägen mit der Therapie von Essstörungen

Kategorie	Unspezifisch	Eher Bulimie	Eher Anorexie	Beide gleich	Keine/ nichts	Nicht bewertbar	fehlend
Unerfahrene	15	4	3	13	0	4	1
Erfahrene	24	1	3	7	2	3	0

Tabelle 38 kann entnommen werden, dass am häufigsten *unspezifische* Themen genannt wurden. Diese bezogen sich in der Mehrzahl auf die Herstellung einer therapeutischen Beziehung bzw. der Evaluation der Therapiemotivation. Die Kategorien *unspezifisch*, *keine/nichts* sowie *nicht bewertbar* (s. Anhang D) stellen hingegen rein technische Kategorien dar, die eingeführt wurden, um die eindeutige Bewertbarkeit jeder Freitextantwort zu ermöglichen. Inhaltlich bedeutsam für die vorliegende Fragestellung sind insbesondere die Häufigkeiten, mit denen Themen angesprochen werden, die eher im Rahmen einer Bulimie-Therapie eingesetzt werden, weil diese wie im aktuellen Fall der Anorexie-Patienten weniger indiziert sind. Aufgrund der geringen Anzahl von Nennungen wird jedoch von einer Auswertung abgesehen, weil keine aussagekräftigen Ergebnisse vor dem Hintergrund der äußerst geringen Datenbasis zu erwarten sind.

5.3.6.2 PH-F₂: Hausaufgaben

Analog zu der Situation bzgl. der Themenvorschläge ist ebenfalls Tabelle 39 zu entnehmen, dass keine ausreichende Datenbasis zur Überprüfung der Abhängigkeit von *Berufserfahrung* und *Überzeugtheit der Prüfhypothese* gegeben ist.

Tabelle 39. Absolute Häufigkeiten bzgl. der Konsistenz von Themenvorschlägen mit der Therapie von Essstörungen

Kategorie	Unspezifisch	Eher Bulimie	Eher Anorexie	Beide gleich	Keine/ nichts	Nicht bewertbar	fehlend
Unerfahrene	7	2	1	21	3	6	0
Erfahrene	15	0	0	15	5	5	0

6 Diskussion

Im letzten Abschnitt der Arbeit sollen zunächst die Ergebnisse der Studie interpretiert und im Hinblick auf ihre Bedeutung für die Hypothesen diskutiert werden (6.1). Sodann schließt sich eine Bewertung der Befunde im Hinblick auf die Generalisierbarkeit an (6.2), die durch eine kritische Würdigung der Arbeit abgerundet wird (6.3). Abschließend werden die Implikationen für die klinische Praxis und die weitere Forschung zusammengefasst (6.4).

6.1 Interpretation der Ergebnisse der vorliegenden Studie

Nach der Bewertung der internen Validität der Studie werden die Ergebnisse getrennt für die drei Einflussvariablen *Überzeugtheit von der Prüfhypothese*, *Berufserfahrung* und *Fachwissen* interpretiert.

6.1.1 Ergebnisinterpretation: Bewertung der internen Validität

Um im Folgenden die Ergebnisse sinnvoll interpretieren zu können, muss zuvor die interne Validität der Studie bewertet werden. Dafür ist Stellung zu beziehen, ob die relevanten Störvariablen eliminiert oder konstant gehalten werden konnten.

Da lediglich 5% der Studienteilnehmer in der Formulierung des Studienziels in eigenen Worten entweder explizit oder implizit zu erkennen gaben, dass sie die Studie in Zusammenhang mit konfirmatorischen Prozessen gebracht haben, und sie deswegen von der Datenanalyse ausgeschlossen wurden, kann die Störvariable *Bewusstsein hinsichtlich konfirmatorischer Prozesse* als eliminiert gelten.

Durch den permanenten Zugriff auf Hintergrundinformationen wurde das *Minimum an prüfhypothesenbezogenem Wissen sichergestellt*. Da außerdem alle Erstsemester aus der Gruppe der Unerfahrenen ausgeschlossen wurden, ist zudem anzunehmen, dass alle Probanden in der Lage waren, von den bereitgestellten Informationen zu profitieren. Die Störvariable gilt als erfolgreich eliminiert. Die Störvariable *Der diagnostische Prozess als soziale Interaktion* wurde durch die Konzeption als Onlinestudie ausgeschaltet. In Abschnitt 5.2.5 wurde der Nachweis erbracht, dass die *randomisierte Zuweisung der Probanden* auf die beiden Gruppen der Variablen *Fachwissen* erfolgreich war, da Alter und Geschlecht wie erwartet gleichverteilt sind. Störeinflüsse durch eine verzerrte Gruppenbildung sind nicht zu erwarten.

Die Interpretierbarkeit der Ergebnisse im Hinblick auf die Hypothesen ist daher gegeben.

6.1.2 Ergebnisinterpretation: Überzeugtheit von der Prüfhypothese

Für den Einfluss der Überzeugtheit von der Prüfhypothese lässt sich die globale Hypothese formulieren „*Je höher die Überzeugtheit von der Prüfhypothese, desto höher die Wahrscheinlichkeit für eine verzerrte Informationsverarbeitung und ein verzerrtes Urteil.*“ Die Befunde zu den einzelnen abhängigen Variablen sollen an dieser Stelle bewertet und mit den einschlägigen empirischen Arbeiten in Beziehung gesetzt werden.

6.1.2.1 PH-A_{ia}: Verzerrtes Urteil

Im Zusammenhang mit dem verzerrten Urteil wurden einerseits die Schlussfolgerung (AV 1) und andererseits die subjektive Sicherheit (AV 2) analysiert.

PH-A_{ia}: Selbstbestätigendes Urteil. Die Annahme, dass die Überzeugtheit von der Prüfhypothese eine Vorhersagekraft für den *confirmation bias* besitzt, hat sich bestätigt, da ein bedeutsamer Anstieg der Wahrscheinlichkeit eines selbstbestätigenden Urteils beim Anstieg der Überzeugtheit gezeigt werden konnte (PH-A_{ia}).

Ähnlich wie in der Studie von Sandifer et al. (1970), die den Einfluss der ersten drei Minuten in einem psychiatrischen Interview unterstreicht, da die spätere Diagnose bereits aus den Aufzeichnungen der Probanden innerhalb der ersten drei Minuten vorhersagbar war, kommt der anfänglichen Wahrnehmung des Patienten und einer in diesem Zusammenhang vermittelten Prüfhypothese bereits eine Bedeutung für das spätere Urteil zu. Wenn nun bereits mithilfe der anfänglichen Überzeugtheit eine Vorhersage der Validität des Urteils gemacht werden kann, deutet dies an, dass nachfolgende Informationen nicht in angemessener Weise in die Urteilsbildung einbezogen werden. Dieser Umstand korrespondiert mit dem Befund von Dailey (1952), dass eine frühzeitige Schlussfolgerung die Fähigkeit behindern kann, von nachfolgenden Daten zu profitieren und diese korrekt in den Eindruck zu integrieren.

Dennoch muss im Hinblick auf die Bedeutsamkeit betont werden, dass die in der vorliegenden Arbeit gefundene Effektstärke für die Überzeugtheit von der Prüfhypothese als klein zu bewerten ist. Die Vorhersage ist dementsprechend nicht fehlerfrei, so dass weitere Einflussfaktoren anzunehmen sind. Allerdings ist auch durchaus von einem robusten Effekt zu sprechen, denn aufgrund der Tatsache, dass in der vorliegenden Arbeit zur Operationalisierung der weiteren Einflussvariablen Fachwissen zwei unterschiedliche Fallgeschichten verwendet wurden und die Gesamtstichprobe hinsichtlich ihres Alters

und ihrer Berufserfahrung vergleichsweise heterogen zusammengesetzt ist, kann ausgeschlossen werden, dass es sich um einen reinen Material- oder Stichprobeneffekt handelt.

Außerdem unterstreicht der Umstand, dass die *subjektive* Überzeugung von der Hypothese diesen kleinen, robusten Effekt aufwies, die von Haverkamp (1993) aufgezeigte Notwendigkeit, in nachfolgenden Forschungsarbeiten diese subjektive Perspektive stärker zu berücksichtigen und nicht stillschweigend zu unterstellen, die Probanden testeten stets eine vorgegebene Hypothese. Aufgrund dieser Sonderstellung der vorliegenden Arbeit ist ein Vergleich mit anderen Studien nicht angezeigt. Analog gilt dies für den Befund bzgl. der *subjektiven Sicherheit*.

PH-A_{2a}: Subjektive Sicherheit. Die Annahme, dass mit einer höheren *Überzeugtheit* von der *Prüfhypothese* auch eine höhere *subjektive Sicherheit* einhergeht (*PH-A_{2a}*), konnte hingegen nicht bestätigt werden. Da die mittlere subjektive Sicherheit allerdings im mittleren Skalenbereich lag (s. Abschnitt 5.3.1.2), kann für das Ausbleiben der systematischen Abhängigkeit weder ein Boden- noch ein Deckeneffekt verantwortlich gemacht werden, weswegen methodische Gründe eher unwahrscheinlich erscheinen und die subjektive Sicherheit von anderen Parametern abzuhängen scheint.

Aus inhaltlicher Sicht war bei Personen mit höherer *Überzeugtheit* von der *Prüfhypothese* erwartet worden, dass diese durch die stärker verzerrte Informationsverarbeitung eher zu einem bestätigenden Urteil kommen und durch dieses verzerrte Bild mit reduzierten Inkonsistenzen ein hohes Vertrauen in die *Prüfhypothese* bzw. die getroffene Entscheidung entwickeln. Für Personen mit niedrigerer *Überzeugtheit* war eine weniger verzerrte Testung unterstellt worden, die u.a. die Wahrnehmung der Inkonsistenzen und in Folge dessen ein korrektes Urteil begünstigen sollte. Dabei wurde nun eine reduzierte subjektive Sicherheit angenommen, denn immerhin steht die eigene Meinung im Widerspruch zu der von einer statushohen Person induzierten *Prüfhypothese*.

Allerdings ist bei dieser Vermutung ein sehr enger, d.h. großer Zusammenhang von der *Überzeugtheit* der *Prüfhypothese* und dem verzerrten Urteil zu unterstellen, der in der vorliegenden Arbeit aber, wie oben diskutiert, nur als klein einzustufen ist. Insofern muss man davon ausgehen, dass es einen nicht unbeträchtlichen Teil von Probanden gibt, die trotz einer hohen *Überzeugtheit* von der *Prüfhypothese* später ein korrektes Urteil fällen. Es könnte nun sein, dass gerade diese Personen eine geringe subjektive Sicherheit aufweisen, weil sie im Laufe des Testungsprozesses einen Hypothesenwechsel vorgenommen haben, der zudem der Meinung der statushohen Person widerspricht. Desweiteren wird es aufgrund des kleinen Zusammenhangs Personen geben, die trotz

einer anfangs geringen subjektiven Sicherheit später das verzerrte Urteil fällen. Um diesen Wechsel zu vollziehen, wird eine große Menge an subjektiv wahrgenommener Evidenz vorgelegen haben müssen, und vermutlich in einer hohen subjektiven Sicherheit resultieren.

Wenn nun ein Teil der Probanden mit anfänglich hoher Überzeugtheit von der Prüfhypothese aufgrund der Bestätigung eine hohe subjektive Sicherheit aufweist, ein anderer Teil jedoch wegen des korrekten Urteils eine niedrige, so reduziert sich die subjektive Sicherheit im Mittel. Analog kann es bei Probanden mit anfänglich geringer Überzeugtheit von Prüfhypothese zu einer im Mittel höheren subjektiven Sicherheit kommen, weil diejenigen, die später das verzerrte Urteil fällen hierbei u.U. besonders sicher sind. Als Resultat mitteln sich die Effekte dann aus, so dass keine Beeinflussung mehr messbar ist.

Es wäre daher möglich, dass der Zusammenhang zwischen der Überzeugtheit von der Prüfhypothese und der subjektiven Sicherheit von der Korrektheit des Urteils moderiert wird, und deshalb in der vorliegenden Arbeit kein erwartungskonformes Ergebnis gefunden werden konnte.

6.1.2.2 *PH-B_{ia}: Asymmetrische Evidenzregeln*

Im Zusammenhang mit der Asymmetrie der Evidenzregeln (AV 3) sind erstens ein Maß für die positive Teststrategie und zweitens vier Maße im Zusammenhang mit der pseudodiagnostischen Vorgehensweise untersucht worden. Da nach Wissen der Verfasserin keine empirischen Arbeiten vorliegen, die die Generierung von Evidenzregeln untersuchen, ist ein Vergleich mit einschlägigen Arbeiten nur sehr bedingt möglich und im Folgenden nur dort angeführt, wo entsprechende Analogien gezogen werden konnten.

Präferenz für positive Evidenzregeln (PH-B_{ia}). Die Erwartung, mit steigender Überzeugtheit von der Prüfhypothese eine zunehmende Präferenz für die positive Teststrategie im Rahmen der Evidenzregeln zu finden, wurde bestätigt. Die beobachtete Effektstärke ist als mittelgroß zu klassifizieren und damit als bedeutsam einzuschätzen.

Interessanterweise wurde in den empirischen Arbeiten zur positiven Teststrategie im Rahmen der Frageformulierung bei erfahrenen Diagnostikern keine primäre Präferenz für die positive Teststrategie gefunden (Hayden, 1987; Strohmer & Chiodo, 1984; Strohmer & Newman, 1983; Wilson Dallas & Baron, 1985). Vielmehr neigten die Diagnostiker dazu, offene Fragen zu generieren, sofern die Aufgabenstellung dies zuließ.

Haverkamp (1993) hatte sich allerdings bereits kritisch dazu geäußert, dass diese scheinbare Präferenz für offene Fragen auch eine Demonstration erwünschten Verhaltens in einem Bewertungskontext oder der Konfundierung von Frageform und Informationsausschöpfung sein könnte, da letztere bei offenen Fragen i.d.R. höher sein wird als bei geschlossenen.

Im Falle der Evidenzregeln, die einen Zusammenhang zwischen Hypothese und (nicht) zu erwartendem Ereignis herstellen, ist diese Option der „offenen“ Formulierung aufgrund der Aufgabenstellung verwehrt, ohne dass dem Hypothesentester eine künstliche Restriktion durch den Versuchsleiter auferlegt wird. Eine Ausgewogenheit kann nur darüber hergestellt werden, dass ähnlich viele positive wie negative Evidenzregeln gebildet werden. Bei einem relativen Anteil von 84% positiven Evidenzregeln ist jedoch von einer außerordentlichen Dominanz der positiven Teststrategie zu sprechen. Dies entspricht sodann auch dem Befund von Haverkamp (1993) für die *self generated hypothesis*, für die das Testungsverhalten auch mehrheitlich (64%) als konsistent mit der Hypothese bewertet wurde und als positive Teststrategie zu klassifizieren ist.

Präferenz für (positive) Evidenzregeln mit nicht diagnostischen Inhalten (PH-B_{2a} & PH-B_{3a}). Aufgrund der nicht hinreichenden Beurteilerübereinstimmung hinsichtlich der Diagnostizität der Inhalte der Evidenzregeln kann über die Bewährung der Hypothesen zur Präferenz für (positive) Evidenzregeln mit nicht diagnostischen Inhalten keine Aussage gemacht werden. Ein Vergleich mit einschlägigen empirischen Arbeiten entfällt daher.

Da kein Hinweis auf die Gültigkeit der Einschätzungen erbracht werden konnte, wurde von einer Auswertung abgesehen, obwohl in der Literatur durchaus der Hinweis zu finden ist, dass auch geringe Koeffizienten für eine hinreichende Übereinstimmung sprechen können, wenn das zu beurteilende Merkmal schwer zu erfassen ist (z.B. Wirtz & Caspar, 2002). Sie begründen dies mit der an sich richtigen Position, dass die Richtlinien für die Einschätzung der Höhe der Koeffizienten notwendiger Weise nur einen groben Anhaltspunkt darstellen können und sachlogische Gesichtspunkte nicht vollständig ersetzen dürfen. Und zweifelsohne ist die Diagnostizität genau ein solch schwer zu erfassendes Merkmal. Selbst im Essstörungskontext, wo die Diagnostizität an Studienergebnissen festgemacht werden kann, dass ein Merkmal *X* eine hohe Korrelation zur Prüfhypothese und eine niedrige (oder gar keine) zur Alternativhypothese aufweist, so bleibt es Auslegungssache der Beurteiler, ab wann die Differenz hinreichend groß ist. Noch uneindeutiger ist dies im Brandstiftungskontext, wo jegliche empirische Basis für die differentielle Auftretenswahrscheinlichkeit fehlt. Insofern überrascht es nicht, dass die Übereinstimmung in dieser Situation geringer war als in der klinischen Diagnostik.

Im Sinne eines konservativen Vorgehens wird von der Option, eine geringe numerische Beurteilerübereinstimmung aus sachlogischen Gründen dennoch für ausreichend zu erklären, Abstand genommen, weil die Basis, die dadurch für die Beurteilung der Bewährung der inhaltlichen Hypothesen geschaffen würde, als nicht ausreichend wissenschaftlich fundiert erachtet wird.

Relevanter erscheint der - vorsichtig als Beobachtung zu kennzeichnende - Umstand, dass selbst im Falle der Essstörungsdiagnostik von den unbeeinflusst agierenden Ratern nur eine akzeptable Übereinstimmung erzielt werden konnte, obwohl die bereitgestellten Hintergrundinformationen die relevanten Angaben zur Diagnostizitätsbestimmung zumindest insofern enthielten, als textuell formuliert war, welche Merkmale im besonderen die Bulimiker auszeichnen und welche die Anorektiker. Diese Angaben müssen daher von den Ratern unterschiedlich interpretiert bzw. gewichtet worden sein. Dies liefert zumindest einen Hinweis dafür, dass die Bewertung der Diagnostizität ein außerordentlich komplexer und kognitiv herausfordernder Prozess ist, denn weil die Rater in den anderen zu beurteilenden Merkmalen deutlich höhere Übereinstimmungen erzielten, muss nicht in erster Linie davon ausgegangen werden, dass spezifische Gründe in den Individuen der Rater für die mangelnde Übereinstimmung hinsichtlich der Diagnostizität liegen.

Berücksichtigung alternativer Hypothesen ($PH-B_{4a}$) und Gewichtung ($PH-B_{5a}$). Mit den Maßen bzgl. der Berücksichtigung alternativer Hypothesen und deren Gewichtung beim Generieren von Evidenzregeln stehen jedoch weitere Aspekte des Testungsverhaltens zur Verfügung, die im Zusammenhang mit dem pseudodiagnostischen Testen nach Trope und Liberman (1996) stehen.

Wenngleich der angenommene negative Zusammenhang zwischen steigender Überzeugtheit von der Prüfhypothese und der Anzahl der insgesamt berücksichtigten Hypothesen die Signifikanz bei einer kleinen Effektstärke verfehlt, ließ sich erwartungskonform beobachten, dass die Anzahl der Evidenzregeln für die Prüfhypothese mit der Überzeugtheit von der Prüfhypothese signifikant zunimmt. Je überzeugter eine Person demnach von ihrer Prüfhypothese ist, desto selektiver wird die Informationssuche, indem die Prüfhypothese anderen Hypothesen gegenüber bevorzugt wird. Letzteres ist erneut als kleiner, robuster Effekt zu interpretieren. Einschlägige Arbeiten für einen Vergleich sind der Verfasserin nicht bekannt.

Fazit. Zusammenfassend ist hinsichtlich der Evidenzregeln anzuführen, dass die Überzeugtheit von der Prüfhypothese einen kleinen, aber robusten Effekt auf die

Asymmetrie der Evidenzregeln sowohl im Hinblick auf den Einsatz der positiven Teststrategie als auch bei der vermehrten Fokussierung auf die Prüfhypothese zu Ungunsten der Alternativen ausübt, indem die Prüfhypothese mit zunehmender Überzeugtheit häufiger Eingang in die Evidenzregeln findet als die Alternativhypothese. Über die Berücksichtigung der Diagnostizität der relevanten Merkmale kann keine Aussage gemacht werden.

6.1.2.3 PH-C: *Hypothesenkonsistente Interpretation uneindeutiger Evidenz*

Die *Überzeugtheit von der Prüfhypothese* hat sich entgegen der Erwartung nicht als geeignet erwiesen, das Ausmaß der Aufwertungen von uneindeutigen Informationen vorherzusagen. Es kam nicht systematisch häufiger dann zu einer stärkeren hypothesenkonsistenten Interpretation, wenn die *Überzeugtheit von der Prüfhypothese* höher war.

Ein tatsächlicher Vergleich mit anderen Studien ist kaum möglich, da die *Überzeugtheit von der Prüfhypothese* keine Entsprechungen in den einschlägigen Arbeiten findet, die mit der hypothesenkonsistenten Interpretation uneindeutiger Evidenz in Verbindung gebracht wurden (Darley & Gross, 1983; Lord et al., 1979; Strohmer et al. 1990, Experiment 2).

Als grundsätzlicher Unterschied ist jedoch zusätzlich festzustellen, dass die Operationalisierung der Uneindeutigkeit in der vorliegenden Arbeit eine andere ist. Hier wurde die Uneindeutigkeit auf Itemebene hergestellt, indem solche Items präsentiert wurden, die sowohl aufgrund von sachlogischen Überlegungen als auch von Ratureinschätzungen als uneindeutig klassifiziert werden konnten. Eine hypothesenkonsistente Interpretation wurde dann festgestellt, wenn einem Item Erklärungswert zugesprochen wurde, den es aufgrund seiner Uneindeutigkeit faktisch nicht haben dürfte.

In den anderen Arbeiten wurde die Uneindeutigkeit eher auf einer übergeordneten Ebene hergestellt, indem eine Teilmenge konsistenter und eine Teilmenge inkonsistenter Informationen vorgegeben wurde, die in ihrer Gesamtheit als uneindeutig zu bewerten sind. Die hypothesenkonsistente Interpretation wurde dadurch angezeigt, dass mehr konsistente als inkonsistente Informationen ausgewählt wurden, wenn es bspw. um die Auflistung der 10 wichtigsten Informationen ging. Insofern ergibt sich hier eine Überschneidung zwischen der hypothesenkonsistenten Interpretation uneindeutiger Evidenz in Sinne von Schulz-Hardt und Köhnken (2000) und der Bestätigungstendenz B sensu Hager und Weißmann (1991), die direkt auf die differentielle Wertigkeitsbeurteilung für konsistente und inkonsistente Items eingehen. Die Diskussion verlagert sich daher sinnvollerweise auf diejenige zur Bestätigungstendenz B.

6.1.2.4 PH-D: Bestätigungstendenz I

Auch die Abhängigkeit der Bestätigungstendenz I von der *Überzeugtheit von der Prüfhypothese* kann mit den vorliegenden Daten nicht bestätigt werden: die Erkennensleistung von Widersprüchen sank nicht wie unterstellt mit zunehmender Überzeugtheit von der Prüfhypothese. Da die Bestätigungstendenzen nach Hager und Weißmann (1991) keinen Einzug in die klassischen Arbeiten zum konfirmatorischen Hypothesentestens gefunden haben, stehen auch für die Bestätigungstendenz I keine direkten Vergleichsarbeiten zur Verfügung.

6.1.2.5 PH-E_{ia}: Bestätigungstendenz B

Die Hypothese, eine zunehmende Bestätigungstendenz B bei zunehmender Überzeugtheit von der Prüfhypothese vorzufinden, konnte sowohl im Hinblick auf die Wichtigkeit als auch den Einfluss auf die Bewährung der Prüfhypothese untermauert werden. Wie erwartet wird die Wichtigkeit von konsistenten Informationen gegenüber inkonsistenten mit zunehmender Überzeugtheit von der Prüfhypothese größer eingeschätzt. Ebenso erhöht sich der Einfluss konsistenter gegenüber inkonsistenten Informationen mit steigender Überzeugtheit von der Prüfhypothese. Letzteres unterliegt einem mittleren bis großen Effekt, ersteres einem kleinen. Die Überzeugtheit von der Prüfhypothese hat sich damit als bedeutsam für das Ausmaß der Bestätigungstendenz B erwiesen.

Wie im Zusammenhang mit der hypothesenkonsistenten Interpretation uneindeutiger Evidenz bereits angedeutet, kann aufgrund der Operationalisierung der Uneindeutigkeit in den Arbeiten von Darley und Gross (1983), Lord et al. (1979) sowie Strohmmer et al. (1990, Experiment 2) eine Parallele zur Bestätigungstendenz B gefunden werden, so dass hier ein zur Literatur stimmiger Befund vorliegt, wenngleich auch hier zu berücksichtigen ist, dass die Zunahme der Bestätigungstendenz B ein Spezifikum der vorliegenden Arbeit ist.

6.1.2.6 PH-F: Verhaltenskonsequenzen

Aussagen über die Bewährung der Hypothese, dass mit zunehmender Überzeugtheit von der Prüfhypothese auch die Wahrscheinlichkeit steigt, nicht indizierte Elemente in die Therapie zu integrieren, sind aufgrund der vorliegenden Daten nicht möglich.

Die Probanden mit relevantem Fachwissen (Essstörung) schlugen am häufigsten (49%) *unspezifische* Themen für die erste Therapiesitzung vor ($PH-F_1$), wobei damit Aspekte wie die Herstellung einer therapeutischen Beziehung und der Evaluation der Therapiemotivation gemeint sind, die bei jeder therapeutischen Intervention relevant sind und somit keine spezifische Funktion für die Behandlung von Essstörungen erfüllen. Solche Themen, die von den Ratern als entweder bulimie- oder anorexiespezifisch eingeschätzt wurden und damit Hinweise für die Hypothese liefern könnten, wurden nur von 11% der Teilnehmer genannt.

In Bezug auf die Hausaufgaben ($PH-F_2$) gaben wiederum 28% der Probanden *unspezifische* Vorschläge ab, wie bspw. die Beschäftigung mit eigenen Zielen und Wünschen. 45% wählten Aufgaben, die sowohl für Bulimiker als auch Anorektiker indiziert sind. Vornehmlich wurde das Führen eines Essstagebuchs genannt. Nur drei Probanden (4%) nannten Maßnahmen, die störungsspezifisch für Anorexie oder Bulimie waren, weswegen die Auswertungsgrundlage entfällt.

Zu diesem frühen Zeitpunkt im (fiktiven) Ablauf der Therapie, nämlich der Therapieplanung nach festgelegter Diagnose legten die Probanden mit ihren Themen- und Hausaufgabenvorschlägen daher eher Grundlagen für die Therapie, was begrüßenswert ist. Es bleibt aber die Frage offen, ob zu einem späteren Zeitpunkt potentielle Verzerrungseffekte auftreten, wie bspw. bei Haverkamp (1993).

6.1.3 Ergebnisinterpretation: Berufserfahrung

Für den Einfluss der Berufserfahrung lässt sich die globale Hypothese formulieren „*Personen mit Berufserfahrung unterscheiden sich nicht von Personen ohne Berufserfahrung im Hinblick auf eine verzerrte Informationsverarbeitung und ein verzerrtes Urteil.*“ Da dem entsprechend für alle bedingten Variablen erwartet wurde, keine Unterschiede zu finden, erfolgt die Darstellung nicht differenziert nach den einzelnen Variablen, sondern gruppiert nach der Erwartungskonformität der Ergebnisse.

Erwartungskonform. Wie erwartet ergaben sich hinsichtlich der meisten bedingten Variablen keine bedeutsamen Unterschiede zwischen den Erfahrenen und den Unerfahrenen. Konkret gab es keine Vorhersagekraft der Berufserfahrung für ein selbstbestätigendes Urteil ($PH-A_{1b}$) und keinen Unterschied hinsichtlich der subjektiven Sicherheit ($PH-A_{2b}$).

Außerdem berücksichtigten die Erfahrenen nicht mehr oder weniger Alternativen als Unerfahrene ($PH-B_{4b}$) und wiesen eine vergleichbar ausgeprägte Präferenz für Evidenzregeln zur Prüfhypothese auf ($PH-B_{5b}$). Die Berufserfahrung schützte die Erfahrenen desweiteren wie erwartet nicht davor, eine vergleichbare Anzahl von Aufwertungen bei der Interpretation uneindeutiger Informationen vorzunehmen wie die Unerfahrenen ($PH-C_b$).

Zusammenfassend kann daher festgehalten werden, dass die Berufserfahrung wie erwartet keinen Einfluss auf die Validität des Urteils sowie das pseudodiagnostische Vorgehen beim Formulieren von Evidenzregeln hatte, wobei zu berücksichtigen ist, dass diese Befunde in ihrer Aussagekraft dahingehend eingeschränkt sind, dass sie auf der Beibehaltung der statistischen Nullhypothese basieren.

Dennoch ist die Prüfung des Einflusses der Berufserfahrung in der vorliegenden Studie vergleichsweise streng gewesen, da es einen bedeutsamen Unterschied zwischen den beiden Gruppen gab, während Haverkamp (1993) und Parmley (2006) die Abwesenheit eines Effekts für die Berufserfahrung anhand des Ausbleibens von Unterschieden in explorativen Untersuchungen ihrer eher homogenen Stichproben ermittelten. Dieselbe Schwäche hatte auch die eigene Vorarbeit (Dreger, 2006) mit sich gebracht, da dort Studierende unterschiedlicher Fachsemester als Unerfahrene bzw. Erfahrene definiert worden waren.

Nicht bewertbar. Erneut sind die Hypothesen im Zusammenhang mit der Diagnostizität der Merkmale der Evidenzregeln ($PH-B_{2b}$ & $PH-B_{3b}$) aufgrund einer unzureichenden Datenbasis nicht prüfbar. Die Argumentation gilt analog zu derjenigen im Zusammenhang mit der Überzeugtheit von der Prüfhypothese (s. Abschnitt 6.1.2.6).

Erwartungskonträr. Es gab jedoch auch erwartungskonträre Befunde. Die Präferenz für positive Evidenzregeln seitens der Erfahrenen war größer als diejenige der Unerfahrenen ($PH-B_{1b}$), außerdem erkannten sie weniger Widersprüche als die Unerfahrenen und unterlagen somit einer stärkeren Bestätigungstendenz I ($PH-D_b$).

In vergleichbarer Weise zeigten sie eine stärkere Bestätigungstendenz B, da sie den konsistenten Informationen eine größere Wichtigkeit in Relation zu den inkonsistenten zuwies als die Unerfahrenen ($PH-E_{1b}$). Betrachtet man hingegen die Bestätigungstendenz B aus der Perspektive der differentiellen Einflusszuschreibung zu konsistenten und inkonsistenten Informationen ($PH-E_{2b}$), so ergibt sich ein anhand des Fachwissens zu differenzierendes Bild: während Erfahrene im Falle von relevantem Fachwissen (Essstörung) den konsistenten Informationen eine größere Bekräftigung der Prüfhypothese zurechnen als den inkonsistenten an Entkräftung, so kehrt sich dieses Verhältnis in der Situation irrelevanten Fachwissens um. Dort wird v.a. der Einfluss

inkonsistenter Informationen betont, während Unerfahrene eine nahezu ausgeglichene Einflusszuschreibung vornehmen. Praktisch bedeutsam ist selbstverständlich die Situation erfahrener Personen mit relevantem Fachwissen, die interessanterweise die meisten Verzerrungen im Sinne der Bestätigungstendenz B mit sich gebracht hat.

Bevor eine inhaltliche Diskussion sinnvoll ist, muss zunächst an den Umstand erinnert werden, dass für die statistische Prüfung der Hypothesen bzgl. der Berufserfahrung zur indirekten Minimierung des Fehlers zweiter Art ein Signifikanzniveau von $\alpha = .20$ gewählt wurde. Dies konfliktiert gerade in der Situation des multiplen Testens durch die große Anzahl an abhängigen Variablen mit der dadurch bedingten Gefahr der Kummulierung des Fehlers erster Art. Gerade bei der nur kleinen Effektstärke im Zusammenhang mit der Präferenz für positive Evidenzregeln ist die Bedeutsamkeit der Berufserfahrung fraglich, zumal sie inhaltlich nicht in das Bild passt, das die Befunde von Martin (2000) aufzeigen, der mehr diskonfirmatorische Fragen bei seinen Erfahrenen fand; keine Unterschiede jedoch bei der Anzahl positiver und neutraler Fragen. Ob ein statistischer Zufallseffekt oder doch die methodische Unterschiedlichkeit für die statistische Signifikanz in der vorliegenden Arbeit verantwortlich ist, kann nicht abschließend beantwortet werden.

Im Zusammenhang mit einer stärker verzerrten Informationsbewertung im Sinne der Bestätigungstendenz I und B sind hingegen mittlere Effektstärken zu beobachten, was einen statistischen Zufallseffekt zumindest etwas unwahrscheinlicher werden lässt. Man könnte sich daher veranlasst sehen, zu folgern, dass mit zunehmender Berufserfahrung der Umgang mit Widersprüchen scheinbar unkritischer wird. Was sich bspw. darin ausdrücken kann, dass sie nicht als solche erkannt bzw. benannt werden, ihre Wichtigkeit geringer eingeschätzt wird und die Entkräftung der Prüfhypothese kleiner ausfällt, wie die drei Maße der vorliegenden Arbeit es nahelegen.

Deshalb drängt sich die Frage auf, wie ein solch negativer Effekt der Berufserfahrung erklärbar ist und wieso das Urteil selbst trotzdem eine vergleichbare Validität aufweisen konnte, obwohl man bei einer scheinbar geringeren Gewichtung von Widersprüchen eine größere Verzerrung erwarten müsste.

Es scheint durchaus plausibel anzunehmen, dass Erfahrene und Unerfahrene eine andere Erwartungshaltung bzgl. der Häufigkeit des Auftretens von Widersprüchen haben, denn selbst wenn bei mehreren Patienten derselben Störung diagnostiziert wird, so müssen sie nicht unbedingt in vollem Umfang eine identische Symptomkonstellation aufweisen. Die Klassifikationssysteme wie das DSM-IV (Saß et al., 2003) tragen der Variabilität der Symptome insofern Rechnung, als es bei einer Vielzahl von Störungen nicht erforderlich ist, dass alle der aufgelisteten Symptome vorliegen. Vielmehr ist häufig die Angabe zu finden,

dass eine Anzahl X der definierten Y Kriterien erfüllt sein müssen, damit die Diagnose gerechtfertigt ist. Außerdem gibt es im Falle nicht vollständig erfüllter Mindestkriterien die Möglichkeit, eine nicht näher bezeichnete Form der jeweiligen Störung zu diagnostizieren. Insofern ist es durchaus Alltag und in Übereinstimmung mit den Klassifikationssystemen, dass eine Diagnose vergeben wird, auch wenn (einzelne) Widersprüche vorliegen. Der Erfahrene erwartet daher durchaus inkonsistente Informationen und sieht in deren Auftreten bei einer isolierten Betrachtungsweise keine absolute Entwertung der Prüfhypothese. Der Unerfahrene wird womöglich eine höhere Konsistenz der Informationen unterstellen, um eine Diagnose zu rechtfertigen und überraschter sein von den inkonsistenten Angaben und ihnen daher einen höheren Wert beimessen. Möglicherweise könnte also eine Erwartungshaltung hinsichtlich der Wahrscheinlichkeit, Widersprüche zu finden, für die geringere Gewichtung der inkonsistenten Informationen angenommen werden.

Aber warum kommt es bei Erfahrenen trotz der scheinbar geringeren Gewichtung der Inkonsistenzen nicht zu einer höheren Rate an selbstbestätigenden Urteilen? Möglicherweise handelt es sich dabei um einen Effekt der Stimuluspräsentation. In den Aufgaben zur Erfassung der Bestätigungstendenz I und B ging es gerade um eine isolierte Bewertung. Die Informationen wurden einzeln präsentiert und sollten für sich genommen betrachtet werden. Dies stellt eine eher einzelheitlich, weil sequentielle Verarbeitung dar. Nach den Erkenntnissen von Jonas, Schulz-Hardt und Frey (2001) besteht bei sequentieller Präsentation bzw. Informationsverarbeitung allerdings ein größeres Verzerrungspotential als bei simultaner Präsentation. Sofern man unterstellen kann, dass die spätere Diagnosestellung wiederum unter eher ganzheitlichen Bedingungen, d.h. durch die gleichzeitige und integrative Betrachtung aller Informationen, so könnten die unterschiedlichen Verarbeitungsstile eine mögliche Erklärung dafür liefern, warum trotz einer Abwertung von inkonsistenten Daten bei sequentieller Betrachtung später kein verzerrtes Urteil resultiert. Die Überlegungen sind an dieser Stelle jedoch spekulativ und bedürfen der empirischen Prüfung in weiteren Arbeiten.

Fazit. Zusammenfassend kann festgehalten werden, dass es keine Hinweise darauf gibt, dass Berufserfahrung alleine davor schützt, Selbstbestätigungstendenzen zu unterliegen. Die meisten Informationsverarbeitungsschritte waren vielmehr unbeeinflusst von der Berufserfahrung, insbesondere die Schlussfolgerung selbst. Weiterer Forschungsbedarf besteht jedoch im Hinblick auf die Bewertung von inkonsistenten Informationen, um nachfolgend abzuklären, ob bei zunehmender Berufserfahrung ein unkritischerer Umgang hiermit entwickelt wird oder ob methodische Umstände die hier vorliegenden Nachteile der Erfahrenen vorgetäuscht haben.

6.1.4 Ergebnisinterpretation: Fachwissen

Für den Einfluss des *Fachwissens* lässt sich die globale Hypothese formulieren „*Personen mit relevantem Fachwissen weisen eine weniger verzerrte Informationsverarbeitung auf und fällen seltener ein selbstbestätigendes Urteil als Personen mit irrelevantem Fachwissen.*“ Grundsätzlich wird demnach eine protektive Wirkung des relevanten Fachwissens unterstellt. Die Darstellung erfolgt daher gruppiert nach der Erwartungskonformität im Hinblick auf die protektive Wirkung.

Erwartete protektive Wirkung. Der einzige erwartungskonforme Effekt hinsichtlich des Fachwissens liegt im Zusammenhang mit der hypothesenkonsistenten Interpretation uneindeutiger Evidenz vor, da Personen mit relevantem Fachwissen seltener die Aussagekraft uneindeutiger Evidenz überschätzten als Personen mit irrelevantem Fachwissen, indem sie sie hypothesenkonsistent interpretierten ($PH-C_c$). Die Wirkung ist zudem als kleiner bis bestenfalls mittlerer Effekt zu klassifizieren.

Keine Unterschiede zwischen Fachwissen relevant und Fachwissen irrelevant. Weder die subjektive Sicherheit ($PH-A_{2c}$), noch eines der Maße zu den asymmetrischen Evidenzregeln weist Unterschiede aufgrund des Fachwissens auf ($PH-B_{1c}$, $PH-B_{4c}$, $PH-B_{5c}$). Entsprechendes gilt für die Bestätigungstendenz B, da Personen mit relevantem und irrelevantem Fachwissen die Wichtigkeit von konsistenten gegenüber inkonsistenten Informationen in ähnlicher Weise bewerteten ($PH-E_{1c}$).

Nicht bewertbar. Erneut sind die Hypothesen im Zusammenhang mit der Diagnostizität der Merkmale der Evidenzregeln ($PH-B_{2c}$ & $PH-B_{3c}$) aufgrund einer unzureichenden Datenbasis nicht prüfbar. Die Argumentation gilt analog zu derjenigen im Zusammenhang mit der Überzeugtheit von der Prüfhypothese (s. Abschnitt 6.1.2.6).

Erwartungskonträre konfirmatorische Wirkung. Ebenfalls im Zusammenhang mit der Bestätigungstendenz B ($PH-E_{2c}$) ist allerdings zu beobachten, dass Probanden mit relevantem Fachwissen (Essstörung) konsistenten Informationen (in Relation zu inkonsistenten) einen noch größeren Einfluss auf die Bewährung der Prüfhypothese zuschreiben als Probanden der Bedingung Fachwissen *irrelevant* (Brandstiftung). Auch die Bestätigungstendenz I ist bei Personen mit relevantem Fachwissen stärker ausgeprägt, da sie mehr Widersprüche übersehen als Personen mit irrelevantem Fachwissen ($PH-D_c$).

Außerdem ist die Wahrscheinlichkeit für ein selbstbestätigendes Urteil in der Bedingung *Fachwissen irrelevant (Brandstiftung)* gegenüber der Bedingung *Fachwissen relevant (Essstörung)* sogar reduziert ($PH-A_{1c}$). Personen mit *irrelevantem Fachwissen (Brandstiftung)* fällten demnach seltener ein selbstbestätigendes Urteil als Personen mit *relevantem Fachwissen (Essstörung)*, denn es kam lediglich bei 18% der Probanden mit irrelevantem Fachwissen zu einem Schuldspruch, während 41% der Teilnehmer mit relevantem Fachwissen die Diagnose Bulimie stellten.

Fazit. Die Hypothese, dass das Fachwissen einen protektiven Einfluss auf eine verzerrte Informationsverarbeitung und ein selbstbestätigendes Urteil hat, konnte sich in der vorliegenden Arbeit nicht bewähren. Vielmehr sind bei zentralen Variablen der Informationsbewertung und der Schlussfolgerung weniger Verzerrungen bei den Personen mit irrelevantem Fachwissen zu beobachten.

Dies widerspricht den Ergebnissen von Krems und Zierer (1994) und wird als methodisches Resultat betrachtet. Wie in Abschnitt 4.1.3 argumentiert, mussten abweichend vom Vorgehen Krems und Zierers (1994) zwei unterschiedliche Fallgeschichten eingesetzt werden, um die Bedingungen mit und ohne relevantes Fachwissen zu schaffen. Der gewählte Brandstiftungsfall weist gegenüber der klinischen Essstörungsdiagnostik jedoch eine spezielle Besonderheit auf, die u.U. erklären könnte, warum hier weniger Fehlurteile resultierten: er liefert durch seine juristische Ausrichtung eine Entscheidungsregel für den Fall, dass ein uneindeutiges Resultatmuster mit sowohl konsistenten als auch inkonsistenten Informationen resultiert: *in dubio pro reo* ist der Angeklagte freizusprechen, wenn die Schuld nicht zweifelsfrei nachzuweisen ist. Dem Fall könnte aufgrund dieses Grundsatzes daher ein Schutzmechanismus gegen den *confirmation bias* innegewohnt haben, der bei der Auswahl des Studienkontexts nicht in Betrachtung gezogen worden war.

Es sind daher Replikationen mit anderen als juristischen Kontexten angezeigt, um in Anlehnung an Krems und Zierer (1994) die Rolle des Fachwissens zu beleuchten. Dabei wäre es durchaus denkbar, die Psychotherapeuten in einer Folgestudie neben einer klinischen Fragestellung eine eignungsdiagnostische Diagnostik bei der Personalauswahl durchführen zu lassen, da nicht anzunehmen ist, dass sie ihr Fachwissen im Sinne von störungsspezifischem Wissen hier einsetzen können.

6.1.5 Explorativ: Interaktion von Berufserfahrung und Fachwissen

Es ist nicht anzunehmen, dass die Berufserfahrung und das Fachwissen in bedeutsamer Weise interagieren, da im Zusammenhang mit den fünf bedingten Variablen und den 11 insgesamt untersuchten Maßen nur ein einziger signifikanter Effekt aufgetreten ist. Dieser wurde im Zusammenhang mit der Bestätigungstendenz B in Abhängigkeit von der Berufserfahrung bereits inhaltlich diskutiert.

Selbst wenn dieser bestenfalls als kleiner Effekt zu beschreibende Unterschied kein Zufallsergebnis des multiplen Testens ohne Adjustierung des Signifikanzniveaus sein sollte, so ist ein kleiner Effekt nicht hinreichend, um Hypothesen über das Zusammenwirken von Berufserfahrung und Fachwissen zu formulieren und in nachfolgenden Studien zu prüfen, zumal das Ausbleiben von Interaktionseffekten mit den Befunden von Krems und Zierer (1994) übereinstimmt.

6.2 Bewertung der Generalisierbarkeit der Ergebnisse

Um die Gültigkeit der Ergebnisse über die vorliegende Studie hinaus abschätzen zu können, werden zunächst die Teilnahmebereitschaft sowie die Repräsentativität hinsichtlich der Berufserfahrung im Hinblick auf mögliche Stichprobenverzerrungen diskutiert. Abschließend werden die interpretatorischen Grenzen aufgrund der Wahl einer Online-Studie aufgezeigt.

6.2.1 Teilnahmebereitschaft

Eine Annäherung an die Schätzung der Teilnahmequote stellt die Beobachtung dar, dass 22,4% derjenigen, die die erste Bildschirmseite bearbeiteten, die Studie vollständig durchführten. Ein Vergleich mit Teilnahmequoten in Studien, die ebenfalls u.a. Psychotherapeuten ansprachen, ist vergleichsweise schwierig. Zwar rekrutierte Parmely (2006) die lizenzierten Therapeuten ebenfalls per Email, indem sie mehrere Verteilerlisten nutzte, berichtet in ihrer Arbeit jedoch keine Abschätzung der Teilnahmebereitschaft.

Auch die Übersichtsarbeit von Weathers, Furlong und Solórzano (1993) über *mail survey research in counseling psychology* ist nur bedingt hilfreich. Dieser ist zwar zu entnehmen, dass der Median der Rücklaufquote bei den 34 in ihrer Arbeit eingeschlossenen Studien bei 63,7% lag, allerdings wurden ausschließlich solche Studien berücksichtigt,

die von Januar 1980 bis Oktober 1989 im *Journal of Counseling Psychology* publiziert wurden und ihren Probanden die Studienunterlagen per Post zukommen und wieder zu sich zurückschicken ließen, während in der vorliegenden Arbeit ein Onlineverfahren genutzt wurde. Außerdem wurde in der Mehrzahl der analysierten Studien zwei oder drei follow-up Kontakte genutzt, was sich allgemein sehr positiv auf die Rücklaufquote auswirkt (Helgeson, Voss & Terpening, 2002). Für die vorliegende Arbeit wurde davon jedoch aufgrund der aktuellen Diskussion um *spam-mails* Abstand genommen.

Die in der Tat eher geringe Teilnahmebereitschaft an der vorliegenden Studie ist aus Sicht der Verfasserin höchstwahrscheinlich zum größten Teil durch die lange Bearbeitungsdauer bedingt. Wie in Abschnitt 3.2.1 verdeutlicht, war es das erklärte Ziel, den gesamten Prozess der Hypothesentestung bzw. Diagnosestellung abzubilden, was nach den Rechercheergebnissen der Verfasserin so nicht unternommen worden war. Es ist daher plausibel, dass der hohe Zeitaufwand, der auf dem Eingangsbildschirm angekündigt wurde, und aus dem Lesen der Hintergrund- und Fallinformationen erkennbar war, eine entscheidende Rolle für die Abbrüche gespielt hat.

Hierin muss aber nicht unbedingt eine Verzerrung der zugrundeliegenden Stichprobe gesehen werden, zumindest keine, die in Zusammenhang mit dem Vortäuschen der Effekte gebracht werden kann. Denn aufgrund des Selektionseffekts, der hauptsächlich auf den Zeitaufwand attribuiert wird, ist anzunehmen, dass vorrangig besonders intrinsisch motivierte Personen teilgenommen haben. Daher ist ebenfalls anzunehmen, dass auch eine hohe Gewissenhaftigkeit bei der Bearbeitung vorliegt und die gefundenen Urteilsverzerrungen daher sehr wahrscheinlich nicht auf Flüchtighkeitsfehler zurückgehen, sondern trotz einer hohen Motivation, Gewissenhaftigkeit und eines Eigeninteresses begangen wurden – ganz wie Schulz-Hardt und Köhnken (2000) es unterstellen. Die Verteilung der Studienabbrüche über den Studienverlauf macht ebenfalls deutlich, dass es zu keinen nennenswerten Abbrüchen mehr kann, wenn ein Proband die Generierung der Evidenzregeln (AV 3) erfolgreich absolviert hatte. Personen, die sich demnach auf die zeitaufwändige, komplexe und kognitiv stark fordernde Aufgabe der Diagnostik einlassen konnten, führten diese weitestgehend bis zum Ende durch.

6.2.2 Repräsentativität hinsichtlich der Variablen Berufserfahrung

Im Folgenden wird getrennt für Erfahrene und Unerfahrene bewertet, inwieweit eine Repräsentativität als notwendige Voraussetzung für die Übertragbarkeit gegeben ist.

Erfahrene. Die Gruppe der Erfahrenen kann hinsichtlich der Geschlechtsverteilung als repräsentativ für die Gruppe der Psychotherapeuten in Deutschland betrachtet werden. In Bezug auf das Alter weicht die Stichprobe von der Gesamtgruppe jedoch insofern ab, als dass eher jüngere Personen an der Studie teilgenommen haben. Psychotherapeuten im Alter von 35 bis 44 Jahren sind in der Stichprobe überrepräsentiert, der Anteil der über 60 Jährigen ist geringer als in der Population. Insofern sind die Personen mit der größten Berufserfahrung weniger repräsentiert. Letzteres hätte jedoch nur dann eine substanzuell einschränkende Konsequenz bzgl. der Gültigkeit der Ergebnisse, wenn es Grund zur Annahme gäbe, dass die Berufserfahrung v.a. dann einen Einfluss ausübt, wenn sie mindestens 25 Jahre beträgt. Empirische Hinweise gibt es hierauf keine, vielmehr wurden die Befunde, die zur Berufserfahrung vorliegen im Rahmen der Ableitung der Fragestellung (s. Abschnitt 3) gerade insoweit verdichtet, als keine Einflüsse der Berufserfahrung erwartet werden. So dass nicht davon ausgegangen werden muss, ein gänzlich anderes Resultatmuster zu erhalten, wenn nicht nur 11% der Probanden über 60 Jahre alt gewesen wären, sondern die in der Population vorherrschenden 23% über 60 Jähriger untersucht worden wären.

Außerdem ist festzuhalten, dass selbst im Falle der konservativen Schätzung der Berufserfahrung der hiesigen Probanden durch eine Approbation vor im Mittel 8 Jahren, die Berufserfahrung und damit die Repräsentativität höher ist als bspw. bei Haverkamp (1993), deren Probanden im Mittel drei Jahre Berufserfahrung aufwiesen, da sie in einschlägigen Trainingsprogrammen rekrutiert wurden. Auch Martins (2000) Probanden, die im Mittel ca. 550 Stunden an psychotherapeutischer Arbeit berichteten, wiesen in diesem Zusammenhang eine deutlich geringere Berufserfahrung und damit Repräsentativität auf. Und während bei Parmley (2006) die Hälfte der Probanden maximal sechs Jahre Berufserfahrung aufwies, so sind es in der vorliegenden Stichprobe nur 31%, die vor maximal 6 Jahre ihre Approbation erhalten haben, wo die mindestens dreijährige Therapieausbildung nicht einmal mit einbezogen ist.

Die Repräsentativität der Stichprobe ist im Hinblick auf die Population als hinreichend zu bewerten und im Vergleich mit anderen empirischen Arbeiten stellt sie sich hoch dar.

Unerfahrene. Dem Umstand, dass in der vorliegenden Studierendenstichprobe mit 88% gegenüber 77% ein höherer Frauenanteil gefunden wurde, wird kein verzerrendes Potential zugesprochen, da es sich einerseits nur um eine kleine Erhöhung des an sich bereits ausgeprägten Frauenüberschusses im Fach Psychologie handelt und andererseits keine Hinweise darauf vorliegen, dass insbesondere Frauen eher zu einem *confirmation*

bias neigen, was zu künstlichen Effekten in der vorliegenden Arbeit führen könnte. Hinweise gegen eine Repräsentativität hinsichtlich des Alters konnten nicht gefunden werden. Die Übertragbarkeit der Befunde für Unerfahrene auf die Gruppe der Psychologiestudierenden in Deutschland ist daher gegeben.

6.2.3 Umsetzung als Online-Studie

Analogstudien wie die vorliegende Arbeit gehen zwar stets mit einer eingeschränkten externen Validität bzw. Generalisierbarkeit einher, haben jedoch gerade zu Beginn der Untersuchung von neuen Konzepten und Paradigmen ihre Berechtigung. Es wird also anerkannt, dass nicht automatisch unterstellt werden darf, dass sich ein vergleichbares Ergebnismuster zeigen wird, wenn ein realer diagnostischer Prozess mit einem echten oder zumindest einem *mock patient* (vgl. Strohmer & Shivy, 1994) untersucht würde. Für den konkreten Diagnosefall der vorliegenden Arbeit könnte man sogar annehmen, dass keine vergleichbaren Effekte gefunden würden, weil man einer Patientin ihr substanzielles Untergewicht im persönlichen Kontakt sofort ansehen würde und dieses die Prüfhypothese Bulimie unmittelbar unwahrscheinlich erscheinen lassen würde. Allerdings war es nicht das Ziel der Arbeit, den *confirmation bias* bei der Diagnosestellung einer Bulimie zu untersuchen. Der verwendete Kontext ist als exemplarisch zu betrachten.

Dass die externe Validität daher durch die Umsetzung als webbasierte Analogstudie eingeschränkt ist, bedeutet keine grundsätzliche Entwertung der vorliegenden Ergebnisse. Sie vermögen es aufgrund der sichergestellten internen Validität, einen tatsächlichen Forschungsbedarf hinsichtlich des Auftretens des *confirmation bias* in der direkten sozialen Interaktion anzuzeigen. Hinweise hierzu werden in Abschnitt 6.4 gegeben.

6.3 Kritische Würdigung

Im Folgenden werden sowohl die Grenzen als auch die Stärken der Arbeit zusammengefasst.

Durchführung als Analogstudie. Die Generalisierbarkeit der Ergebnisse ist aufgrund der Durchführung als webbasierte Analogstudie wie bereits diskutiert als eingeschränkt zu bewerten und stellt damit eine relevante Grenze der Bedeutsamkeit der Ergebnisse dar, die es in Folgestudien zu analysieren gilt.

Rolle des Fachwissens ungeklärt. Unbefriedigend ist desweiteren der Umstand, dass durch die Wahl eines juristischen Fallbeispiels in der Bedingung Fachwissen irrelevant unbeabsichtigt vermutlich ein solcher Kontext gewählt wurde, der einer Selbstbestätigung entgegenwirkt, da durch die Implementierung des Grundsatzes *in dubio pro reo* eine Entscheidungsregel für uneindeutige Situation mitgeliefert wird, die im klinisch-psychologischen Kontext so nicht vorliegt.

Ganzheitlichkeit des Hypothesentestungsprozesses. Als eine wesentliche Stärke der vorliegenden Arbeit ist anzuführen, dass der gesamte Ablauf des Hypothesentestungsprozess betrachtet wurde, indem Maße für das Testungsverhalten auf allen Ebenen des Modells des (sozialen) Hypothesentestens nach Trope und Liberman (1996) bzw. Hager und Weißmann (1991) sowie dem diagnostischen Prozess (Jäger, 1982) erhoben wurden, wohingegen bislang eher entweder die Informationssuche oder die Informationsbewertung einzeln untersucht wurden. So wurde die positive Teststrategie im Einklang mit der Forderung von Haverkamp (1993) losgelöst von der Frageformulierung betrachtet und dabei zum ersten Mal die Generierung der Evidenzregeln in Betracht gezogen.

Berücksichtigung der subjektiven Perspektive. Außerdem ist durch die Berücksichtigung der subjektiven Überzeugtheit von der Prüfhypothese eine größere Realitätsnähe geschaffen worden als in der Situation der reinen *experimenter provided hypothesis* (Haverkamp, 1993). Denn so ist zwar die Übermittlung von Verdachtsdiagnosen seitens eines überweisenden Hausarztes durchaus realistisch, aber wie die vorliegenden Befunde zeigen, ist es ungerechtfertigt davon auszugehen, dass auf ein und dieselbe Prüfhypotheseninduktion dasselbe Informationsverarbeitungsmuster folgt.

Repräsentativität der Stichprobe. Eine weitere Stärke stellt die gewonnene Stichprobe dar, da sie als groß und repräsentativ zu bewerten ist. So können regionale Effekte aufgrund der Teilnahme von Personen aus 14 Bundesländern weitgehend ausgeschlossen werden. Außerdem kann in Bezug auf die Teilstichprobe der Erfahrenen begründetermaßen von einer Gruppe berufserfahrener Personen gesprochen werden. Ausgebliebene Unterschiede zwischen Erfahrenen und Unerfahrenen sind demnach nicht auf eine zu große Ähnlichkeit zurückzuführen, wie dies bspw. in der eigenen Vorarbeit (Dreger, 2006) aber auch bei Martin (2000) der Fall gewesen sein könnte.

6.4 Bedeutung der Ergebnisse für Praxis und Forschung

Es können sowohl Implikationen für die klinische Praxis als auch die Forschung im Zusammenhang mit dem *confirmation bias* abgeleitet werden.

6.4.1 Implikationen für die klinische Praxis

Für die klinische Praxis sind v.a. die Befunde hinsichtlich der Berufserfahrenen relevant, die die Prüfhypothese der Bulimie testeten. Insgesamt 41% der Personen, dieser Bedingung mit relevantem Fachwissen bestätigten die Prüfhypothese später trotz des Vorliegens konträrer Befunde und der Möglichkeit, die diagnostischen Kriterien jederzeit einzusehen. Und dies waren keineswegs nur die Unerfahrenen, vielmehr stellten einerseits 40% der Erfahrenen und andererseits 43% der Unerfahrenen die falsche Diagnose. Ein erfahrener Psychotherapeut ist damit zumindest bei der Beurteilung schriftlichen Materials über einen Patienten nicht automatisch gegen Selbstbestätigungstendenzen immun. Das Phänomen des *confirmation bias* tritt daher auch im Bereich klinisch-psychologischer Diagnostik durch hinreichend erfahrene Therapeuten auf. Insofern muss für einen nicht unbedeutenden Teil der Probanden die Diagnose *confirmation bias* gestellt werden (vgl. Mendel et al., 2011).

Da die Befunde hinsichtlich der Überzeugtheit von der Prüfhypothese zeigen, dass die anfängliche Einschätzung einer möglichen Hypothese Konsequenzen für den gesamten Verlauf des Testungsprozesses hat, unterstreicht dies die Wichtigkeit, gerade zu Beginn des diagnostischen Prozesses eine offene Grundhaltung einzunehmen. Dies ist am besten dadurch umzusetzen, dass ganz bewusst alternative Hypothesen generiert werden. Der hierin zu investierende Aufwand sollte umso höher sein, je überzeugter man anfänglich von einer bestimmten Hypothese ist. Dies setzt eine Aufklärung darüber voraus, eine hohe Überzeugtheit als Warnzeichen zu verstehen, das es zu beachten gilt.

Auch die Ergebnisse hinsichtlich der Evidenzregeln sprechen dafür, gerade dem Beginn der Hypothesentestung mehr Aufmerksamkeit zu schenken. Aus der Beobachtung, dass über 25% der im Essstörungskontext generierten Evidenzregeln nicht instruktionskonform waren, kann abgeleitet werden, dass es sich um eine vergleichsweise unvertraute Aufgabe gehandelt haben muss. Im Einklang mit dieser Interpretation steht auch die hohe Abbruchquote im Zusammenhang mit dieser Aufgabe: haben noch 47% der ursprünglich den Studienlink anklickenden Probanden die Instruktionen zu den Evidenzregeln gelesen, so haben nur 25% der Ursprungsteilnehmer die Evidenzregeln

tatsächlich auch formuliert. Es kam an dieser Stelle daher fast zu einer Halbierung der Teilnehmerzahl. Beide Ergebnisse lassen nun vermuten, dass der hohe Aufwand bzw. eine Unvertrautheit eine Rolle gespielt haben. Woraus wiederum ableitbar wäre, dass in der Praxis üblicherweise kein vergleichbarer Schritt unternommen wird. In Übereinstimmung mit dem Trainingsziel von vom Schemm (2008b) ist daher zu fordern, mit der Vermeidung des *confirmation bias* bereits im Zusammenhang mit dem Formulieren von Evidenzregeln zu beginnen. Dies sollte einen unverzichtbaren Teil des diagnostischen Prozesses darstellen und ist vergleichbar mit der Forderung, bei der Personalauswahl vor dem Führen von Einstellungsinterviews ein Anforderungsprofil zu erstellen, um einen Anforderungsbezug der zu stellenden Fragen zu sichern, da dieser neben der Strukturiertheit des Interviews eine der beiden zentralen Variable für die Validität der Entscheidung darstellt (Schuler, 2000). Zugleich bietet das Aufstellen der Evidenzregeln eine Möglichkeit, die zuvor generierten Alternativhypothesen aktiv in den Hypothesentestungsprozess einzubeziehen und nicht als leere Hüllen mit Alibifunktion stehen zu lassen.

Eine weitere Implikation ergibt sich aus den Hinweisen auf einen möglicherweise unkritischeren Umgang der Erfahrenen mit widersprüchlichen Informationen. Selbst wenn ungeklärt bleibt, ob ein methodisches Phänomen oder ein tatsächlicher Nachteil zu Lasten der Erfahrenen vorlag, sollte im Sinne eines konservativen, den *confirmation bias* vermeidenden Vorgehens, Vorsorge getroffen werden. Eine zentrale Rolle könnten an dieser Stelle erneut die Alternativhypothesen spielen. Denn wenn die widersprüchlichen Befunde selbst es nicht direkt schaffen, die Überzeugtheit von der Prüfhypothese zu senken, was sich in deren Abwertung widerspiegelt, so wäre es aber vielleicht indirekt über eine Stärkung der Alternativhypothese möglich. Wenn die Inkonsistenzen nicht einfach als tolerable Gegebenheiten interpretiert werden, sondern zusätzlich ihre Bedeutung für die Alternativhypothese bewertet würde und an dieser Stelle eine Bekräftigung erfolgt, könnte sich dies wie angedeutet protektiv auswirken.

Aus diesen Ausführungen wird deutlich, dass der Schlüssel zur Vermeidung des *confirmation bias* v.a. in der konsequenten Beachtung von Alternativhypothesen über den gesamten Prozess der Hypothesenprüfung hinweg liegt.

6.4.2 Implikationen für die Forschung

Für die Forschung stellen sich nun mindestens drei weitere Fragen. Es soll in der Folge daher erstens Stellung hinsichtlich der Bewährung des Modells des konfirmatorischen Hypothesentestens (Schulz-Hardt & Köhnken, 2000) in der klinisch-psychologischen

Diagnostik bezogen werden. Zweitens wird erläutert, welche weiteren Einflussgrößen forschungsrelevant sind, bevor drittens auf *consider the opposite* Paradigmen zur Vermeidung des *confirmation bias* eingegangen wird.

Bewährung des Modells des konfirmatorischen Hypothesentestens in der klinisch-psychologischen Diagnostik. Was ist nun hinsichtlich der empirischen Bewährung bzw. weiterem Forschungsbedarf hinsichtlich der einzelnen Verzerrungsmechanismen des Modells des konfirmatorischen Hypothesentestens (Schulz-Hardt & Köhnken, 2000) im klinisch-psychologischen Bereich zu resümieren?

In Bezug die *Überschätzung a priori Wahrscheinlichkeit* ist anzuführen, dass der im Wesentlichen durch die sozialpsychologische Arbeit von Hodgins und Zuckerman (1993) belegte Verzerrungsmechanismus noch nicht im klinischen-psychologischen Bereich untersucht wurde, was in der vorliegenden Arbeit mit dem Vorhandensein der Prävalenzraten aus epidemiologischen Studien begründet wurde. Dies unterstellt jedoch, dass in der Praxis tatsächlich diese Zahlen zugrundegelegt werden, die an den unausgelesenen Stichproben ermittelt wurden und nicht vielmehr institutionsspezifische Basisraten relevant sind, die in der Form vorliegen „*Jeder sechste Patient, der mir vorgestellt wird, ist bulimisch*“, und von der Selektivität der Stichprobe abhängen. Diese subjektiven Prävalenzraten werden vermutlich immer über den objektiven liegen und könnten damit grundsätzlich eine Überschätzung beinhalten. Die *selektiven Gedächtnisprozesse* sind umfassend belegt, sowohl im Rahmen sozialpsychologischer Arbeiten (Snyder & Cantor, 1979; Stangor & McMillan, 1992) als auch für den klinischen Bereich (Arkes & Harkness, 1980; Strohmmer et al. 1990), sogar in einer Arbeit mit einem *mock client* in der Praxis der Therapeuten (Strohmmer & Shivy, 1994).

Hinsichtlich der *hypothesenkonsistenten Interpretation uneindeutiger Evidenz* ist eine inhaltliche Überschneidung mit den Bestätigungstendenzen B und I nach Hager und Weißmann (1991) festzustellen, wobei letztere differenzierter sind und in der vorliegenden Arbeit besser gestützt werden konnten. Eine Integration der Konzepte bzw. entsprechender Operationalisierungen in Folgestudien erscheint daher sinnvoll. Für den *affirmation bias* ist v.a. im klinischen Bereich Forschungsbedarf festzustellen, weil es sich um ein robustes Phänomen handelt, dass entsprechend der Ausführungen von Turk und Salovey (1985), insbesondere deshalb relevant ist, weil die Kliniker aufgrund ihres Arbeitskontextes Pathologisches erwarten, was das Erkennen der Anwesenheit der Symptome wahrscheinlicher macht als das Wahrnehmen ihrer Abwesenheit. Der konfirmatorische Einfluss eines *positiven bzw. pseudodiagnostischen Vorgehens* muss

als gesichert beschrieben werden, dennoch ist die Fokussierung auf den Prozess der Frageformulierung nachfolgend endgültig aufzugeben. Außerdem liegen kaum Arbeiten über die Bildung von Hypothesen und Alternativen vor, die das pseudodiagnostische Arbeiten weiter beleuchten könnten. Der motivationale Aspekt der *asymmetrischen Schwellenwahl* ist im klinisch-psychologischen Bereich bislang nicht untersucht.

Diejenigen Verzerrungsmechanismen, die untersucht wurden, können daher alle als relevant bezeichnet werden, wobei nicht außer Acht gelassen werden darf, dass in der Mehrzahl kleine Effektstärken vorgefunden wurden. Es muss also davon ausgegangen werden, dass der Kreis der bedingenden Variablen für eine Selbstbestätigung noch größer gezogen werden muss.

Weitere Einflussgrößen. In Zusammenhang mit Abschnitt 2.4 wurden bereits allgemeine Ergänzungsmöglichkeiten in Bezug auf das Modell des konfirmatorischen Hypothesentestens vorgestellt. Im Zusammenhang mit den vorliegenden Ergebnissen ergeben sich insbesondere Implikationen in Bezug auf den intern-repräsentierten Anfangsverdacht, der als Überzeugtheit von der Prüfhypothese operationalisiert wurde und sich als relevant erwiesen hat. Nachfolgend aufzuklären ist in diesem Zusammenhang die Frage, warum Menschen mit einer so unterschiedlichen Überzeugtheit auf identisches Material reagieren. Mit dem Befund von Jochumsen (2009) liegt ein Hinweis auf den Einfluss des Persönlichkeitsmerkmals *need for cognition* vor, was weiter zu verfolgen ist.

Rassin (2008) schlägt außerdem vor, interindividuelle Unterschiede im Hinblick auf die Anfälligkeit für den *confirmation bias* zu untersuchen und stellt ein 10 Item umfassendes Selbstbeurteilungsinstrument vor, welches höhere Werte bei solchen Personen anzeigt, die in verschiedenen Urteilsbildungsaufgaben eher zu konfirmatorischen denn diskonfirmatorischen Antworten neigen. Inwieweit diese 10 Items aber geeignet sind, den Facettenreichtum des Phänomens abzubilden, muss dahin gestellt bleiben. Dennoch ist es erstaunlich, dass die Forschungsbemühungen um die interindividuellen Unterschiede in der Vergangenheit so gering waren. Außerdem sind der Verfasserin seit Rassins Publikation keine vergleichbaren Arbeiten bekannt.

Die Persönlichkeit des Hypothesentesters in den Mittelpunkt nachfolgender Forschungsüberlegungen zu stellen, ist nach Ansicht der Verfasserin daher unbedingt erforderlich. Außerdem dürfen auch situative Aspekte wie die kontextbezogene emotionale Involviertheit bzgl. der Prüfhypothese nicht außer Acht gelassen werden (vom Schemm, 2008a).

In dubio pro reo als Ansatzpunkt für die Vermeidung des confirmation bias.

Um die Annahme zu überprüfen, dass der Grundsatz *in dubio pro reo* im juristischen Kontext bereits eine gewisse Intervention gegen den *confirmation bias* darstellen könnte, wäre folgendes Design denkbar: die vorliegende Fallbeschreibung aus dem Brandstiftungsfall könnte Probanden mit der Bitte vorgelegt werden, sich in die Rolle des Schöffen hineinzusetzen und zu einer Einschätzung hinsichtlich der Schuld bzw. Unschuld zu kommen. Alle Probanden erhalten zudem spezielle Instruktionen zu ihrer Aufgabe als Schöffen. Eine Kontrollgruppe könnte man auf ihre Pflicht zur Objektivität und Unvoreingenommenheit hinweisen – ganz im Sinne einer *be unbiased*-Bedingung aus der *consider the opposite* Forschung (z.B. Lord et al., 1984), die Experimentalgruppe erhält entsprechende Erläuterungen hinsichtlich des *in dubio pro reo*. Weitere Experimentalbedingungen könnten auch spezifische Alternativerklärungen für die Entstehung des Feuers enthalten, z.B. einen technischen Defekt. Erkenntnisse hinsichtlich der Wirksamkeit von *consider the opposite* Elementen wären sehr hilfreich, denn selbstverständlich darf die Forschung nicht dort aufhören, wo Unzulänglichkeiten festgestellt wurden. Denn um das Wissen um den *confirmation bias* zum Wohle der Einzelnen bzw. der Gemeinschaft einsetzen zu können, bedarf es des ergänzenden Wissens um entsprechende Interventionsmaßnahmen.

Ihren Abschluss soll die Arbeit nun dort finden, wo sie begonnen hat. Wenn eingangs argumentiert wurde, dass *Irren menschlich ist*, so gilt dies auch im Hinblick darauf, dass ein Hypothesentester mit seiner anfänglichen Überzeugtheit in eine zu prüfende Annahme irren kann. Nicht darin liegt das Problem, sondern in der Immunisierung der Überzeugtheit. Es ist daher darauf zu achten, dass man Fehler erkennen und falsche Einschätzungen revidieren kann. Ist dies nicht der Fall und behält man Fehleinschätzungen bei, ist dies mit den Worten von Cicero nicht länger menschlich, sondern vielmehr teuflisch, lautet sein Zitat vollständig doch:

Errare humanum est, sed in errare perseverare diabolicum.

Literatur

- Adobe Systems Incorporated. (2012). *Adobe Flash Player PC Penetration*. Retrieved February 09, 2012 from http://www.adobe.com/products/player_census/flashplayer/PC.html
- Amelang, M., & Zielinski, W. (2002). *Psychologische Diagnostik und Intervention*. Berlin: Springer.
- Amelang, M., Bartussek, K., Stemmler, G., & Hagemann, D. (2006). *Differentielle Psychologie und Persönlichkeitsforschung*. Stuttgart: Kohlhammer.
- Ammann, A. (1999). *Entstehung realitätsfremder Überzeugungen im Kontext des sexuellen Missbrauchs von Kindern*. Unveröffentlichte Diplomarbeit, Christian-Albrechts-Universität, Kiel.
- Amnesty International. (2007). *Richey case: reaction to news of release*. Retrieved February 09, 2012 from http://www.amnesty.org.uk/news_details.asp?NewsID=17579
- Arkes, H. R. (1981). Impediments to accurate clinical judgement and possible ways to minimize their impact. *Journal of Consulting and Clinical Psychology*, 49(3), 323–330.
- Arkes, H. R., & Harkness, A. R. (1980). Effects of making a diagnosis on subsequent recognition of symptoms. *Journal of Experimental Psychology: Human Learning and Memory*, 6(5), 568–575.
- Asch, S. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology*, 41, 258–290.
- Ask, K., & Granhag, P. (2005). Motivational sources of confirmation bias in criminal investigations: The need for cognitive closure. *Journal of Investigative Psychology and Offender Profiling*, 2(1), 43–63.
- Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2008). *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*. Berlin: Springer.
- Baer, A. (2002). *Die Entstehung von Falschbeschuldigungen im Kontext sexuellen Kindesmissbrauchs: Eine empirische Untersuchung mit Eltern 3-6 jähriger Kinder*. Unveröffentlichte Diplomarbeit, Christian-Albrechts-Universität, Kiel.
- Barnikol, K. (2004). *Suggestivfragen in Interviews: Auswirkungen von Erwartungen, Vorkenntnissen und Ängstlichkeit auf die Häufigkeit und Intensität von Suggestivfragen*. Unveröffentlichte Diplomarbeit, Christian-Albrechts-Universität zu Kiel.
- Bassok, M., & Trope, Y. (1984). People's strategies for testing hypotheses about another's personality: Confirmatory or diagnostic? *Social Cognition*, 2, 199–216.
- BDP. (1999). *Ethische Richtlinien der Deutschen Gesellschaft für Psychologie e.V. und des Berufsverbands Deutscher Psychologinnen und Psychologen e.V.* Zugriff am 05.03.2012 http://www.bdp-verband.org/bdp/verband/clips/BDP_Ethische_Richtlinien_2005.pdf

- Bless, H., Wänke, M., Bohner, G., Fellhauer, R., & Schwarz, N. (1994). Need for Cognition: Eine Skala zur Erfassung von Engagement und Freude bei Denkaufgaben. *Zeitschrift für Sozialpsychologie*, 25(2), 147–154.
- Bördlein, C. (2000). Die Bestätigungstendenz - Warum wir (subjektiv) immer Recht behalten. *Skeptiker*, 13(3), 132–138.
- Borgida, E., & DeBono, K. G. (1989). Social hypothesis testing and the role of expertise. *Personality and Social Psychology Bulletin*, 15(2), 212–221.
- Borkenau, P., & Ostendorf, F. (2008). *NEO-Fünf Faktoren-Inventar nach Costa und McCrae*. Göttingen: Hogrefe.
- Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler*. Berlin: Springer.
- Burwitz, U. (2005). *Auswirkungen von Verdacht und interpersonellem Vertrauen auf das verbale Verhalten von Interviewern in Interviews mit Kindern im Vorschulalter*: Unveröffentlichte Diplomarbeit, Christian-Albrechts-Universität, Kiel.
- Carstens, T. (2001). *Falschanklagen in Abhängigkeit von der Salienz der Konsequenzen*: Unveröffentlichte Diplomarbeit, Christian-Albrechts-Universität, Kiel.
- Casscells, W., Schoenberger, A., & Graboys, T. B. (1978). Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine*, 299(19), 999–1001.
- Caverni, J. P., Fabre, J. M., & Gonzalez, M. (1990). *Cognitive Biases - Advances in Psychology*. Oxford: North-Holland.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Lawrence Erlbaum.
- Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgement under certainty. *Cognition*, 58, 1–73.
- Cramb, A. (2007, December 20). Kenny Richey to walk free after plea deal. *Telegraph*. Retrieved February 09, 2012 from <http://truthinjustice.org/kenny-richey-free.htm>
- Crown, S. (1975). "On being sane in insane places": A comment from England. *Journal of Abnormal Psychology*, 84(5), 453–455.
- Dailey, C. A. (1952). The effects of premature conclusion upon the acquisition of understanding a person. *Journal of Psychology: Interdisciplinary and Applied*, 33, 133–152.
- Dalbert, C. (1996). *AUS. Autoritarismus-Skala*. Retrieved February 09, 2012 from <http://www.erzwiss.uni-halle.de/gliederung/paed/ppsych/sdaus.pdf>
- Darley, J., & Gross, P. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, 44(1), 20–33.
- Davis, D. A. (1976). On being detectably sane in insane places: Base rates and psychodiagnosis. *Journal of Abnormal Psychology*, 85(4), 416–422.
- Davison, G., & Neale, J. (2002). *Klinische Psychologie* (6., vollständig überarbeitete und aktualisierte Auflage). Weinheim: Verlagsgruppe Beltz.

- Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth*. New York: Free Press.
- Devine, P., Hirt, E., & Gehrke, E. (1990). Diagnostic and confirmation strategies in trait hypothesis testing. *Journal of Personality and Social Psychology*, 58(6), 952–963.
- DGPs. (2007). *Richtlinien zur Manuskriptgestaltung* (3., überarbeitete und erweiterte Auflage). Göttingen: Hogrefe.
- Doherty, M., Mynatt, C., Tweney, R., & Schiavo, M. (1979). Pseudodiagnosticity. *Acta Psychologica*, 43(2), 111–121.
- Dreger, B. (2006). *Diagnostiker in der Falle? Confirming the Confirmation Bias*. Unveröffentlichte Diplomarbeit, Christian-Albrechts-Universität, Kiel.
- Dreger, T. (2008). SciPoll_FX. [Computer Software].
- Edwards, W. (1983). Human cognitive capabilities, representativeness, and ground rules for research. In P. C. Humphreys, O. Svenson, & A. Vari (Eds.), *Advances in psychology: Analyzing and aiding decision processes* (pp. 507–513).
- Effler, M. (1986). Hypothesenbestätigende Strategien mit Hilfe von Suggestivfragen: Fakt oder Artefakt? *Psychologische Beiträge*, 28(3-4), 441–456.
- Evet, S., Devine, P., Hirt, E., & Price, J. (1994). The role of hypothesis and the evidence in the trait hypothesis testing process. *Journal of Experimental Social Psychology*, 30(5), 456–481.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191.
- Fichter, M., & Quadflieg, N. (1999). *Strukturiertes Inventar für Anorektische und Bulimische Essstörungen nach DSM-IV und ICD-10*.
- Fichter, M. & Warschburger, P. (2002). Essstörungen. In F. Petermann (Hrsg.) *Lehrbuch der klinischen Kinderpsychologie und -psychotherapie* (S. 561-571). Göttingen: Hogrefe.
- Fiedler, K. (1983). Beruhen Bestätigungsfehler nur auf einem Bestätigungsfehler? Eine Replik auf Gadenne 1982. *Psychologische Beiträge*, 25, 280–286.
- Fiedler, K. (2004). Illusory correlation. In R. F. Pohl (Ed.), *Cognitive illusions. A handbook on fallacies and biases in thinking, judgement and memory* (pp. 97–114). Hove: Psychology Press.
- Fiedler, K., Hemmeter, U., & Hofmann, C. (1984). On the origin of illusory correlations. *European Journal of Social Psychology*, 14(2), 191–201.
- Gadenne, V. (1982). Confirmation bias and rationality of cognitive processes. *Psychologische Beiträge*, 24(1), 11–25.
- Gadenne, V., & Oswald, M. (1986). Entstehung und Veränderung von Bestätigungstendenzen beim Testen von Hypothesen. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 33(3), 360–374.

- Ganzach, Y. (2000). The weighing of pathological and non-pathological information in clinical judgement. *Acta Psychologica*, *104*(1), 87–101.
- Garb, H. N. (1989). Clinical judgement, clinical training, and professional experience. *Psychological Bulletin*, *105*(3), 387–396.
- Garb, H. N., & Lutz, C. (2001). Cognitive complexity and the validity of clinicians' judgements. *Assessment*, *8*(1), 111–115.
- Gigerenzer, G., & Engel, C. (2006). *Heuristics and the law*. Cambridge, MA: MIT Press.
- Gila, A., Castro, J., Toro, J., & Salamero, M. (1998). Subjective body-image dimensions in normal and anorexic adolescents. *British Journal of Medical Psychology*, *71*(2), 175–184.
- Griggs, R. A., & Cox, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, *73*, 407–420.
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta analysis. *Psychological Assessment*, *12*, 19–30.
- Hagemann, J. (2008). *Hauptschule oder lieber nicht? Mechanismen der Urteilsverzerrung im Bereich der Schulleistungsdiagnostik und der Zusammenhang mit dem „need for cognitive closure“*. Unveröffentlichte Diplomarbeit, Christian-Albrechts-Universität, Kiel.
- Hager, W., & Weißmann, S. (1991). *Bestätigungstendenz in der Urteilsbildung*. Göttingen: Hogrefe.
- Halder-Sinn, P. (1993). Fehlerhafte Urteilsheuristiken in Sachverständigengutachten. *Monatsschrift für Kriminologie und Strafrechtsreform*, *76*(1), 44–49.
- Haverkamp, B. E. (1993). Confirmatory bias in hypothesis testing for client-identified and counselor self-generated hypotheses. *Journal of Counseling Psychology*, *40*, 303–315.
- Haverkamp, B. E. (1994). Cognitive bias in the assessment phase of the counseling process. *Journal of Employment Counseling*, *31*(4), 155–167.
- Hayden, D. C. (1987). Counselor and client responses to hypothesis-testing strategies. *Journal of Counseling Psychology*, *34*(2), 149–156.
- Helgeson, J. G., Voss, K. E., & Terpening, W. D. (2002). Determinants of mail-survey response: Survey design factors and respondent factory. *Psychology & Marketing*, *19*(3), 303–328.
- Herbert, D., Nelson, R., & Herbert, J. (1988). Effects of psychodiagnostic labels, depression severity, and instructions on assessment. *Professional psychology: research and practice*, *19*(5), 496–502.
- Hill, C., Memon, A., & McGeorge, P. (2008). The role of confirmation bias in suspect interviews: A systematic evaluation. *Legal and Criminological Psychology*, *13*(2), 357–371.

- Hirt, E. R., & Markman, K. D. (1995). Multiple explanation: A consider-an-alternative strategy for debiasing judgements. *Journal of Personality and Social Psychology*, 69(6), 1069–1086.
- Hodgins, H., & Zuckerman, M. (1993). Beyond selecting information: Biases in spontaneous questions and resultant conclusions. *Journal of Experimental Social Psychology*, 29(5), 387–407.
- Hogarth, R. (1981). Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. *Psychological Bulletin*, 90, 197–217.
- Inbau, F., Reid, J., Buckley, J., & Jayne, B. (2001). *Criminal interrogation and confessions*. Gaithersburg, MD: Aspen.
- Jacobi, C., Thiel, A., & Paul, T. (2002). *Kognitive Verhaltenstherapie bei Anorexia nervosa und Bulimia nervosa*. Weinheim: Verlagsgruppe Beltz.
- Jäger, R. S. (1982). Diagnostische Urteilsbildung. In K. J. Groffmann & L. Michel (Hrsg.), *Serie II: Psychologische Diagnostik: Bd. 1: Grundlagen psychologischer Diagnostik. Enzyklopädie der Psychologie* (S. 295–375). Göttingen: Hogrefe.
- Jochumsen, N. (2009). *Person, Situation und Hypothese: Welche dispositionellen und situativen Merkmale beeinflussen die Ausbildung eines intern-repräsentierten Anfangsverdachts als Ausgangspunkt für das konfirmatorische Hypothesentesten?* Unveröffentlichte Diplomarbeit, Christian-Albrechts-Universität, Kiel.
- Jonas, E., Schultz-Hardt, S., & Frey, D. (2001). Confirmation bias in information seeking under simultaneous vs sequential information presentation. *Journal of Experimental Psychology*, 48(3), 239–247.
- Kahnemann, D., & Tversky, A. (1972). Subjective probability: a judgment of representativeness. *Cognitive Psychology*, 3, 430–454.
- Kassin, S., Goldstein, C., & Savitsky, K. (2003). Behavioral confirmation in the interrogation room: On the dangers of presuming guilt. *Law and Human Behavior*, 27(2), 187–203.
- Klauer, K. C. (1988). Consider-the-opposite-Strategie für Korrektur und Analyse von Halo-Fehlern. *Zeitschrift für experimentelle und angewandte Psychologie*, 35(4), 597–627.
- Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, 94(2), 211–228.
- Konitzky, S. (2009). *Evaluation des Trainingsprogrammes MIA*. Unveröffentlichte Diplomarbeit, Christian-Albrechts-Universität, Kiel.
- Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, 6(2), 107–118.
- Krems, J., & Zierer, C. (1994). Sind Experten gegen kognitive Täuschungen gefeit? Zur Abhängigkeit des confirmation bias von Fachwissen. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 41(1), 98–115.
- Kruglanski, A., & Mayselless, O. (1988). Contextual effects in hypothesis testing: The role of competing alternatives and epistemic motivations. *Social Cognition*, 6(1), 1–20.

- Kurpius, D. J., Benjamin, D., & Morran, D. K. (1985). Effects of teaching a cognitive strategy on counselor trainee internal dialogue and clinical hypothesis formulation. *Journal of Counseling Psychology, 32*, 263–271.
- Langer, E. J., & Abelson, R. P. (1974). A patient by any other name...: Clinician group difference in labeling bias. *Journal of Consulting and Clinical Psychology, 42*(1), 4–9.
- Leary, M. R., & Miller, R. S. (1986). *Social psychology and dysfunctional behavior*. New York: Springer.
- Leibetseder, M., Laireiter, A.-R., Riepler, A., & Köller, T. (2001). E-Skala: Fragebogen zur Erfassung von Empathie – Beschreibung und psychometrische Eigenschaften. *Zeitschrift für Differentielle und Diagnostische Psychologie, 22*(1), 70–85.
- Lilienfeld, S. O., Spitzer, R. L., & Miller, M. B. (2005). Rosenhan revisited - the scientific credibility of Lauren Slater's pseudopatient diagnosis study. *Journal of Nervous and Mental Disease, 193*(11), 745–746.
- Limbrecht, K. (2008). *Konfirmatorisches Hypothesentesten. Systematische Urteilsverzerrung in Abhängigkeit von der individuellen Ausprägung des Persönlichkeitskonstruktes "Need for Cognition"*. Unveröffentlichte Diplomarbeit, Christian-Albrechts-Universität, Kiel.
- Loftus, E., & Palmer, J. (1974). Reconstruction of automobile destruction: an example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior, 13*, 585–589.
- Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgement. *Journal of Personality and Social Psychology, 47*, 1231–1243.
- Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37*(11), 2098–2109.
- Martin, J. M. (2000). *Confirmation bias in the therapy session: the effects of expertise, external validity, instruction set, confidence and diagnostic accuracy*. Unveröffentlichte Dissertation, University of Memphis, Memphis.
- Maxwell, A. E. (1977). Coefficients of agreement between observers and their interpretation. *British Journal of Psychiatry, 130*, 79–83.
- Mendel, R., Traut-Mattausch, E., Jonas, E., Leucht, S., Kane, J. M., Maino, K. et al. (2011). Confirmation bias: why psychiatrists stick to wrong preliminary diagnosis. *Psychological Medicine, 41*(12), 2651–2659.
- Merton, R. K. (1948). The self-fulfilling prophecy. *Antioch Review, 8*, 193–210.
- Miller, M. J. (1985). Counselor as hypothesis tester: Some implications of research. *Journal of Counseling and Development, 63*, 276–278.
- Millon, T. (1975). Reflections on Rosenhan's "on being sane in insane places". *Journal of Abnormal Psychology, 84*(5), 456–461.

- Modes, P. (2005). *Auswirkungen von Verdacht auf das nonverbale Verhalten von Interviewern in Interviews mit Kindern im Vorschulalter*: Unveröffentlichte Diplomarbeit, Christian-Albrechts-Universität zu Kiel.
- Mumma, G. H., & Wilson S. B. (1995). Procedural debiasing of primacy/anchoring effects in clinical-like judgements. *Journal of Clinical Psychology*, 51(6), 841–853.
- Mussweiler, T., Strack, F., & Pfeiffer, T. (2000). Overcoming the inevitable anchoring effect: Considering the opposite compensates for selective accessibility. *Personality and Social Psychology Bulletin*, 26(3), 1142–1150.
- Newman, J., Wolff, W. T., & Hearst, E. (1980). The feature-positive effect in adult human subjects. *Journal of Experimental Psychology*, 6(5), 630–650.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175–220.
- O'Brien, B. (2009). Prime suspect: An examination of factors that aggravate and counteract confirmation bias in criminal investigations. *Psychology, Public Policy, and Law*, 15(4), 315–334.
- Oswald, M. E. (in press). When hypotheses override therapists' information about the client. In F. Caspar (Ed.), *The inner processes of psychotherapists. Innovations in clinical training*. Oxford: Oxford University Press.
- Oswald, M. E., & Grosjean. (2004). Confirmation bias. In R. F. Pohl (Ed.), *Cognitive illusions. A handbook on fallacies and biases in thinking, judgement and memory* (pp. 79–96). Hove: Psychology Press.
- Oswald, M., & Gadenne, V. (1986). Der Einfluss von Speicherung und Suchstrategie auf die Erinnerungsleistung hypothesekonformer und -konträrer Informationen. *Sprache & Kognition*, 5(4), 195–201.
- Parmley, M. C. (2006). *The effects of the confirmation bias on diagnostic decision making*. Unveröffentlichte Dissertation, Drexel University Philadelphia, Philadelphia.
- Paul, T., & Thiel, A. (2005). *Eating Disorder Inventory 2 (EDI-2)*. Göttingen: Hogrefe.
- Pepinsky, J. B., & Pepinsky, N. (1954). *Counseling theory and practice*. New York: Ronald Press.
- Pines, H. A., & Julian, J. W. (1972). Effects of task and social demands on locus of control differences in information processing. *Journal of Personality and Social Psychology*, 40(3), 407–416.
- Pohl, R. (2004). Introduction: Cognitive illusions. In R. F. Pohl (Ed.), *Cognitive illusions. A handbook on fallacies and biases in thinking, judgement and memory* (pp. 1–20). Hove: Psychology Press.
- Popper, K. R. (1959). *The logic of scientific discovery*. New York: Basic Books.
- Potchen, E. J. (2006). Measuring observer performance in chest radiology: some experiences. *Journal of the American College of Radiology*, 3(6), 423–432.

- Pottick, K., Kirk, S., Hsieh, D., & Tian, X. (2007). Judging mental disorder in youths: Effects of client, clinician, and contextual differences. *Journal of Consulting and Clinical Psychology, 75*(1), 1–8.
- Pudel, V., & Westenhöfer, J. (1991). *Ernährungspsychologie: Eine Einführung*. Göttingen: Hogrefe.
- Rabichow, H. G., & Pharis, M. E. (1974). Rosenhan was wrong: The staff was lousy. *Clinical Social Work Journal, 2*(4), 271–278.
- Rassin, E. (2008). Individual differences in the susceptibility to confirmation bias. *Netherlands Journal of Psychology, 64*(2), 87–93.
- Rassin, E., Eerland, A., & Kuijpers, I. (2010). Let's find the evidence: An analogue study of confirmation bias in criminal investigations. *Journal of Investigative Psychology and Offender Profiling, 7*(3), 231–246.
- Rosenhan, D. L. (1973). On Being Sane In Insane Places. *Science, 179*, 250–258.
- Rosenhan, D. L. (1975). The contextual nature of psychiatric diagnosis. *Journal of Abnormal Psychology, 84*(5), 462–474.
- Rosenthal, R. (1976). *Experimenter effects in behavioral research*. New York: Irvington/Wiley.
- Sandifer, M. G., Hordern, A., & Green, L. M. (1970). The psychiatric interview: The impact of the first three minutes. *American Journal of Psychiatry, 126*(7), 968–973.
- Saß, H. (1999). Die Verankerung der Diagnostik in der Psychiatrie. In R. S. Jäger & F. Petermann (Hrsg.), *Psychologische Diagnostik*. Weinheim: Psychologische Verlags Union.
- Saß, H., Wittchen, H.-U., Zaudig, M., & Houben, I. (2003). *Diagnostisches und Statistisches Manual Psychischer Störungen. (DSM-IV-TR): Textrevision*. Göttingen: Hogrefe.
- Schemm, K. vom. (2008a). *Auf der Suche nach dem Missing Link: Welche Bedeutung hat das Testen sozialer Hypothesen für das suggestive Befragungsverhalten in Interviews?* Frankfurt: Verlag für Polizeiwissenschaft.
- Schemm, K. vom. (2008b). *MIA-Training*, Unveröffentlichtes Manuskript, Christian-Albrechts-Universität, Kiel.
- Schemm, K. vom, Dreger, B., & Köhnken, G. (2008). Suggestion und konfirmatorisches Testen sozialer Hypothesen in Befragungssituationen. *Forensische Psychiatrie, Psychologie, Kriminologie, 2*, 20–27.
- Schneller, K. & Schneider, W. (2004). *Bundesweite Befragung der Absolventinnen und Absolventen des Jahres 2003 im Studiengang Psychologie im Auftrag der DGPs*. Zugriff am 09.02.2012. http://www.dgps.de/studium/danach/_download/bericht.pdf
- Schuler, H. (2000). *Psychologische Personalauswahl*. Göttingen: Hogrefe.
- Schulz-Hardt, S., & Köhnken, G. (2000). Wie ein Verdacht sich selbst bestätigen kann: Konfirmatorisches Hypothesentesten als Ursache von Falschbeschuldigungen wegen sexuellen Kindesmissbrauchs. *Praxis der Rechtspsychologie, 10*(1), 60–88.

- Schweizer, M. (2005). *Kognitive Täuschungen vor Gericht. Eine empirische Studie*, Unveröffentlichte Dissertation, Universität Zürich, Zürich.
- Scott, W. A. (1969). Structure of natural cognitions. *Journal of Personality and Social Psychology*, *12*, 261–278.
- Semin, G. R., & Strack, F. (1980). The plausibility of the implausible: a critique of Snyder and Swann (1978). *European Journal of Social Psychology*, *10*(4), 379–388.
- Semin, G., & De Poot, C. (1997). Bringing partiality to light: Question wording and choice as indicators of bias. *Social Cognition*, *15*(2), 91–106.
- Shepherd, E. (1988). Developing interview skills. In P. Southgate (Ed.), *New directions in police training* (pp. 170–188). London: HMSO.
- Shipko, S. (1982). A correct diagnosis of a pseudopatient in an inpatient setting. *Professional psychology: research and practice*, *13*(3), 329.
- Skov, R., & Sherman, S. (1986). Information gathering processes: Diagnosticity, hypothesis-confirmatory strategies and perceives hypothesis confirmation. *Journal of Experimental Social Psychology*, *22*(2), 93–121.
- Slater, L. (2004). *Opening skinner's box: Great psychological experiments of the twentieth century*. New York: W W Norton & Co.
- Slater, L. (2005). Reply to Spitzer and colleagues. *Journal of Nervous and Mental Disease*, *193*(11), 743–744.
- Slovic, P., & Fischhoff, B. (1977). On the psychology of experimental surprises. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(4), 544–551.
- Smith, J., & Kida, T. (1991). Heuristics and biases: Expertise and task realism in auditing. *Psychological Bulletin*, *109*(3), 472–489.
- Snyder, M., & Cantor, N. (1979). Testing hypotheses about other people: The use of historical knowledge. *Journal of Experimental Social Psychology*, *15*(4), 330–342.
- Snyder, M., & Swann, W. (1978). Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology*, *36*(11), 1202–1212.
- Spengler, P. M., & Strohmer, D. C. (1994). Clinical judgemental biases: The moderating roles of counselor cognitive complexity and counselor client preferences. *Journal of Counseling Psychology*, *41*(1), 8–17.
- Spengler, P. M., Strohmer, D. C., Dixon, D. N., & Shivy, V. A. (1995). A scientist-practitioner model of psychological assessment: Implications for training, practice and research. *The Counseling Psychologist*, *23*(3), 506–534.
- Spengler, P. M., White, M. J., Ægisdóttir, S., Maugherman, A. S., Anderson, L. A., Cook, R. S. et al. (2009). The meta-analysis of clinical judgment-project: Effects of experience on accuracy. *The Counseling Psychologist*, *37*(3), 350–399.
- Spitzer, R. L. (1975). On pseudoscience in science, logic in remission, and psychiatric diagnosis: A critique of Rosenhan's "On being sane in insane places". *Journal of Abnormal Psychology*, *84*(5), 413–418.

- Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social development literatures. *Psychological Bulletin, 111*(1), 42–61.
- Statistisches Bundesamt. (2009). *Bildung und Kultur. Studierende an Hochschulen. WS 2008/2009*. (Fachserie 11, Reihe 4.11). Zugriff am 09.02.2012. <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Publikationen/Fachveroeffentlichungen/BildungForschungKultur/Hochschulen/StudierendeHochschulenEndg2110410097004,property=file.pdf>
- Statistisches Bundesamt. (2011). *Gesundheitsberichterstattung des Bundes*. Zugriff am 09.02.2012. http://www.gbe-bund.de/oowa921-install/servlet/oowa/aw92/dboowasys921.xwdevkit/xwd_init?gbe.isgbetol/xs_start_neu/&p_aid=3&p_aid=55203821&nummer=697&p_sprache=D&p_indsp=-&p_aid=62104763
- Stoltenberg, C. D., Pace, T. M., Kashubeck-West, S., Biever, J. L., Patterson, T., & Welch, I. D. (2000). Training models in counseling psychology: Scientist-practitioner versus practitioner-scholar. *The Counseling Psychologist, 28*, 622–640.
- Strack, F., & Deutsch, R. (2002). Urteilsheuristiken. In D. Frey & M. Irle (Hrsg.), *Motivations-, Selbst- und Informationsverarbeitungstheorien: Bd. III. Theorien der Sozialpsychologie* (S. 352–384). Bern: Huber.
- Strohmer, D. C., & Chiodo, A. L. (1984). Counselor hypothesis testing strategies: The role of initial impressions and self-schema. *Journal of Counseling Psychology, 31*(4), 510–519.
- Strohmer, D. C., & Newman, L. J. (1983). Counselor hypothesis-testing strategies. *Journal of Counseling Psychology, 30*(4), 557–565.
- Strohmer, D. C., & Shivy, V. A. (1994). Bias in counselor hypothesis testing: Testing the robustness of counselor confirmatory bias. *Journal of Counseling & Development, 73*(2), 191–197.
- Strohmer, D. C., Shivy, V. A., & Chiodo, A. L. (1990). Information processing strategies in counselor hypothesis testing: The role of selective memory and expectancy. *Journal of Counseling Psychology, 37*(4), 465–472.
- Strong, S. R. (1968). Counseling: An interpersonal influence process. *Journal of Counseling Psychology, 15*(3), 215–224.
- Swann, W., & Giuliano, T. (1987). Confirmatory search strategies in social interaction: How, when, why and with what consequences. *Journal of Social and Clinical Psychology, 5*(4), 511–524.
- Temerlin, M. K. (1968). Suggestion effects in psychiatric diagnosis. *Journal of Nervous and Mental Disease, 147*(4), 349–353.
- Trope, Y., & Bassok, M. (1983). Information-gathering strategies in hypothesis-testing. *Journal of Experimental Social Psychology, 19*(6), 560–576.
- Trope, Y., & Liberman, A. (1996). *Social hypothesis testing: Cognitive and motivational mechanisms*. (Kruglanski, A., Ed.). New York: Guilford Press.

- Trope, Y., & Mackie, D. (1987). Sensitivity to alternatives in social hypothesis-testing. *Journal of Experimental Social Psychology*, 23(6), 445–459.
- Trope, Y., Bassok, M., & Alon, E. (1984). The questions lay interviewers ask. *Journal of Personality and Social Psychology*, 52(1), 90–106.
- Turk, D. C., & Salovey, P. (1985). Cognitive structures, cognitive processes, and cognitive behavior modification: I. Client issues. *Cognitive Therapy and Research*, 9, 1–17.
- Tversky, A. & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131.
- Vogelgesang, F., Schulz-Hardt, S., & Köhnken, G. (2006). *Schuldig bei Verdacht: Entstehung von Falschbeschuldigungen durch Selbstbestätigungstendenzen beim Testen sozialer Hypothesen. DFG-Neuantrag*. Georg-August-Universität und Christian-Albrechts-Universität, Göttingen und Kiel.
- Walker, B., & Spengler, P. M. (1995). Clinical judgment of major depressions in AIDS patients: The effects of clinician complexity and stereotyping. *Professional psychology: research and practice*, 26, 269–273.
- Wason, P. (1960). On the failure to eliminate hypothesis in a conceptual task. *The Quarterly Journal of Experimental Psychology*, 12, 129–140.
- Wason, P., & Johnson-Laird, P. (1972). *The psychology of reasoning: Structure and content*. Cambridge, MA: Harvard University Press.
- Weathers, P. F. M. & S. D. (1993). Mail survey research in counseling psychology: Current practice and suggested guidelines. *Journal of Counseling Psychology*, 40(2), 238–244.
- Weiner, B. (1975). “On being sane in insane places”: A process (attributional) analysis and critique. *Journal of Abnormal Psychology*, 84(5), 433–441.
- Wietersheim, J. von. (2003). Bulimia nervosa. In R. H. Adler, J. M. Hermann, K. Köhle, W. Langewitz, O. W. Schonecke, T. Uexküll, & von W. Wesiack (Hrsg.), *Psychosomatische Medizin* (S. 715–716). München: Urban & Fischer.
- Wilson Dallas, M. E., & Baron, R. S. (1985). Do psychotherapists use a confirmatory strategy during interviewing? *Journal of Social and Clinical Psychology*, 3(1), 106–122.
- Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen: Hogrefe.
- Wittek, N. (2003). *Untersuchungen zur Körperschemastörung bei Anorexia nervosa*. Unveröffentlichte Dissertation, Julius-Maximilians-Universität, Würzburg.
- Wyer, R., & Scrull, T. (1986). *Memory and cognition in its social context*. Hillsdale, NJ: Lawrence Erlbaum.

Abbildungsverzeichnis

Abbildung 1.	Integrative Darstellung dreier Konzeptionen des Ablaufs einer Hypothesentestung	29
Abbildung 2.	Veranschaulichung der konfirmatorischen Mechanismen.....	30
Abbildung 3.	Konsequenzen eines Hypothesentestungsprozesses	50
Abbildung 4.	Unterscheidung zwischen induzierter Hypothese und internem Anfangsverdacht	62
Abbildung 5.	Bildschirmansicht zur Einordnung in die Berufsgruppe.....	99
Abbildung 6.	Überschneidung der diagnostischen Kriterien von Bulimie und Anorexie (Subtypus Binge-Eating/Purging) im Sinne des DSM-IV	103
Abbildung 7.	Veranschaulichung Konstruktion und Auswahl der Zusatzitems.....	106
Abbildung 8.	Bildschirmansicht der Instruktion zur Generierung der sechs Evidenzregeln sowie der Eingabefelder (Fachwissen relevant)	113
Abbildung 9.	Verteilung der Studienabbrüche über den Studienverlauf.....	136
Abbildung 10.	Altersverteilung von Studienteilnehmern (Stichprobe) und Psychotherapeuten in Deutschland (Population) zum Studienzeitpunkt.....	143
Abbildung 11.	Boxplot zur Veranschaulichung des Zusammenhangs der Überzeugtheit von der Prüfhypothese und der Schlussfolgerung in den Bedingungskombinationen von Fachwissen und Berufserfahrung.....	146
Abbildung 12.	Veranschaulichung des Interaktionseffekts zwischen Berufserfahrung und Fachwissen.....	162
Abbildung C-1.	Erscheinungsbild Online-Studie	278
Abbildung C-2.	Bedingungsinvarianter Begrüßungsbildschirm mit Abfrage des beruflichen Status	279
Abbildung C-3.	Bedingungsinvariante Erläuterungen zu Inhalt, Ablauf und Vertraulichkeit.....	280
Abbildung C-4.	Bedingungsinvariante technischen Erläuterungen zur Bearbeitung der Studie.....	280
Abbildung C-5.	Bedingung Fachwissen relevant: Einführung in die zu bearbeitende klinische Fragestellung	281
Abbildung C-6.	Bedingung Fachwissen irrelevant: Einführung in die zu bearbeitende kriminalistische Fragestellung	281
Abbildung C-7.	Bedingung Fachwissen relevant: Hinweis auf die Hintergrundinformationen, die über einen Button zu öffnen sind	281

Abbildung C-8.	Bedingung Fachwissen irrelevant: Hinweis auf die Hintergrundinformationen, die über einen Button zu öffnen sind	281
Abbildung C-9.	Bedingungsinvarianter Hinweis auf die Fallbeschreibung, die über einen Button zu öffnen ist	281
Abbildung C-10.	Bedingung Fachwissen relevant: Abfrage der Überzeugtheit von der Prüfhypothese	282
Abbildung C-11.	Bedingung Fachwissen irrelevant: Abfrage der Überzeugtheit von der Prüfhypothese	282
Abbildung C-12.	Bedingung Fachwissen relevant: Abfrage der AV 3 (Generieren von Evidenzregeln)	282
Abbildung C-13.	Bedingung Fachwissen irrelevant: Abfrage der AV 3 (Generieren von Evidenzregeln)	283
Abbildung C-14.	Bedingung Fachwissen relevant: Instruktionen zur AV 4 (Bestätigungstendenz B)	283
Abbildung C-15.	Bedingung Fachwissen irrelevant: Instruktionen zur AV 4 (Bestätigungstendenz B)	283
Abbildung C-16.	Bedingung Fachwissen relevant: Exemplarische Bildschirmansicht zur Abfrage der AV 4 (Bestätigungstendenz B) ...	284
Abbildung C-17.	Bedingung Fachwissen irrelevant: Exemplarische Bildschirmansicht zur Abfrage der AV 4 (Bestätigungstendenz B) ...	284
Abbildung C-18.	Bedingungsinvariant: Täuschung darüber, dass der PC nun zufällig einen Blickwinkel auswählte, der im Folgenden zu beachten sei	284
Abbildung C-19.	Bedingung Fachwissen relevant: Hinweis, welcher Blickwinkel ausgewählt wurde. Alle Probanden erhielten den Blickwinkel Bulimie.	284
Abbildung C-20.	Bedingung Fachwissen irrelevant: Hinweis, welcher Blickwinkel ausgewählt wurde. Alle Probanden erhielten den Blickwinkel schuldig.....	285
Abbildung C-21.	Bedingung Fachwissen relevant: Instruktionen zur AV 5a + AV 5b.	285
Abbildung C-22.	Bedingung Fachwissen irrelevant: Instruktionen zur AV 5a + AV 5b.	285
Abbildung C-23.	Bedingungsinvariant: Exemplarische Bildschirmansicht zur Abfrage der AV 5a + AV 5b	285
Abbildung C-24.	Bedingung Fachwissen relevant: Abfrage der Schlussfolgerung (AV 1)	286
Abbildung C-25.	Bedingung Fachwissen irrelevant: Abfrage der Schlussfolgerung (AV 1)	286
Abbildung C-26.	Bedingungsinvariante Abfrage der subjektiven Sicherheit (AV 2)	286

Abbildung C-27. Bedingung Fachwissen relevant: Abfrage der Themen für die erste Therapiesitzung und zu verordnenden Hausaufgaben (AV 6) ...	286
Abbildung C-28. Bedingungsinvariante Abfrage von Altern, Geschlecht und Wohnort	287
Abbildung C-29. Abfrage der biographischen Daten, wenn auf dem Begrüßungsbildschirm angegeben wurde, Studierender zu sein	288
Abbildung C-30. Abfrage der biographischen Daten, wenn auf dem Begrüßungsbildschirm angegeben wurde, Psychologischer Psychotherapeut und/oder Kinder- und Jugendlichenpsychotherapeut zu sein	289
Abbildung C-31. Abfrage der biographischen Daten, wenn auf dem Begrüßungsbildschirm angegeben wurde, Facharzt zu sein	289
Abbildung C-32. Abfrage der biographischen Daten, wenn auf dem Begrüßungsbildschirm angegeben wurde, als Diplom-Psychologe ohne Therapieausbildung im klinischen Bereich tätig zu sein	290
Abbildung C-33. Abfrage zur Charakterisierung des Arbeitsumfelds	290
Abbildung C-34. Bedingungsinvariante Abfrage zur SV Bewusstsein hinsichtlich konfirmatorischer Prozesse.....	291
Abbildung C-35. Bedingungsinvariante Danksagung und Hinweis auf die Erhöhung des Spendenzählers	291
Abbildung C-36. Bedingungsinvariante Aufklärung und Verabschiedung der Studienteilnehmer	291

Tabellenverzeichnis

Tabelle 1.	Veranschaulichung der Validität eines Urteils in Abhängigkeit des wahren Zustandes.....	58
Tabelle 2.	Häufigkeit der Eignungsurteile in Abhängigkeit von der Valenz der Hypothese.....	65
Tabelle 3.	Überblick über debiasing techniques beim Hypothesentesten.....	77
Tabelle 4.	Maßnahmen zur Erreichung einer ergebnisoffenen Grundhaltung und der Einbeziehung von Alternativhypothesen im Trainingsprogramm MiA (vom Schemm, 2008b).....	83
Tabelle 5.	Psychologische Hypothesenkomplexe zur Überzeugtheit von der Prüfhypothese, der Berufserfahrung und dem Fachwissen.....	96
Tabelle 6.	Einschätzung der Wahrscheinlichkeit des Vorliegens einer Bulimie bzw. Anorexie sowie des Ausmaßes, in dem die Informationen charakteristisch für Bulimie bzw. Anorexie sind.....	105
Tabelle 7.	Einschätzung der Wahrscheinlichkeit der möglichen Urteilsoptionen sowie des Ausmaßes, in dem die Informationen charakteristisch für die jeweiligen Optionen sind.....	111
Tabelle 8.	Beispiele zur Veranschaulichung der positiven vs. negativen Teststrategie bei den Evidenzregeln.....	114
Tabelle 9.	Beispiele zur Veranschaulichung von diagnostischen und nicht diagnostischen Inhalten von Evidenzregeln.....	114
Tabelle 10.	Begründung der Einteilung in konsistente vs. inkonsistente Items (Essstörung).....	116
Tabelle 11.	Untersuchung der Ausbalancierung der konsistenten vs. inkonsistenten Items (Essstörung).....	117
Tabelle 12.	Untersuchung der Ausbalancierung der konsistenten vs. inkonsistenten Items (Brandstiftung).....	117
Tabelle 13.	Begründung der Einteilung in konsistente vs. inkonsistente Items (Brandstiftung).....	118
Tabelle 14.	Übersicht über die (Reihenfolge der) verwendeten Items für AV 5 (Essstörung).....	121
Tabelle 15.	Übersicht über die (Reihenfolge der) verwendeten Items für AV 5 (Brandstiftung).....	123
Tabelle 16.	Übersicht über die verwendeten Maße des Effekts und deren verbale Klassifizierung.....	133
Tabelle 17.	Beurteilerübereinstimmung für Konformität, Teststrategie & Diagnostizität pro Evidenzregel.....	139

Tabelle 18.	Beurteilerübereinstimmung für die Anzahl berücksichtigter Hypothesen bzw. Anzahl Evidenzregeln für Prüf- und Alternativhypothese	140
Tabelle 19.	Beurteilerübereinstimmung für die Einschätzung der Konsistenz der vorgeschlagenen Themen bzw. Hausaufgaben im Hinblick auf eine typische Behandlung von Bulimikern bzw. Anorektikern	140
Tabelle 20.	Deskriptive Kennwerte des Alters (in Jahren) und des Geschlechts in Abhängigkeit vom Fachwissen	145
Tabelle 21.	Zusammenfassung der logistischen Regressionsanalyse	148
Tabelle 22.	Mittelwerte und Streuungen der subjektiven Sicherheit in den Bedingungskombinationen von Berufserfahrung und Fachwissen	149
Tabelle 23.	Tafel der zweifaktoriellen Varianzanalyse zur Überprüfung der Abhängigkeit der subjektiven Sicherheit von der Berufserfahrung und dem Fachwissen	149
Tabelle 24.	Mittelwerte und Streuungen der Variable „Überwiegen der Evidenzregeln im Sinne einer positiven Teststrategie“ in den Bedingungskombinationen von Berufserfahrung und Fachwissen	150
Tabelle 25.	Tafel der Kovarianzanalyse zur statistischen Prüfung der PH-B ₁ zum Überwiegen positiver Evidenzregeln	151
Tabelle 26.	Mittelwerte und Streuungen der Anzahl berücksichtigter Hypothesen in den Bedingungskombinationen der Variablen Berufserfahrung und Fachwissen	153
Tabelle 27.	Tafel der Kovarianzanalyse zur statistischen Prüfung der PH-B ₄ zur Anzahl berücksichtigter Hypothesen	153
Tabelle 28.	Mittelwerte und Streuungen der Variablen Überwiegen der Evidenzregeln zur Prüfhypothese in den Bedingungskombinationen der Variablen Berufserfahrung und Fachwissen	154
Tabelle 29.	Tafel der Kovarianzanalyse zur statistischen Prüfung der PH-B5 zum Überwiegen von Evidenzregeln zur Prüfhypothese	155
Tabelle 30.	Mittelwerte und Streuungen der Variablen Anzahl der Aufwertungen in den Bedingungskombinationen der Variablen Berufserfahrung und Fachwissen	156
Tabelle 31.	Tafel der zweifaktoriellen Varianzanalyse zur statistischen Prüfung der PH-C zur Anzahl von Aufwertungen	156
Tabelle 32.	Mittelwerte und Streuungen der Variablen Anzahl der nicht erkannten Widersprüche in den Bedingungskombinationen der Variablen Berufserfahrung und Fachwissen	157
Tabelle 33.	Tafel der zweifaktoriellen Varianzanalyse zur statistischen Prüfung der PH-D zur Anzahl der nicht erkannten Widersprüche	158

Tabelle 34. Mittelwerte und Streuungen der Variablen Anzahl der nicht erkannten Widersprüche in den Bedingungskombinationen der Variablen Berufserfahrung und Fachwissen.....	159
Tabelle 35. Tafel der Kovarianzanalyse zur statistischen Prüfung der PH-E ₁ zur Beurteilung der Wichtigkeit von konsistenten und inkonsistenten Informationen	160
Tabelle 36. Mittelwerte und Streuungen der Summe des Einflusses von konsistenten und inkonsistenten Items in den Bedingungskombinationen der Variablen Berufserfahrung und Fachwissen	160
Tabelle 37. Tafel der Kovarianzanalyse zur statistischen Prüfung der PH-E ₂ zur Beurteilung des Einflusses von konsistenten und inkonsistenten Informationen	161
Tabelle 38. Absolute Häufigkeiten bzgl. der Konsistenz von Themenvorschlägen mit der Therapie von Essstörungen.....	163
Tabelle 39. Absolute Häufigkeiten bzgl. der Konsistenz von Themenvorschlägen mit der Therapie von Essstörungen.....	163
Tabelle A-1. Überblick über die 39 Items sowie deren Kennwerten aus dem gemeinsamen Rating	234
Tabelle B-1. Überblick über die 42 Items sowie deren Kennwerten aus dem gemeinsamen Rating	266

Anhang A - Fallmaterialien Essstörung

Fallbeschreibung Fachwissen relevant (Essstörung)	211
Hintergrundwissen Fachwissen relevant (Essstörung)	213
Materialien zur Eignungsprüfung der Fallbeschreibung.....	223
Konstruktion der Zusatzinformationen_Instruktionen.....	229
Konstruktion der Zusatzinformationen_Items und Kennwerte	233

Fallbeschreibung Fachwissen relevant (Essstörung)

Fallinformationen

Marisa (18;6) wurde von ihrer Mutter in der Ambulanz einer Psychosomatischen Klinik für Jugendliche und Junge Erwachsene vorgestellt. Die Mutter von Marisa habe sich zunächst an den Hausarzt der Tochter gewandt, da sie wegen des von ihr beobachteten sozialen und emotionalen Rückzugs der Tochter, deren niedergeschlagenen Stimmung und den Problemen mit dem Essverhalten besorgt gewesen sei. Diesbezüglich habe die Mutter beobachtet, dass Marisa in den letzten Wochen immer wieder zu- und abgenommen habe, wobei die Tochter sich selbst nicht als zu dünn, sondern eher als zu dick einschätze. Der Arzt habe Auffälligkeiten im Elektrolythaushalt von Marisa festgestellt und die Vorstellung in einer Klinik empfohlen, da er den Verdacht hatte, dass eine Bulimie-Erkrankung vorliegt. In der Klinik wurde ein Erstgespräch geführt, nach dem Marisa stationär in der Klinik aufgenommen wurde.

Im Rahmen des Aufnahmegesprächs habe die Patientin nach Aussage des Interviewers teilweise abwesend gewirkt, den Blickkontakt gemieden und sich dem Gespräch entzogen. Ihre einzige Einlassung habe darin bestanden, zu äußern, dass sie durchaus mit Genuss essen wolle. Die folgenden Informationen stammen daher von der Mutter.

Marisa lebe mit ihrer jüngeren Schwester, mit der sie sich gut verstehe, bei den Eltern. Die Mutter sei Hausfrau, der Vater verdiene den Lebensunterhalt der Familie als Beamter. Die Mutter berichtete, dass die Schwangerschaft und die Geburt von Marisa unkompliziert verlaufen seien. Außerdem habe es keine Auffälligkeiten in der frühkindlichen Entwicklung gegeben und Marisa sei nicht ernsthaft krank gewesen. Gefragt nach der Familienanamnese, gab die Mutter an, dass sie selbst wegen depressiver Episoden und Beginn einer adipösen Entwicklung in psychotherapeutischer Behandlung gewesen sei. Ansonsten gebe es keine psychischen und medizinischen Auffälligkeiten in der Familie. Über den schulischen Werdegang von Marisa berichtete die Mutter, dass sie regelgerecht eingeschult worden sei und heute die 13. Klasse des Gymnasiums besuche. Sie sei immer gut in die Klassengemeinschaft integriert gewesen und habe durchweg gute bis sehr gute Leistungen erzielt. Sie erledige ihre Hausaufgaben sehr gewissenhaft und wende viel Zeit für die Schule auf, wenn sie nicht gerade einer ihrer vielen sportlichen Aktivitäten nachgehe. Dazu gehörten bspw. das morgendliche Joggen sowie das Schwimmtraining.

Bezüglich der aktuellen Problematik gab die Mutter an durch den sozialen Rückzug von Marisa beunruhigt zu sein. Sie beschrieb, dass Marisa Freundschaften vernachlässige, z.B. vermehrt Verabredungen ablehne. Auch aus dem Familienalltag ziehe Marisa sich mehr und mehr zurück. So habe Marisa keinerlei Interesse mehr an Familienfeiern sowie gemeinsamen Unternehmungen und suche nach Ausreden, um nicht teilnehmen zu müssen.

Am meisten missfalle der Mutter, dass Marisa sich auch häufig dem gemeinsamen Essen der Familie zu entziehen versuche. Mit den Argumenten, sie habe schon gegessen oder habe keinen Hunger, versuche Marisa, das gemeinsame Beisammensein zu vermeiden. Wenn die Eltern sie dazu zwingen, dann setze Marisa sich zwar zu ihnen und esse auch etwas, allerdings stünde sie dann auch so schnell wie möglich wieder auf.

Auch die wechselhaften Stimmungen der Tochter machten der Mutter Sorgen. Teilweise sei Marisa sehr reizbar und werde schnell wütend, so dass es zu heftigen Streitereien zwischen ihr und den Eltern komme. Andererseits habe die Mutter beobachtet, dass Marisa auch häufig niedergeschlagen wirke. Neben dem Desinteresse an gemeinsamen Aktivitäten, mache sie häufig den Eindruck, traurig zu sein. Einem Gespräch über ihre Gefühle entziehe sie sich.

Hintergrundwissen Fachwissen relevant (Essstörung)

5 Essstörungen

In vielen Kulturen spielt Essen eine große Rolle. Es gibt unzählige Feinschmecker-Restaurants und außerdem widmen sich zahlreiche Zeitschriften sowie Fernsehsendungen der Zubereitung von Lebensmitteln. Andererseits gibt es viele Übergewichtige und Diäten sind an der Tagesordnung. Der Wunsch vieler Menschen, schlanker zu sein, hat einen florierenden Wirtschaftszweig hervorgerufen. Aufgrund dieses starken Interesses an Lebensmitteln und am Essen überrascht es nicht, dass dieses Verhalten auch zu Störungen führen kann.

Im Folgenden sollen nun die wichtigsten Merkmale der **Anorexia nervosa** und der **Bulimia nervosa** beschrieben werden. Diese beiden Diagnosen haben mehrere gemeinsame Merkmale, darunter vor allem die starke Angst, übergewichtig zu sein. Auch die Komorbiditäten der Störungen sind ähnlich und werden deshalb im Anschluss gemeinsam ausgeführt.

Anorexia bezeichnet einen schweren Appetitverlust und **nervosa** bedeutet, dass die Gründe dafür emotionaler Natur sind. Der Begriff *Anorexie* ist etwas irreführend, da viele Anorektikerinnen nicht ihren Appetit oder das Interesse am Essen verlieren. Im Gegenteil – die meisten hungern zwar, befassen sich jedoch intensiv mit Essen, lesen ständig Kochbücher und Kochen für ihre Familien die besten Gerichte.

Das Wort **Bulimia** ist vom Griechischen abgeleitet und bedeutet *Ochsenhunger*. Patientinnen mit Bulimia nervosa zeichnen sich durch einen Wechsel von gezügeltem Essen und Fressanfällen mit anschließendem Purgingverhalten aus (Adler, 2003). Die interozeptive Wahrnehmung der Bulimie-Patientinnen ist gestört, so dass Sättigungs- und Hungergefühle auch ganz verschwunden sein können. Meist sind sie normalgewichtig oder leicht übergewichtig und ihr Ziel ist es, nicht zuzunehmen. Aufgrund des Wechsels von Phasen gezügelten Essens und Fressanfällen kann das Gewicht von Patientinnen mit Bulimia nervosa stark schwanken.

5.1 Anorexia nervosa

Tabelle 1 nennt die Kriterien für Anorexia nervosa nach DSM-IV.

Tab. 1: Diagnostische Kriterien für Anorexia nervosa nach DSM-IV (Sass et al., 1998)	
A	Weigerung, das Minimum des für Alter und Körpergröße normalen Körpergewichts zu halten (z.B. der Gewichtsverlust führt dauerhaft zu einem Körpergewicht von weniger als 85% des zu erwartenden Gewichts; oder das Ausbleiben einer während der Wachstumsperiode zu erwartenden Gewichtszunahme führt zu einem Körpergewicht von weniger als 85% des zu erwartenden Gewichts).
B	Ausgeprägte Ängste vor einer Gewichtszunahme oder davor, dick zu werden , trotz bestehenden Untergewichts.
C	Störung in der Wahrnehmung der eigenen Figur und des Körpergewichts , übertriebener Einfluss des Körpergewichts oder der Figur auf die Selbstbewertung oder Leugnen des Schweregrades des gegenwärtigen geringen Körpergewichts.
D	Bei postmenarchalen Frauen das Vorliegen einer Amenorrhoe , d.h. das Ausbleiben von mindestens drei aufeinander folgenden Menstruationszyklen. (Kriterium nicht anwendbar bei hormoneller Verhütung.)
	Restriktiver Typus (F50.00): Während der aktuellen Episode der Anorexia nervosa hat die Person keine regelmäßigen <i>Fressanfälle</i> gehabt oder hat kein <i>Purging</i> -Verhalten (d.h. selbstinduziertes Erbrechen oder Missbrauch von Laxantien, Diuretika oder Klistieren) gezeigt.
	Binge-Eating/Purging-Typus (F50.01): Während der aktuellen Episode der Anorexia nervosa hat die Person regelmäßig Fressanfälle gehabt und hat Purgingverhalten (d.h. selbstinduziertes Erbrechen oder Missbrauch von Laxantien, Diuretika oder Klistieren) gezeigt.

⇒ **Zwei Subtypen.** Für die Differenzierung in einen restriktiven Typus und einen Binge-Eating/Purging-Typus sprechen zahlreiche Unterschiede zwischen diesen beiden Untertypen. Der Binge-Eating/Purging-Typus scheint psychopathologischer zu sein, d.h. die Betroffenen weisen mehr

Klinische Störungsbilder kompakt

Persönlichkeitsstörungen und impulsives Verhalten auf und ziehen sich vermehrt von ihrem sozialen Umfeld zurück (Davison & Neale, 2002).

⇒ **Epidemiologie.** Die Anorexia nervosa setzt typischerweise in den frühen bis mittleren Jugendjahren ein und zwar häufig nach einer Diät und einem belastenden Ereignis. Bei Frauen ist die Störung etwa zehnmal so häufig wie bei Männern, wobei die Lebenszeitprävalenz bei etwas unter einem Prozent liegt (Hsu, 1990).

⇒ **Gestörte Körperwahrnehmung.** Auch wenn die Betroffenen schon stark abgemagert sind, behaupten sie immer noch, übergewichtig zu sein, oder dass bestimmte Körperteile, insbesondere Bauch, Gesäß und Oberschenkel, noch zu dick seien. Sie überschätzen also trotz ihres Untergewichts ihre Körperdimensionen. Ihre Selbstachtung hängt stark davon ab, dass sie dünn bleiben und bevorzugt noch mehr abnehmen.

Die gestörte Körperwahrnehmung wurde auf unterschiedliche Weise, meist jedoch mit Hilfe eines Fragebogens, wie dem Essstörungsinventar (EDI-Eating Disorder Inventory) erfasst (Paul & Thiel, 2005). Bei einer anderen Methode zur Erfassung der Körperbildstörung werden den Patientinnen Zeichnungen von Frauenfiguren mit unterschiedlichem Körpergewicht vorgelegt, und sie müssen diejenige ankreuzen, die ihrer eigenen Figur am ähnlichsten ist, und diejenige, die für sie eine Idealfigur darstellt. Während Patientinnen mit Anorexia nervosa ihre eigene Figur überschätzen und eine dünne als ihre Idealfigur auswählen, zeigen Patientinnen mit Bulimia nervosa (s.u.) diese gestörte Wahrnehmung nicht (Wittek, 2003). In diesem Zusammenhang fällt auch auf, dass Magersüchtige vom Körper ausgehende Reize verändert wahrnehmen, verglichen mit nicht essgestörten Menschen. So wird Hunger meist verleugnet (Adler et al., 2003).

⇒ **Leistungsorientierung.** Ein weiteres Merkmal der Anorexie ist der Kontrollzwang. Viele Patientinnen zeigen eine ausgeprägte Leistungsorientierung, die mit asketischen Idealen einhergeht. Bemühungen um sehr gute Erfolge in Schule oder Beruf können ebenso wie sportliche Aktivitäten dazu dienen, das Selbstwertgefühl zu stabilisieren (Jacobi et al., 2000). Die Leistungsorientierung ist bei bulimischen Patientinnen nicht in diesem Ausmaß vorhanden.

5 Essstörungen

- ⇒ **Compliance.** Die Patientinnen empfinden keinen Leidensdruck und verleugnen die Krankheit und ihre Folgen. Sie sind dementsprechend auch nicht bereit, sich einer Therapie zu unterziehen und zeigen nur geringe Compliance (Pudel & Westenhöfer, 1991). Vielmehr ist es so, dass sich viele mit ihrer skelettartigen Erscheinung identifizieren und darum kämpfen, sich dieses Aussehen zu erhalten (Adler et al., 2003). Manche Patienten entfernen sich selbst ihre Magensonden, wenn eine künstliche Ernährung angeordnet wird.
- ⇒ **Körperliche Veränderungen.** Das Hungern und die Verwendung von Laxantien haben zahlreiche schädliche Auswirkungen auf die körperliche Gesundheit der Betroffenen. Häufig sinken Blutdruck und Herzfrequenz, es kommt zu Nieren- sowie Magen-Darm-Problemen, die Knochenmasse nimmt ab, die Haut trocknet aus und der Hormonhaushalt ändert sich. Beim Binge- Eating/Purging-Typus kann es zu Gewebeerletzungen in Magen und Rachen sowie zu Verlust von Zahnschmelz kommen.
- ⇒ **Sexualität.** Anorektische Patientinnen haben in den meisten Fällen keine intimen sexuellen Kontakte. Sofern solche Kontakte noch bestehen, was insgesamt bei bulimischen Frauen häufiger der Fall ist, ist die sexuelle Erlebnisfähigkeit oft eingeschränkt (Jacobi et al., 2000). In einer Untersuchung an Frauen im durchschnittlichen Alter von 24 Jahren wurde festgestellt, dass 20 Prozent von ihnen noch keinen Geschlechtsverkehr und über die Hälfte entweder keinen Orgasmus oder kaum sexuelle Appetanz hatten (Raboch & Faltus, 1991).

5.2 Bulimia nervosa

- ⇒ **Epidemiologie.** Die Bulimia nervosa setzt normalerweise in den späten Jugendjahren oder im frühen Erwachsenenalter ein. Die Prävalenz in der Bevölkerung wird auf etwa 1 bis 2 Prozent geschätzt (Gotesdam & Agras, 1995). Etwa 90% der Betroffenen sind Frauen. Viele waren vor Beginn der Störung leicht übergewichtig, und die Fressanfälle begannen oft während einer Diät (Davison & Neale, 2002).

Klinische Störungsbilder kompakt

⇒ **Fressanfälle.** Die Fressanfälle finden meist im Geheimen statt, können durch Stress und die damit verbundenen negativen Emotionen ausgelöst werden; sie dauern, bis die Betroffenen sich unangenehm voll fühlen (Grilo et al., 1994). Während eines Anfalls haben die Betroffenen das Gefühl, nicht mehr kontrollieren zu können, wie viel sie essen. Gewöhnlich wählen die Patienten Nahrungsmittel, die sich schnell essen lassen, insbesondere Süßigkeiten wie Eis und Kuchen.

Die Patientinnen empfinden normalerweise Scham und Ekel über ihre Fressanfälle und versuchen, sie zu verbergen. Sie berichten, dass sie während der Anfälle die Kontrolle verlieren, was sogar bis zum Erleben einer Art dissoziativen Zustands gehen kann, bei dem ihnen möglicherweise nicht mehr bewusst ist, was sie tun oder fühlen (Davison & Neale, 2002).

⇒ **Angst vor Gewichtszunahme.** Wenn der Fressanfall vorüber ist, führen Ekel, starkes Unbehagen und Angst vor Gewichtszunahme zum zweiten Schritt, der Entleerung mit dem Ziel, die Kalorienaufnahme wieder rückgängig zu machen. In den meisten Fällen lösen die Betroffenen den Würgereiz aus, um erbrechen zu können. Manche Patientinnen machen auch übermäßigen Gebrauch von Laxantien und Diuretika, fasten und betreiben exzessiv Sport.

Wie die Patientinnen mit Anorexia nervosa haben auch die Patientinnen mit Bulimia nervosa Angst zuzunehmen, und ihre Selbstachtung hängt stark davon ab, ob sie in der Lage sind, ihr normales Gewicht zu halten.

⇒ **Compliance.** Dementsprechend groß ist der Leidensdruck bei diesen Patientinnen. Haben sie die Krankheit einmal offenbart, ist die Bereitschaft zu einer Therapie groß und die Patientinnen befolgen die therapeutischen Anweisungen zuverlässig (Adler et al., 2003).

⇒ **Körperliche Begleiterscheinungen.** Das häufige Erbrechen kann zu Kaliummangel führen und durch den übermäßigen Gebrauch von Laxantien und Diuretika kann es zu Veränderungen im Elektrolythaushalt kommen. Wie schon bei der Anorexia nervosa erwähnt, kann es durch das Erbrechen zu Gewebeverletzungen in Magen und Rachen und einem Verlust an Zahnschmelz kommen.

5 Essstörungen

Tabelle 2 nennt die Kriterien für Bulimia nervosa nach dem DSM-IV.

Tab. 2: Diagnostische Kriterien für Bulimia nervosa nach DSM-IV (Sass et al., 1998)

A	<p>Wiederholte Episoden von Fressattacken. Eine Fressattacken-Episode ist gekennzeichnet durch beide der folgenden Merkmale:</p> <ol style="list-style-type: none"> 1. Verzehr einer Nahrungsmenge in einem bestimmten Zeitraum (z.B. innerhalb eines Zeitraums von 2 Stunden), wobei diese Nahrungsmenge erheblich größer ist, als die Menge, die die meisten Menschen in einem vergleichbaren Zeitraum und unter vergleichbaren Bedingungen essen würden. 2. Das Gefühl, während der Episode die Kontrolle über das Essverhalten zu verlieren (z.B. das Gefühl, weder mit dem Essen aufhören zu können, noch Kontrolle über Art und Menge der Nahrung zu haben).
B	<p>Wiederholte Anwendung von unangemessenen, einer Gewichtszunahme gegensteuernden Maßnahmen, wie z.B. Selbstinduziertes Erbrechen, Missbrauch von Laxantien, Diuretika, Klistieren oder anderen Arzneimitteln, Fasten oder übermäßige körperliche Betätigung.</p>
C	<p>Die Fressattacken und das unangemessene Kompensationsverhalten kommen drei Monate lang im Durchschnitt mindestens zweimal pro Woche vor.</p>
D	<p>Figur und Körpergewicht haben einen übermäßigen Einfluss auf die Selbstbewertung.</p>
E	<p>Die Störung tritt nicht ausschließlich im Verlauf von Episoden einer Anorexia nervosa auf.</p>
	<p>Purging-Typus: Die Person induziert während der aktuellen Episode der Bulimia nervosa regelmäßig Erbrechen oder missbraucht Laxantien, Diuretika oder Klistiere.</p> <p>Nicht-Purging-Typus: Die Person hat während der aktuellen Episode der Bulimia nervosa andere unangemessene, einer Gewichtszunahme gegensteuernde Maßnahmen gezeigt wie beispielsweise Fasten oder übermäßige körperliche Betätigung, hat aber nicht regelmäßig Erbrechen induziert oder Laxantien, Diuretika oder Klistiere missbraucht.</p>

Klinische Störungsbilder kompakt

⇒ **Sexualität.** Wie schon erwähnt haben bulimische Patientinnen eher intime Kontakte als anorektische Patientinnen. Sie versuchen zum Teil durch zunächst betont unkompliziertes, offenes und zugewandtes Verhalten, über Angst- und Insuffizienzgefühle hinwegzutäuschen (Jacobi et al., 2000). Bulimische Patientinnen ziehen sich sozial deutlich weniger zurück als anorektische Patientinnen.

5.3 Komorbiditäten bei Anorexia und Bulimia nervosa

Beide Störungen treten oft zusammen mit anderen Störungen auf, vor allem affektive Störungen, Persönlichkeitsstörungen (insbesondere der Borderline-Störung), Angststörungen, Substanzmissbrauch und Verhaltensstörungen (Davison & Neale, 2002). Aus dem Kreis der affektiven Störungen steht die depressive Verstimmung im Vordergrund. Diese wird durch häufig vorhandenen sozialen Rückzug und Isolierung zusätzlich verstärkt. Kendler und Mitarbeiter konnten eine Lebenszeitprävalenz der Major Depression mit Essstörungen von 51% eruieren (Kendler et al., 1991). Für eine Angststörung beträgt die Prävalenz bei anorektischen Patientinnen zwischen 23% und 66% und bei bulimischen Patientinnen zwischen 25% und 75%. Bei bis zu 40% der bulimischen Patientinnen findet sich außerdem eine Alkoholabhängigkeit (Mitchell et al., 1991).

Sowohl Anorexia nervosa als auch Bulimia nervosa treten in bestimmten Familien gehäuft auf. Bei Verwandten ersten Grades von jungen Frauen mit Anorexia nervosa tritt die Störung etwa viermal so häufig auf wie bei der Allgemeinbevölkerung (Strober et al., 1990). Für die Bulimia nervosa gelten ähnliche Werte (Kassett et al., 1989). Außerdem weisen Verwandte von Patienten mit Essstörungen überdurchschnittlich häufig Symptome von Essstörungen auf, die jedoch nicht alle Diagnosekriterien erfüllen (Lilienfeld et al., 1998). Außerdem wurden in den meisten Studien über Anorexie und Bulimie für eineiige Zwillinge höhere Konkordanzraten ermittelt als für zweieiige (Fichter & Noegel, 1990).

5.4 Diagnostische Verfahren

- ⇒ **Eating Disorder Inventory [EDI].** Das Eating Disorder Inventory ist ein Selbstbeurteilungsinstrument zur Erfassung von essstörungsrelevanten Verhaltensweisen und Einstellungen. Es umfasst verschiedene, für Patientinnen mit Anorexia und Bulimia nervosa typische psychologische Charakteristika. Die acht Skalen des EDI lauten *Schlankheitsstreben, Bulimie, Körperliche Zufriedenheit, Ineffektivität, Perfektionismus, zwischenmenschliches Misstrauen, Interozeption* und *Angst vor dem Erwachsenwerden*. Die neuere Version des EDI-2 ist um drei zusätzliche Skalen, *Askese, Impulsregulation* und *soziale Unsicherheit* ergänzt worden (Jacobi et al., 2000).
- ⇒ **Strukturiertes Interview für Anorektische und Bulimische Essstörungen [SIAB].** Mit dem *Strukturierten Interview für Anorektische und Bulimische Essstörungen* wurde ein deutschsprachiges Instrument für die Untersuchung von Patienten mit psychogenen Essstörungen vorgelegt. Es besteht aus einem Selbsteinschätzungsfragebogen für die Patientin sowie einem Interviewteil für den Untersucher und beinhaltet die Diagnosekriterien des DSM-IV und der ICD-10. Die jeweils 87 Fragen des Interviews bzw. des Fragebogens berücksichtigen neben der typischen anorektischen und bulimischen Symptomatik auch Depressionen, Ängste, Zwänge sowie andere, für eine klinisch valide Untersuchung relevante Symptombereiche (Fichter & Quadflieg, 1999).

Klinische Störungsbilder kompakt

Literatur

- Adler, R.H., Hermann, J.M., Köhle, K., Langewitz, W., Schonecke, O.W., von Uexküll, T. & Wesiack, W. (2003). *Psychosomatische Medizin* (6.Auflage). München: Urban & Fischer.
- Davison, G.C. & Neale, J.M. (2002). *Klinische Psychologie* (6. Auflage). Weinheim: Beltz.
- Fichter, M.M. & Noegel, R. (1990). Concordance for bulimia nervosa in twins. *International Journal of Eating Disorders*, 9, 425-436.
- Fichter, M.M. & Quadflieg, N. (1999). *Strukturiertes Inventar für Anorektische und Bulimische Essstörungen (SIAB). Fragebogen (SIAB-S) und Interview (SIAB-EX) nach DSM-IV und ICD-10 Handanweisung*. Göttingen: Hogrefe
- Grilo, C.M., Shiffman, S. & Carter-Campbell, J.T. (1994). Binge eating antecedents in normal weight nonpurging females: Is there Consistency?. *International Journal of Eating Disorders*, 16, 239-249.
- Gotesdam, K.G. & Agras, W.S. (1995). General population-based epidemiological survey of eating disorders in Norway. *International Journal of Eating Disorders*, 18, 119-126.
- Hsu, L.K.G. (1990). *Eating disorders*. New York: Guilford.
- Jacobi, C., Thiel, A. & Paul, T. (2002). *Kognitive Verhaltenstherapie bei Anorexia nervosa und Bulimia nervosa*. Weinheim: Verlagsgruppe Beltz.
- Kassett, J.A., Gershon, E.S. & Maxwell, M.E. (1989). Psychiatric disorders in the first-degree relatives of probands with bulimic disorders. *American Journal of Psychiatry*, 146, 1468-1471.
- Kendler, K.S., MacLean, C., Neale, M., Kessler, R., Heath, A. & Eaves, L. (1991). The genetic epidemiology of bulimia nervosa. *American Journal of Psychiatry*, 148, 1627-1637.
- Mitchell, J., Specker, S. & de Zwaan, M. (1991). Comorbidity and medical complications in bulimia nervosa. *Journal of clinical psychiatry*, 52, 13-21.
- Paul, T. & Thiel, A. (2005). *Eating Disorder Inventory-2*. Deutsche Version. Göttingen: Hogrefe.
- Pudel, V. & Westenhöfer, J. (1991). *Ernährungspsychologie*. Eine Einführung. Göttingen: Hogrefe.
- Raboch, J. & Faltus, F. (1991). Sexuality of Woman with Anorexia nervosa. *Acta Psychiatrica Scandinavica*, 84, 9-11.
- Sass, H., Wittchen, H.U. & Zaudig, M. (1998). *Diagnostisches und Statistisches Manual Psychischer Störungen. DSM-IV*. Göttingen: Hogrefe.
- Strober, M., Lampert, C., Morrell, W., Burroughs, J. & Jacobs, C. (1990). A controlled family study of anorexia nervosa: Evidence of family aggregation and lack of shared transmission with affective disorders. *International Journal of Eating Disorders*, 9, 239-253.
- Wittek, N. (2003). *Untersuchungen zur Körperschemastörung bei Anorexia nervosa*. Unveröffentlichte Dissertation, Medizinische Fakultät der Bayrischen Julius-Maximilians-Universität zu Würzburg.

Weiterführende Literatur

- Academy for Eating Disorders (2005). *Eating disorders review*. Oxford: Radcliffe Publ.
- Buddeberg-Fischer, B. (2000). *Früherkennung und Prävention von Essstörungen. Essverhalten und Körpererleben von Jugendlichen*. Stuttgart: Schattauer.
- Grilo, M.G. (2006). *Eating and weight disorders*. Hove: Psychology Press.
- Herzof, W., Munz, D. & Kächele, H. (2004). *Essstörungen: Therapieführer und psychodynamisches Behandlungskonzept*. Stuttgart: Schattauer.
- Schweiger, U., Peters, A. & Skipos, V. (2003). *Essstörungen*. Stuttgart: Thieme.
- Stahr, I., Barb-Priebe, I. & Schulz, E. (2007). *Essstörungen und die Suche nach Identität: Ursachen, Entwicklung und Behandlungsmöglichkeiten*. Weinheim: Juventa Verlag.

Materialien zur Eignungsprüfung der Fallbeschreibung

Christian-Albrechts-Universität zu Kiel
INSTITUT FÜR PSYCHOLOGIE
Differentielle & Diagnostische Psychologie
Dipl.-Psych. Britta Dreger



Liebe/r Teilnehmer/in,

wir planen derzeit eine Studie zur diagnostischen Urteilsbildung im Kontext von klinischen Störungsbildern. Dazu haben wir eine Fallgeschichte einer jungen Patientin konstruiert und wollen nun vor Studienbeginn die Eigenschaften der verwendeten Informationen überprüfen.

Wir wären Ihnen sehr dankbar, wenn Sie sich 15-20 Minuten Zeit nehmen würden, um uns bei diesem Vorhaben zu unterstützen und damit einen wertvollen Beitrag für die Forschung zu leisten.

Sie bleiben in dieser Befragung vollkommen anonym, denn wir fragen Sie nirgends nach Ihrem Namen, sondern bitten Sie lediglich um einige allgemeine Angaben zu Ihrer Person:

Wie alt sind Sie?

..... Jahre

Ihr Geschlecht:

- weiblich
 männlich

Sind Sie Studierende/r?

- ja
 nein

Sind Sie bereits Diplom-Psychologe/in?

- ja
 nein
-

Vielen Dank für Ihre bisherigen Antworten! Sie erhalten nun die Fallinformationen. Bitte lesen Sie sich diese aufmerksam durch und beantworten Sie die Fragen auf den folgenden beiden Seiten!

Fallinformationen

Marisa (18;6) wurde von ihrer Mutter in der Ambulanz einer Psychosomatischen Klinik für Jugendliche und Junge Erwachsene vorgestellt. Die Mutter von Marisa habe sich zunächst an den Hausarzt der Tochter gewandt, da sie wegen des von ihr beobachteten sozialen und emotionalen Rückzugs der Tochter, deren niedergeschlagenen Stimmung und den Problemen mit dem Essverhalten besorgt gewesen sei. Diesbezüglich habe die Mutter beobachtet, dass Marisa in den letzten Wochen immer wieder zu- und abgenommen habe, wobei die Tochter sich selbst nicht als zu dünn, sondern eher als zu dick einschätze. Der Arzt habe Auffälligkeiten im Elektrolythaushalt von Marisa festgestellt und die Vorstellung in einer Klinik empfohlen, da er den Verdacht hatte, dass eine Bulimie-Erkrankung vorliegt. In der Klinik wurde ein Erstgespräch geführt, nach dem Marisa stationär in der Klinik aufgenommen wurde.

Im Rahmen des Aufnahmegesprächs habe die Patientin nach Aussage des Interviewers teilweise abwesend gewirkt, den Blickkontakt gemieden und sich dem Gespräch entzogen. Ihre einzige Einlassung habe darin bestanden, zu äußern, dass sie durchaus mit Genuss essen wolle. Die folgenden Informationen stammen daher von der Mutter.

Marisa lebe mit ihrer jüngeren Schwester, mit der sie sich gut verstehe, bei den Eltern. Die Mutter sei Hausfrau, der Vater verdiene den Lebensunterhalt der Familie als Beamter. Die Mutter berichtete, dass die Schwangerschaft und die Geburt von Marisa unkompliziert verlaufen seien. Außerdem habe es keine Auffälligkeiten in der frühkindlichen Entwicklung gegeben und Marisa sei nicht ernsthaft krank gewesen. Gefragt nach der Familienanamnese, gab die Mutter an, dass sie selbst wegen depressiver Episoden und Beginn einer adipösen Entwicklung in psychotherapeutischer Behandlung gewesen sei. Ansonsten gebe es keine psychischen und medizinischen Auffälligkeiten in der Familie. Über den schulischen Werdegang von Marisa berichtete die Mutter, dass sie regelgerecht eingeschult worden sei und heute die 13. Klasse des Gymnasiums besuche. Sie sei immer gut in die Klassengemeinschaft integriert gewesen und habe durchweg gute bis sehr gute Leistungen erzielt. Sie erledige ihre Hausaufgaben sehr gewissenhaft und wende viel Zeit für die Schule auf, wenn sie nicht gerade einer ihrer vielen sportlichen Aktivitäten nachgehe. Dazu gehörten bspw. das morgendliche Joggen sowie das Schwimmtraining.

Bezüglich der aktuellen Problematik gab die Mutter an durch den sozialen Rückzug von Marisa beunruhigt zu sein. Sie beschrieb, dass Marisa Freundschaften vernachlässige, z.B. vermehrt Verabredungen ablehne. Auch aus dem Familienalltag ziehe Marisa sich mehr und mehr zurück. So habe Marisa keinerlei Interesse mehr an Familienfeiern sowie gemeinsamen Unternehmungen und suche nach Ausreden, um nicht teilnehmen zu müssen.

Am meisten missfalle der Mutter, dass Marisa sich auch häufig dem gemeinsamen Essen der Familie zu entziehen versuche. Mit den Argumenten, sie habe schon gegessen oder habe keinen Hunger, versuche Marisa, das gemeinsame Beisammensein zu vermeiden. Wenn die Eltern sie dazu zwingen, dann setze Marisa sich zwar zu ihnen und esse auch etwas, allerdings stünde sie dann auch so schnell wie möglich wieder auf.

Auch die wechselhaften Stimmungen der Tochter machten der Mutter Sorgen. Teilweise sei Marisa sehr reizbar und werde schnell wütend, so dass es zu heftigen Streitereien zwischen ihr und den Eltern komme. Andererseits habe die Mutter beobachtet, dass Marisa auch häufig niedergeschlagen wirke. Neben dem Desinteresse an gemeinsamen Aktivitäten, mache sie häufig den Eindruck, traurig zu sein. Einem Gespräch über ihre Gefühle entziehe sie sich.

Bitte beantworten Sie nun folgende Fragen:

Im Folgenden sind die Störungen aufgeführt, die in der Einrichtung, in die Marisa aufgenommen wurde, schwerpunktmäßig behandelt werden. Bitte geben Sie jeweils an, für wie wahrscheinlich Sie das Vorliegen der genannten Störung bei der konkreten Patientin aufgrund der Fallinformationen halten.

Kreuzen Sie bitte eine Zahl an und machen Sie kein Kreuz zwischen die Ziffern!

<u>Störungsbilder</u>	Sehr unwahrscheinlich						Sehr wahrscheinlich	
Depression	0	1	2	3	4	5	6	
Bulimie	0	1	2	3	4	5	6	
Angststörung	0	1	2	3	4	5	6	
Burn-Out	0	1	2	3	4	5	6	
Anorexie	0	1	2	3	4	5	6	

Für das weitere Vorgehen in der Klinik wäre es nun aus verwaltungstechnischer Sicht wichtig, dass Sie sich für eine Störung entscheiden, die Sie zunächst weiter abklären möchten, d.h. dass Sie zum jetzigen Zeitpunkt eine Verdachtsdiagnose stellen. Bitte geben Sie daher an welche Störung Ihrer Meinung nach am wahrscheinlichsten vorliegt. Treffen Sie Ihre Entscheidung aufgrund der vorliegenden Fallinformationen.

Ich stelle folgende Verdachtsdiagnose: _____

<u>Sicherheit</u>	Gar nicht sicher						Sehr sicher	
Wie sicher sind Sie sich bei Ihrer Entscheidung?	0	1	2	3	4	5	6	

Bitte beantworten Sie auch die folgenden Fragen zu den Fallinformationen:

<u>Lesbarkeit</u>	Sehr schlecht						Sehr gut	
	lesbar						lesbar	
Wie schätzen Sie die Lesbarkeit des Textes ein?	0	1	2	3	4	5	6	

<u>Plausibilität</u>	Gar nicht						Sehr	
	plausibel						plausibel	
Wie plausibel finden Sie die Informationen?	0	1	2	3	4	5	6	

Unabhängig von Ihrer Verdachtsdiagnose: Wie schätzen Sie die Fallinformationen ein?

<u>Bulimie</u>	Gar nicht						Sehr	
	charakteristisch						charakteristisch	
Wie charakteristisch sind die Informationen für eine Bulimie-Erkrankung?	0	1	2	3	4	5	6	

<u>Anorexie</u>	Gar nicht						Sehr	
	charakteristisch						charakteristisch	
Wie charakteristisch sind die Informationen für eine Anorexie-Erkrankung?	0	1	2	3	4	5	6	

Ist Ihnen beim Lesen des Textes noch etwas aufgefallen?

Vielen herzlichen Dank für Ihre Mithilfe!

Konstruktion der Zusatzinformationen_Instruktionen

Liebe Rater!

Ich benötige zu meinem Studienmaterial einige Einschätzungen von unabhängigen und hypothesenblinden Ratern zur Validierung. Es existieren einerseits 39 Informationen, die aus einer Studie im Zusammenhang mit Essstörungen stammen und meinen eigentlichen VPn vorgelegt werden, die das Vorliegen einer Bulimie bei der jungen Patientin überprüfen sollen. Andererseits existieren 42 Informationen, die für die Schuld oder Unschuld eines Angeklagten in einem Brandstiftungsfall herangezogen werden. Es liegen daher insgesamt 81 Items vor, die in Bezug auf 5 Merkmale einzuschätzen sind, die im Folgenden erläutert werden.

Konsistenz. Auf einer Skala von *-5=sehr inkonsistent* über *0=neutral/irrelevant* bis *+5=sehr konsistent* soll eingeschätzt werden, wie konsistent, d.h. konform die Information mit einer bestimmten Annahme ist. Im Essstörungsfall geht es um die Frage, wie konsistent die Angaben im Falle einer vorliegenden Bulimie wären, d.h. ob sie zu dem typischen Krankheitsbild passen (konsistent) oder eigentlich nicht erwartet werden, wenn die Störung vorliegt (inkonsistent). Wenn die Informationen in dieser Hinsicht neutral sind oder gar nichts mit der Essstörung zu tun haben, soll die null angekreuzt werden. Im Brandstiftungsfall geht es analog um die Frage, ob man das Vorliegen der Information erwarten würde, wenn der Angeklagte schuldig ist (konsistent) oder gerade nicht (inkonsistent).

Wichtigkeit. Auf einer Skala von *0=gar nicht* bis *100=sehr wichtig* soll die Wichtigkeit des Items eingeschätzt werden. Es geht bei der Beurteilung darum, wie viel Nutzen man aus den einzelnen Informationen ziehen kann, um zu entscheiden, ob jemand unter einer Bulimie leidet bzw. einer Brandstiftung schuldig ist. Wenn eine Information kann dazu keinen Beitrag leisten, ist sie unwichtig, sonst je nach Nutzen unterschiedlich wichtig. Es soll bei dieser Einschätzung unterstellt werden, dass die gegebenen Informationen tatsächlich wahr sind, d.h. eine vollkommene Zuverlässigkeit aufweisen (s.u.)

Zuverlässigkeit. Auf einer Skala von *0=gar nicht* bis *100=vollkommen* soll die Zuverlässigkeit der Information eingeschätzt werden. Damit ist gemeint, wie fehlerfrei und damit reliable ein Item ist. Eine Information kann dabei auf unterschiedliche Weise fehlerbehaftet sein, bspw. weil es sich um eine subjektive Einschätzung handelt, weil man etwas nur durch Hören-Sagen kennt, weil man eine Motivation hat, zu lügen. Es geht daher letztendlich darum, wie viel Glauben man der Information schenken möchte.

Diagnostizität. Auf einer Skala von *0=gar nicht* bis *100= sehr diagnostisch* ist das Ausmaß der Diagnostizität des Items einzuschätzen. Ein Merkmal ist dann diagnostisch, wenn es unter der

angenommenen Hypothesen (entweder Bulimie oder Schuld des Angeklagten) eine andere Auftretenswahrscheinlichkeit hat als unter der Alternative (entweder Anorexie oder Unschuld). Ist es bspw. einerseits wahrscheinlich, dass eine extravertierte Person häufig auf Partys auf dem Tisch tanzt, und gleichzeitig andererseits unwahrscheinlich, dass eine introvertierte Person dies tut, so ist das Merkmal „auf dem Tisch tanzen“ diagnostisch. Das Merkmal „Bücher lesen“ hingegen ist nicht diagnostisch, denn dass ein Introvertierter zu einem Buch greift, ist genauso wahrscheinlich wie bei einem Extravertierten. Man muss sich also immer fragen, ob das genannte Merkmal unter der einen Hypothese genauso wahrscheinlich ist wie unter der anderen. Je unterschiedlicher die Wahrscheinlichkeiten sind, desto diagnostischer sind die Items.

Relevanz. Hinsichtlich des fünften Merkmals ist eine kategoriale Einschätzung abzugeben, d.h. die Information ist einer von 4 Merkmalsklassen zuzuordnen: eindeutiger Beleg, brauchbarer Hinweis, Widerspruch, irrelevante Information.

Ein „**eindeutiger Beleg**“ zeichnet sich dadurch aus, dass die betreffende Angabe vorliegen muss, damit das Urteil gerechtfertigt ist. Der Befund „*Person A erreicht auf der Skala Extraversion ein überdurchschnittliches Ergebnis.*“ kann bspw. einen eindeutigen Beleg darstellen. Im Fall des Essstörungskontexts ist die z.B. immer dann gegeben, wenn die Information ein Merkmal enthält, das in den diagnostischen Kriterien genannt wird.

Ein „**brauchbarer Hinweis**“ zeichnet sich dadurch aus, dass die betreffende Sachlage/ die betreffende Aussage vorliegen kann, wenn das Urteil gerechtfertigt ist, obwohl es nicht ausreicht, dass diese Information vorliegt. Es kommt gehäuft vor, jedoch nicht notwendigerweise. Die Information „*Person A wurde auf einer Party gesehen*“ wäre in diesem Zusammenhang ein Hinweis.

Ein **Widerspruch** zeichnet sich dadurch aus, dass die betreffende Angabe den Vorgaben für das Urteil widerspricht bzw. gewöhnlich nicht vorliegt. Die Beobachtung „*Person A kam nicht zur Party, um zuhause ein Buch zu lesen*“ wäre ein Widerspruch für Extraversion.

Eine **irrelevante Information** bezieht keinerlei Zusammenhang mit dem Urteil – ob sie vorliegt oder nicht, hat keinen Einfluss. *Dass Person A am liebsten Erdbeereis mag, ist irrelevant* für Extraversion.

Zunächst bitte ich darum, die Items im Zusammenhang mit der Essstörung zu bewerten. Dies geschieht pro Merkmal immer blockweise, d.h. erst werden alle Items präsentiert, um sie hinsichtlich der Konsistenz einzuschätzen. Dann werden die Items wiederholt und hinsichtlich des Informationsgehalts bewertet etc. Im Anschluss daran geschieht dasselbe für die

Brandstiftungsinformationen. Aufgrund der häufigen Wiederholungen der Items entsteht die große Anzahl an auszufüllenden Seiten. Ich hoffe jedoch, durch das blockweise Arbeiten das Kriterium, dass ihr jeweils für die Einschätzung anlegt, konstant halten zu können. Haltet euch also bitte unbedingt an die blockweise Darstellung und macht nur zwischen den verschiedenen Merkmalen eine Pause; möglichst nicht innerhalb eines Merkmalsblocks.

Es geht bei den Bewertungen um eure subjektive Einschätzung. Ein richtig oder falsch gibt es nicht, alleine eure Bewertung des Items zählt. Ihr erfüllt den Zweck des Ratings also am besten, indem ihr so spontan wie möglich eure persönliche Einschätzung abgebt. Dennoch ist es im Kontext der Essstörung erforderlich, dass ihr euch vor dem Rating (neben der Fallbeschreibung) den beigefügten Informationstext durchlest und im Zweifelsfall immer wieder zu Rate zieht. Er liefert euch das Fachwissen, das für die Entscheidung notwendig ist. Ein vergleichbares Fachwissen ist im Brandstiftungsfall nicht relevant, allerdings ist auch hier ein genaues Studierendes der Fallbeschreibung wichtig, da viele Personen an der Geschichte beteiligt sind und es essentiell ist, dass ihr bei den Bewertungen immer genau wisst, wer welche Rolle spielt.

Hinweis: Die Items sind nicht der Reihenfolge nach nummeriert, das hat für euch keine Bedeutung!

Zur Bewertung setzt ihr jeweils ein in die Spalte, die eure Meinung am besten wiedergibt!

Konstruktion der Zusatzinformationen_Items und Kennwerte

Tabelle A-1. Überblick über die 39 Items sowie deren Kennwerten aus dem gemeinsamen Rating

Nr.	Item	Konsistenz			Wichtigkeit		Zuverlässigkeit		Diagnostizität		Relevanz
		M (SD)	t'	p	d	M (SD)	M (SD)	M (SD)	M (SD)	Code ²	
1	Die Patientin gab an, dass sie in ihrer Freizeit auf einem Ponyhof mithelfe.	0.56 (0.88)	1.89	.05	0.64	4.44 (7.27)	76.67 (24.50)	12.22 (21.10)	I	8/9	
2	Im SIAB danach befragt, gab die Patientin zu, ab und zu unter ihrem Essverhalten zu leiden.	3.11 (1.27)	7.35	.00	2.45	65.56 (30.87)	75.56 (12.36)	54.44 (38.12)	H	8/9	
3	Auf die Frage nach ihren Plänen für die Zeit nach dem Abitur, gab sie an, studieren wolle.	0.00 (0.00)	---	---	---	5.56 (8.82)	78.89 (23.15)	3.33 (7.07)	I	9/9	
4	Die Mutter hatte Gelegenheit, ihre Tochter heimlich bei einem Heißhungeranfall zu beobachten und gewann den Eindruck, ihre Tochter habe nicht mehr mit dem Essen aufhören können.	4.44 (1.13)	11.80	.00	3.93	85.56 (23.51)	58.89 (24.21)	46.67 (45.55)	-	-	
5	Die Patientin bestätigte, vor allem nicht zunehmen zu wollen. Es wäre für sie nicht so schwierig, wenn sie in Zukunft nicht stark abnehmen würde.	3.56 (1.24)	8.63	.00	2.87	66.67 (23.98)	61.11 (20.28)	54.44 (32.83)	H	7/9	
6	Die Patientin erklärte, keine Diät zu halten, um dem Zunehmen entgegen zu wirken.	-0.56 (3.05)	-0.55	.60	0.18	40.00 (29.58)	28.89 (20.28)	47.78 (38.01)	W	5/9	
7	Mit dem <i>subjective body dimensions apparatus</i> wird die Einschätzung des Ausmaßes einzelner Körperteile erfasst. Die Patientin schätzte alle zur Beurteilung vorgegebenen Körperteile (Bauch, Arme und Beine) weitaus größer ein, als sie objektiv waren.	-2.78 (2.54)	-3.28	.01	1.09	72.22 (27.29)	84.44 (13.33)	73.33 (29.58)	W	6/9	
8	Die Patientin gab an, dass sie zur Entspannung am ehesten zu einem guten Buch greife.	0.00 (0.00)	---	---	---	2.22 (4.41)	87.78 (22.79)	1.11 (3.33)	I	9/9	
9	Ihre Mutter berichtete, dass Marisa sich nicht bewusst sei, dass sie ihrem Körper schaden könnte.	0.44 (1.88)	0.71	.25	0.23	36.67 (34.28)	43.33 (25.00)	20.00 (29.16)	---	-	

Anmerkungen.

¹ Die Anzahl der Freiheitsgrade beträgt jeweils $df = 8$ ² Code: I= irrelevante Information; H= brauchbarer Hinweis; W: Widerspruch; B= eindeutiger Beleg; - = keine Mehrheit; Ü = Übereinstimmung von X/9 Ratern

Nr.	Item	Konsistenz			Wichtigkeit		Zuverlässigkeit		Diagnostizität		Relevanz
		M (SD)	t ¹	p	d	M (SD)	M (SD)	M (SD)	M (SD)	Code ²	
10	Die Patientin führte aus, ihren Schwächen vornehmlich im künstlerisch-kreativen Bereich zu sehen.	0.00 (0.00)	---	---	---	0.00 (0.00)	68.89 (28.04)	3.33 (7.07)	I	9/9	
11	Das Ergebnis des SIAB machte deutlich, dass die Patientin keine Krankheitseinsicht hat.	-1.78 (2.68)	-1.99	.04	0.66	58.89 (35.16)	81.11 (9.28)	47.78 (32.32)	W	6/9	
12	Die Antworten im EDI ergaben, dass die Patientin keine Schuldgefühle hat, wenn sie eine große Menge Nahrung zu sich genommen hat.	-3.67 (1.23)	-8.98	.00	2.98	62.22 (29.91)	67.78 (22.24)	22.22 (25.39)	W	9/9	
13	Im SIAB gab Marisa an, dass ihr Selbstvertrauen von ihrer Figur abhängt.	4.22 (0.97)	13.03	.00	4.35	75.56 (32.83)	84.44 (14.24)	11.11 (11.67)	--	-	
14	Ihre Mutter berichtete, dass Marisa als Kind große Angst vor Hunden gehabt habe	0.11 (0.33)	1.00	.17	0.33	0.00 (0.00)	77.50 (28.66)	1.11 (3.33)	I	9/9	
15	Marisas Zahnarzt konnte keine Veränderungen am Zahnschmelz feststellen.	-2.00 (1.87)	-3.21	.01	1.07	63.33 (24.50)	93.33 (13.23)	25.56 (33.95)	W	8/9	
16	Die Patientin war nach Ansicht des Bezugstherapeuten durchaus zu einer Mitarbeit in der Therapie bereit.	2.56 (1.24)	6.20	.00	2.06	57.50 (27.12)	78.89 (23.69)	57.78 (29.91)	H	7/9	
17	Gebeten, eine ihrer Stärken zu nennen, gibt sie an, gut zuhören zu können.	0.11 (0.34)	1.00	.17	0.32	0.00 (0.00)	75.56 (26.03)	1.11 (3.33)	I	9/9	
18	Die Patientin sagte, dass ihr Selbstwertgefühl von ihrem Körpergewicht abhängig sei.	3.89 (1.05)	11.07	.00	3.70	73.33 (31.62)	84.44 (26.03)	8.89 (10.54)	B	5/9	
19	Die Patientin hatte kein Normalgewicht, denn bei einer Körpergröße von 1,75m und einem Gewicht von 50 kg lag ihr BMI bei 16,3.	-3.33 (0.87)	-11.55	.00	3.83	77.78 (27.74)	100.00 (0.00)	65.56 (40.35)	W	9/9	
20	Im CFT-20-R zeigte sich eine überdurchschnittliche Intelligenz.	0.00 (0.00)	-	-	-	5.56 (8.82)	92.22 (8.33)	3.33 (10.00)	I	9/9	

*Anmerkungen.*¹ Die Anzahl der Freiheitsgrade beträgt jeweils $df = 8$ ² Code: I= irrelevante Information; H= brauchbarer Hinweis; W: Widerspruch; B= eindeutiger Beleg; - = keine Mehrheit; Ü = Übereinstimmung von X/9 Ratern

Nr.	Item	Konsistenz			Wichtigkeit		Zuverlässigkeit		Diagnostizität		Relevanz
		M (SD)	t'	p	d	M (SD)	M (SD)	M (SD)	M (SD)	Code ²	
21	Laut Mutter tritt bei der Patientin Angst auf, weiter zuzunehmen, wenn sie ein Pfund zugenommen habe.	3.11 (1.36)	6.84	.00	2.29	60.00 (25.00)	63.33 (16.58)	30.44	H	5/9	
22	Die Patientin wurde sehr ungehalten, als sie befürchtete, ihr Tagebuch auf der Station verloren zu haben. Gefragt danach, über welches eigene Verhalten sie sich am meisten ärgere, berichtete sie, dass sie ein Angebot über einen Auslandsaufenthalt abgelehnt habe, um bei ihrem Freund zu bleiben.	0.00 (0.00)	-	-	-	7.78 (10.93)	75.56 (25.55)	4.44 (7.27)	I	9/9	
23	Mit der Patientin wurde ein Computer Body Image Test durchgeführt. Bei diesem Verfahren können die Patienten am Computer bei einem Foto von sich selbst die Körperdimensionen so verändern, dass es ihrem eigenen Bild von sich selbst entspricht. Die Patientin schätzte ihre Körperdimensionen weitaus größer ein, als sie sind. Das heißt, dass sich eine bedeutsame Differenz zwischen den wahrgenommenen und den tatsächlichen Proportionen.	0.11 (0.33)	1.00	.35	0.33	2.22 (6.67)	66.67 (28.72)	2.22 (4.41)	I	5/9	
24	Die Patientin war nach Angabe der Kunsttherapeutin nicht bereit, in der Therapie mitzuarbeiten.	-2.67 (2.35)	-3.41	.04	1.14	71.11 (28.92)	83.33 (11.18)	73.33 (34.64)	W	6/9	
25	Die Patientin war unzufrieden mit ihrem Essverhalten.	2.78 (1.79)	4.66	.00	1.55	58.89 (34.44)	70.00 (31.23)	38.89 (33.71)	H	8/9	
26	Die Patientin glaubte nicht, ihrem Körper durch ihr Essverhalten zu schaden.	-1.33 (1.00)	-4.00	.00	1.33	33.33 (29.58)	80.00 (22.36)	26.67 (31.62)	W	6/9	
27	Über die Geschwisterbeziehung sagte die Mutter, dass die beiden die üblichen Streitigkeiten untereinander hätten, aber sich in jugendtypischer Weise gegen die Eltern verbündeten.	-0.33 (2.06)	-0.49	.32	0.16	43.33 (30.41)	53.33 (29.16)	34.44 (31.27)	W	5/9	
28		0.00 (0.00)	-	-	-	0.00 (0.00)	67.78 (26.35)	1.11 (3.33)	I	9/9	

Anmerkungen.

¹ Die Anzahl der Freiheitsgrade beträgt jeweils $df = 8$

² Code: I= irrelevante Information; H= brauchbarer Hinweis; W: Widerspruch; B= eindeutiger Beleg; - = keine Mehrheit; Ü = Übereinstimmung X/9 Ratern

Nr.	Item	Konsistenz			Wichtigkeit		Zuverlässigkeit		Diagnostizität		Relevanz	
		M (SD)	t'	p	d	M (SD)	M (SD)	M (SD)	M (SD)	Code ²	Ü ²	
29	Große Mengen an Nahrung nahm die Patientin nur dann zu sich, wenn sie alleine war.	3.44 (1.59)	6.50	.00	2.16	71.11 (28.92)	68.89 (25.22)	51.11 (46.49)	H	6/9		
30	Die Patientin war gedanklich häufig mit Diäten beschäftigt.	2.11 (2.32)	2.74	.03	0.91	61.11 (32.19)	67.78 (26.35)	15.56 (14.24)	H	7/9		
31	Die Patientin sagte, dass sie einen Freund habe, mit dem sie auch intime sexuelle Kontakte habe	2.44 (1.33)	5.50	.00	1.83	52.22 (29.06)	57.78 (29.06)	63.33 (23.98)	H	7/9		
32	Sie sagte außerdem, dass sie sich häufig mit dem Thema Essen, Kochen und Kaloriengehalt beschäftigt.	1.00 (2.83)	1.06	.16	0.35	61.11 (31.40)	71.11 (22.05)	38.89 (34.80)	H	5/9		
33	Ihre Schwester gab an, dass Marisa leicht reizbar sei.	0.11 (1.17)	0.29	.39	0.09	20.00 (18.71)	50.00 (30.00)	7.78 (9.72)	I	9/9		
34	Die Mutter berichtete, dass Marisa Klassensprecherin gewesen sei.	0.56 (0.73)	2.29	.03	0.76	12.22 (14.81)	86.67 (17.32)	6.67 (10.00)	I	7/9		
35	Die Patientin gab an, ihr sexuelles Bedürfnis und Verlangen als nicht normal wahrzunehmen.	0.44 (1.24)	1.08	.16	0.35	38.89 (27.13)	65.56 (29.20)	26.67 (27.39)	---	---		
36	Im EDI wurde deutlich, dass die Patientin das Gefühl hatte, dass in der Familie nicht jede, sondern nur Bestleistungen gut genug seien.	-0.67 (1.66)	-1.21	.13	0.40	31.25 (22.32)	78.89 (19.65)	12.22 (17.87)	I	5/9		
37	Die Mutter sagte, dass ihre Tochter noch keine intimen sexuellen Kontakte gehabt habe.	-0.56 (0.88)	-1.89	.05	0.64	24.44 (24.04)	17.78 (18.56)	13.33 (22.36)	I	5/9		
38	Gefragt nach der Aussage der Mutter, sie vernachlässige ihre Freundschaften, gab die Patientin an, sie habe viel Zeit mit ihrem Freund verbracht.	0.33 (1.80)	0.55	.30	0.18	25.56 (18.11)	44.44 (16.67)	31.11 (28.48)	I	6/9		
39	Die Patientin erklärte, dass sie sich selbst ablehnen würde, wenn sie zu dick wäre.	3.56 (1.24)	8.63	.00	2.87	85.56 (23.51)	83.33 (18.03)	8.89 (10.54)	H	5/9		

Anmerkungen.

¹ Die Anzahl der Freiheitsgrade beträgt jeweils $df = 8$

² Code: I= irrelevante Information; H= brauchbarer Hinweis; W: Widerspruch; B= eindeutiger Beleg; - = keine Mehrheit; Ü = Übereinstimmung von X /9 Ratern
Zusatzanmerkung. Die Items 1-18 wurden im Zuge der AV 4 (Bestätigungstendenz B) eingesetzt. Die Items 19-30 für die AV 5 a+b (Bestätigungstendenz I sowie hypothesenkonstante Interpretation uneindeutiger Evidenz). Die Items 31-39 wurden nicht verwendet.

Anhang B - Fallmaterialien Brandstiftung

Fallbeschreibung Fachwissen irrelevant (Brandstiftung)	241
Hintergrundwissen Fachwissen irrelevant (Brandstiftung).....	245
Materialien zur Eignungsprüfung der Fallbeschreibung.....	255
Konstruktion der Zusatzinformationen_Instruktionen.....	261
Konstruktion der Zusatzinformationen_Items und Kennwerte	265

Fallbeschreibung Fachwissen irrelevant (Brandstiftung)

Fallinformationen

In der Tatnacht zum 30. Juni 2006 um 4 Uhr 15 morgens brach im zweiten Stock eines Wohnhauses in Mannheim ein Feuer aus. Die zweijährige Katharina, die sich zu diesem Zeitpunkt alleine in der Wohnung aufgehalten hatte, starb an einer Rauchvergiftung. Der 22-jährige Peter Wagner wurde bald darauf als dringend tatverdächtig festgenommen und wegen Brandstiftung mit Todesfolge in Tateinheit mit Einbruch angeklagt.

Um die Transparenz der nachfolgenden Schilderungen vor, während und nach dem Tatgeschehen zu gewährleisten, soll die folgende Abbildung 1 die Wohnsituation der beteiligten Personen veranschaulichen.



Abb 1: Wohnsituation und beteiligte Personen.

Frau Schröder zog am 15. Juni 2006 in das Gebäude in dem der Brand später stattfand. Zu diesem Zeitpunkt lernte sie Herrn Wagner kennen, mit dem sie sehr schnell eine sexuelle Beziehung einging. Auch mit einem anderen Nachbarn, Herrn Braun, hatte sie Ende Juni eine einmalige sexuelle Begegnung, bevor sie Herrn Lange kennenlernte.

Am Sonntag den 29. Juni, also am Abend der Tatnacht, nahm Frau Schröder ihren neuen Freund Herrn Lange mit zu einer Feier in der nahegelegenen Wohnung von Frau König; einige Nachbarn, darunter auch Herr Wagner und Frau Bauer, nahmen an dieser Party teil. Während dieser Feier küsste Frau Schröder Herrn Lange öffentlich und erklärte Herrn Wagner, sie wolle nun mit Herrn Lange zusammen sein. Um 1 Uhr morgens verließen Frau Schröder und Herr Lange gemeinsam die Feier und gingen zu Frau Schröders Wohnung.

Drei Zeugen, die ebenfalls auf der Feier waren, berichteten später, Herr Wagner habe die Absicht geäußert, er wolle den Wohnbereich [in dem auch Frau Schröder lebte] niederbrennen. Alle drei Zeugen erklärten, Herr Wagner sei sehr wütend gewesen und mindestens ein Zeuge glaubte, diese Wut liege an Frau Schröders neuer Beziehung.

Ein weiterer Zeuge gab an, dass Herr Wagner gesagt habe: „Wenn ich sie [Frau Schröder] nicht haben kann, kann es keiner.“ Frau König, die Gastgeberin dieser Feier, konnte sich erinnern, Herr Wagner habe ihr gegenüber geäußert: „Noch bevor die Nacht vorbei ist, wird eine Wohnung [des betreffenden Gebäudes] niederbrennen“. Dies bestätigten auch zwei weitere Zeuginnen.

Kurz nach 3 Uhr morgens fragte ein männlicher Bekannter Frau Bauer, ob sie nicht noch mit ihm zu ihm wolle. Sie erwiderte, sie habe keinen Babysitter, der auf ihre Tochter Katharina aufpassen könnte. Daraufhin bot Herr Wagner an, auf das Kind Acht zu geben, sodass Frau Bauer schließlich die Einladung ihres Bekannten annahm. Frau Bauer wies Herrn Wagner darauf hin, die Wohnung nicht zu verlassen, da sie keinen Schlüssel dabei hätte, die sie ihm überlassen könnte.

Die Staatsanwaltschaft geht davon aus, dass Herr Wagner auf seine Ex- Freundin Frau Schröder und deren neuen Freund Herrn Lange wütend war, deren Wohnung sich direkt unter der Wohnung von Frau Bauer, in der er Babysitten sollte, befand. Das Feuer habe er somit absichtlich gelegt, in der Hoffnung, dass das Feuer auch auf die Wohnung der Ex-Freundin, die sich zum Zeitpunkt des Brandes mit ihrem neuen Freund dort aufhielt, übergreifen würde.

Der Tathergang wurde folgendermaßen rekonstruiert: Herr Wagner habe die Wohnung von Frau Bauer verlassen und sei in das nahegelegene Gewächshaus eingebrochen, um dort Benzin und Verdünnungsmittel zu stehlen. Da er keinen Schlüssel zur Wohnung gehabt habe, sei er über den Elektrizitätskasten auf den Balkon von Frau Bauer geklettert. In der Wohnung habe er das Benzin und das Verdünnungsmittel verteilt und angezündet und sei anschließend mit den leeren Kanistern über den Balkon geflohen.

Hintergrundwissen Fachwissen irrelevant (Brandstiftung)

Informationen zum Thema Brandstiftung

1 Informationen zu §§ 306-306c:

1.1 §306 Brandstiftung

Tab.1 § 306 des StGB (Brandstiftung)

<p>(1) Wer fremde</p> <ol style="list-style-type: none"> 1 Gebäude oder Hütten, 2 Betriebsstätten oder technische Einrichtungen, namentlich Maschinen, 3 Warenlager oder -vorräte, 4 Kraftfahrzeuge, Schienen-, Luft- oder Wasserfahrzeuge, 5 Wälder, Heiden oder Moore oder land-, ernährungs- oder forstwirtschaftliche Anlagen oder Erzeugnisse <p>in Brand setzt oder durch eine Brandlegung ganz oder teilweise zerstört, wird mit Freiheitsstrafe von einem Jahr bis zu zehn Jahren bestraft.</p> <p>(2) In minder schweren Fällen ist die Strafe Freiheitsstrafe von sechs Monaten bis zu fünf Jahren.</p>
--

1.2 § 306a Schwere Brandstiftung

Tab.2 § 306a des StGB (Schwere Brandstiftung)

<p>(1) Mit Freiheitsstrafe nicht unter einem Jahr wird bestraft, wer</p> <ol style="list-style-type: none"> 1 ein Gebäude, ein Schiff, eine Hütte oder eine andere Räumlichkeit, die der Wohnung von Menschen dient, 2 eine Kirche oder ein anderes der Religionsausübung dienendes Gebäude oder 3 eine Räumlichkeit, die zeitweise dem Aufenthalt von Menschen dient, zu einer Zeit, in der Menschen sich dort aufzuhalten pflegen, <p>in Brand setzt oder durch eine Brandlegung ganz oder teilweise zerstört.</p>
--

Informationen zum Thema Brandstiftung

- (2) Ebenso wird bestraft, wer eine in § 306 Abs. 1 Nr. 1 bis 6 bezeichnete Sache in Brand setzt oder durch eine Brandlegung ganz oder teilweise zerstört und dadurch einen anderen Menschen in die Gefahr einer Gesundheitsschädigung bringt.
- (3) In minder schweren Fällen der Absätze 1 und 2 ist die Strafe Freiheitsstrafe von sechs Monaten bis zu fünf Jahren.

1.3 § 306b Besonders schwere Brandstiftung**Tab.3 § 306b des StGB (Besonders schwere Brandstiftung)**

- (1) Wer durch eine Brandstiftung nach § 306 oder § 306a eine schwere Gesundheitsschädigung eines anderen Menschen oder eine Gesundheitsschädigung einer großen Zahl von Menschen verursacht, wird mit Freiheitsstrafe nicht unter zwei Jahren bestraft.
- (2) Auf Freiheitsstrafe nicht unter fünf Jahren ist zu erkennen, wenn der Täter in den Fällen des § 306a
- 1 einen anderen Menschen durch die Tat in die Gefahr des Todes bringt,
 - 2 in der Absicht handelt, eine andere Straftat zu ermöglichen oder zu verdecken oder
 - 3 das Löschen des Brandes verhindert oder erschwert.

1.4 § 306c Brandstiftung mit Todesfolge**Tab.4 § 306c des StGB (Brandstiftung mit Todesfolge)**

Verursacht der Täter durch eine Brandstiftung nach den §§ 306 bis 306b wenigstens leichtfertig den Tod eines anderen Menschen, so ist die Strafe lebenslange Freiheitsstrafe oder Freiheitsstrafe nicht unter zehn Jahren.

1.5 § 306d Fahrlässige Brandstiftung

Informationen zum Thema Brandstiftung

Tab.5 § 306d des StGB (Fahrlässige Brandstiftung)

- | |
|--|
| <p>(1) Wer in den Fällen des § 306 Abs. 1 oder des § 306a Abs. 1 fahrlässig handelt oder in den Fällen des § 306a Abs. 2 die Gefahr fahrlässig verursacht, wird mit Freiheitsstrafe bis zu fünf Jahren oder mit Geldstrafe bestraft.</p> <p>(2) Wer in den Fällen des § 306a Abs. 2 fahrlässig handelt und die Gefahr fahrlässig verursacht, wird mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft.</p> |
|--|

Literatur

Deutsches Strafgesetzbuch

<http://www.gesetze-im-internet.de/stgb/index.html>

2 Der Grundsatz: „in dubio pro reo“

Wird gegen einen Bürger ein Strafverfahren durchgeführt, so sieht dieser sich zum Teil **massiven Eingriffen in seine Freiheitssphäre** (Freiheitsentzug) ausgesetzt. Diese Beschränkungen werden ihm dabei durch den Staat auferlegt. Von einer negativen Statusveränderung eines Bürgers wird gesprochen, wenn dieser Eingriff in die Freiheitssphäre eine gewisse Intensität erreicht und nicht nur zu einer kurzfristigen Einschränkung seiner Rechte führt. Diese Intensität ist erreicht, wenn die Freiheitsbeschränkung des Bürgers Wirkungen erzeugt, die ihn vor den Augen anderer stigmatisieren könnten (Rufschädigung). Denn in diesem Falle ist der Bürger erkennbar **in seiner sozialen Stellung innerhalb der Gesellschaft betroffen, d.h. in seinem negativ Status verändert**. Die bedeutsamste Statusveränderung im Strafverfahren ist die Entscheidung, dass der Beschuldigte zu verurteilen ist.

Der „in dubio pro reo“ (idpr) Satz gilt dort, wo die **Voraussetzungen für eine negative Statusveränderung nicht zweifelsfrei gegeben** sind. Zu entscheiden ist, ob der gegenwärtige Status des Betroffenen zu seinem Nachteil abgeändert wird. Diese Fragestellung kennt nur zwei Antworten: Entweder es wird der belastende Status zugesprochen oder der Betroffene verbleibt in seinem gegenwärtigen Status. Auf die Verurteilung bezogen

Informationen zum Thema Brandstiftung

bedeutet dass, dass die **Entscheidung, ob der Betroffene verurteilt wird nur mit „Ja“ oder „nein“ zu beantworten** ist.

⇒ **Zum Zweifel im Zweifelssatz.** Der Begriff des Zweifels lässt sich umschreiben als die **Unentschiedenheit zwischen einander widersprechenden Möglichkeiten** und ist demnach dort zu verzeichnen, wo bei zwei Alternativen sowohl die eine, als auch die andere als möglich angesehen wird. Bezieht sich ein Zweifel auf Tatsachen, so kann dieser nur darauf beruhen, dass Tatsachen nicht vollständig oder nicht widerspruchlos aufgeklärt wurden. Denn in Wirklichkeit bestehen diese Tatsachen, oder sie bestehen nicht. (z.B. hat ein Prüfling in einer Klausur geschummelt oder eben nicht. Möglicherweise kann dies nicht zweifelsfrei bewiesen werden; daran ob die Tat begangen wurde oder nicht, ändert dies allerdings nichts.) Zweifel im Sinne des „idpr“ setzen voraus, dass trotz Bemühungen seitens der Untersuchungsinstanzen hinreichende Ermittlungen anzustellen, keine Aufklärung der Tatsachen erzielt werden konnte. In dem Zusammenhang mit „idpr“ werden häufig Formulierungen wie „verbleibender Zweifel“ oder „nicht zu überwindender Zweifel“ verwendet. **Zweifel dürfen demnach nicht durch weitere Ermittlungen auflösbar erscheinen** oder durch einseitige, unzureichende Würdigung der ermittelnden Umstände hervorgerufen sein. Zweifel müssen in diesem Sinne endgültig sein. Einzelne Umstände, vermögen einen solchen Zweifel nicht zu begründen. **Erst nach Ermittlung und Würdigung** (und damit auch der Ermittlung) **aller Umstände**, die für den Schluss auf die Tatsache von Bedeutung sind, sei im Falle eines verbleibenden Zweifels an der Täterschaft Raum für den „idpr“ Grundsatz. Zweifel im Sinne von „idpr“ meint nicht einen solchen, der im Hinblick auf eine einzelne Indizientatsache besteht. Die Frage, ob die Behauptung einer Indizientatsache bewiesen, widerlegt oder nicht widerlegbar ist, kann nur im Rahmen einer **Gesamtwürdigung des Beweisstoffes** ihre Beantwortung finden. Gemeint ist dabei, dass alle Umstände, die unmittelbar für oder gegen die aus dem Vorliegen der Indizientatsache zu ziehende Schlussfolgerung sprechen, miteinbezogen werden. So mögen am Tatort gefundene Fasern nicht den zweifelsfreien Schluss auf eine Tatortanwesenheit ermöglichen, zugleich aber unzweifelhaft

Informationen zum Thema Brandstiftung

belegen, dass die Fasern nur von der Kleidung des Beschuldigten stammen können.

- ⇒ **Die Voraussetzungen des „idpr“- Satzes.** Eine Anwendung des „idpr“- Grundsatzes kommt dann in Betracht, **wenn Tatsachen nicht zweifelsfrei festgestellt** werden können, deren Vorliegen oder Fehlen Voraussetzung dafür ist, dass der Status des Beschuldigten negativ verändert werden kann.
- ⇒ **Zweifel und Überzeugung. Zweifel ist dort gegeben, wo es nicht gelungen ist,** sich von dem Vorliegen oder Nichtvorliegen eines bestimmten Geschehensverlaufs **eine Überzeugung zu bilden.** Danach müssen sich Überzeugung und Zweifel gegenseitig ausschließen. Bei der Bestimmung der Überzeugung geht es darum, dass **Risiko eines Fehlurteils einzugrenzen bzw. zu rechtfertigen.** Grundlage einer Überzeugungsbildung ist das in rechtsstaatlich gebotener Weise ermitteltes Tatsachenmaterial, welches für den in der Anklage behaupteten Sachverhalt von Bedeutung ist. Die **Überzeugungsbildung muss** in nachvollziehbarer Weise **vermittelbar an Dritte sein.** Es ist deshalb Aufgabe des Tatrichters bei Zweifeln bezüglich des Tatgeschehens, nicht nur darzulegen, dass es die Möglichkeit eines anderen Geschehensablaufs gibt, sondern auch verständlich machen zu können, weshalb dieser abweichende Geschehensablauf in Anbetracht der getroffenen Feststellung für diesen Fall Relevanz hat.

Literatur

Zopf, J. (1999). *Der Grundsatz „in dubio pro reo“*. Baden-Baden: Nomos Verlagsgesellschaft.

3 Das Standardvorgehen der Kriminalpolizei bei einem Brand

In der Brandermittlung gibt es drei große Felder. Bei etwa ein Drittel der Brände handelt es sich um vorsätzliche Brandstiftung, ein weiteres Drittel geht auf fahrlässige Brandstiftung zurück und bei dem letzten Drittel handelt es sich um Fälle, bei denen technische Brandursachen vorliegen oder die Ursache nicht geklärt werden konnte.

LS

Informationen zum Thema Brandstiftung

Die Brandermittler haben den Auftrag **neutral** an den Brand heranzugehen. Als erstes wird der Brandort durch die Schutzpolizei aufgesucht, die daraufhin ihre Erkenntnisse an die zuständige Kriminalpolizei weitergibt. Anschließend wird der **Brandort beschlagnahmt**. Nachdem die Feuerwehr den Brandort übergibt, muss sichergestellt werden, dass keine unbefugten Personen den Brandort mehr betreten. Entsprechend wird der **Brandort sichergestellt** (abgesperrt). Als nächstes werden die **Zeugen festgestellt**. Die Brandbetroffenen und die Feuerwehr werden zu dem Brand befragt. Die Identität von vermissten Personen, Verletzten und Leichen wird festgestellt und im Falle der letzten beiden Personengruppen, deren Unterbringungsort in Erfahrung gebracht. Weiter werden **Informationen zur Gefährdungslage** am Brandort eingeholt, beispielsweise, ob es bautechnische Gefährdungen gibt oder sich noch toxische Substanzen oder Gase am Brandort befinden. Anschließend wird ein Bericht zum Brandort angefertigt.

⇒ **Folgemaßnahmen.** Nachdem der Brandort besichtigt wurde, folgt die **Spurensicherung und Dokumentation**. Der Brandort wird untersucht. Dabei wird der Brandschutt sichergestellt und gegebenenfalls Rekonstruktionen von Zimmereinrichtung vorgenommen, mit dem Ziel herauszufinden wo der Brand genau ausgebrochen ist und warum es gerade dort angefangen hat zu brennen. (Möbelrekonstruktion: Stand zum Beispiel dort ein laufendes TV- Gerät, das möglicherweise einen Defekt hatte?) Dabei kommen zwei grundsätzliche Herangehensweisen in Betracht. Einmal das **Eliminierungsverfahren**. Hierbei wird aus vielen theoretischen Brandursachen, eine nach der anderen ausgeschlossen. Noch angewendet wird folgendes Vorgehen: Nachdem die Ermittler sich den Brandausbruchsort angesehen und festgestellt haben, wo das Feuer angefangen hat, wird **konzentriert nach begründbaren technischen Ursachen gesucht**. Z.B. wird betrachtet, ob am Ausbruchort (möglicherweise fehlerhafte) Elektrogeräte gestanden haben. Erst wenn technische Gründe als Brandursache ausgeschlossen werden können, wird nach einem Täter gesucht. Ist man z.B. durch Zeugenaussagen bereits vorher auf einen möglichen Brandstifter aufmerksam geworden, gilt es noch immer zuerst neutral nach technischen Ursachen zu suchen und diese

Informationen zum Thema Brandstiftung

auszuschließen, bevor man eine Brandstiftung in Betracht zieht. Können **technische Gründe definitiv ausgeschlossen** werden, muss ermittelt werden, ob es sich um eine **fahrlässige oder eine vorsätzliche Brandstiftung** handelt. Indizien bzw. Beweise am Tatort geben darüber Aufschluss. Entsprechende Sachverständige stellen dies durch kriminaltechnische Untersuchungen fest. Vor Ort wird ermittelt, ob es Reste von brennbaren Flüssigkeiten am Brandort gibt. Erst, wenn solche gefunden werden konnten, werden die Proben für genauere Untersuchungen zum Labor geschickt. Zur Untersuchung vor Ort können **Brandmittelspürhunde** eingesetzt werden oder ein **Photo-Ionisation-Detektor (PID)**. Proben vom Brandort, die in ein Labor verschickt werden, werden dabei in gasdichten Kunststoffbehältern transportiert. Gibt es einen **Beschuldigten** für die Legung des Brandes, werden **Ermittlungen gegen diesen eingeleitet**. Z.B. wird ihm Blut entnommen, seine Fingerabdrücke werden festgehalten und mit denen vom Brandort verglichen und es werden DNA-Analysen vorgenommen. Auch wird untersucht, ob wohlmöglich an seiner Kleidung die gleiche brennbare Flüssigkeit gefunden werden konnte, wie (wenn vorhanden) am Brandort.

4 Psychische Störungen, die für den Fall von Bedeutung sein könnten

Im relevanten Fall wird erwähnt, dass eine der beteiligten Personen unter folgenden Persönlichkeitsstörungen leiden würde. Zur Einschätzung der Bedeutung für deine Untersuchungen, werden hier noch einmal die Diagnosekriterien aus dem DSM – IV erwähnt:

4.1 Die Antisoziale Persönlichkeitsstörung

Tab.1 Diagnostische Kriterien für die Antisoziale Persönlichkeitsstörung aus Saß H., Wittchen, H.-U., Zaudig, M. (1998) Diagnostisches und statistisches Manual psychischer Störungen (DMS-IV). Göttingen:Hogrefe.)

Es besteht ein tiefgreifendes Muster von Missachtung und Verletzung der Rechte anderer, dass seit dem 15. Lebensjahr auftritt. Mindestens 3 der folgenden Kriterien müssen erfüllt sein:

7

Informationen zum Thema Brandstiftung

1	Versagen, sich in bezug auf gesetzmäßiges Verhalten gesellschaftlichen Normen anzupassen , was sich in wiederholten Begehen von Handlungen äußert, die einen Grund für eine Festnahme darstellen,
2	Falschheit , die sich in weiderholtem Lügen, dem Gebrauch von Decknamen oder dem Betrügen anderer zum persönlichen Vorteil oder Vergnügen äußert,
3	Impulsivität oder Versagen vorausschauend zu planen, Reizbarkeit und Aggressivität , die sich in wiederholten Schlägereien oder Überfällen äußert,
4	Rücksichtslose Missachtung der eigenen Sicherheit bzw. der Sicherheit anderer ,
5	Durchgängige Verantwortungslosigkeit , die sich in wiederholten Versagen zeigt, eine dauerhafte Tätigkeit auszuüben oder finanziellen Verpflichtungen nachzukommen
6	Fehlende Reue , die sich in Gleichgültigkeit oder Rationalisierung äußert, wenn die Person andere Menschen gekränkt, misshandelt oder bestohlen hat.
A	Die Person ist mindestens 18 Jahre alt.
B	Eine Störung des Sozialverhaltens war bereits vor Vollendung des 15. Lebensjahres erkennbar.
C	Das Antisoziale Verhalten tritt nicht ausschließlich im Verlauf einer Schizophrenie oder Manischen Episode auf.

⇒ Zu C. **Die vier Hauptkriterien der Störung des Sozialverhaltens in Kindheit und Jugend** sind:

- Feindselige Aggression gegenüber Menschen und Tieren,
- Zerstörung fremden Eigentums,
- Unehrlichkeit und Diebstähle sowie

4.2 **Die Borderline-Persönlichkeitsstörung**

Tab.2 Diagnostische Kriterien für die Borderline Persönlichkeitsstörung aus Saß H., Wittchen, H.-U., Zaudig, M. (1998) Diagnostisches und statistisches Manual psychischer Störungen (DMS-IV). Göttingen:Hogrefe.)

Informationen zum Thema Brandstiftung

Ein tiefgreifendes Muster von Instabilität in zwischenmenschlichen Beziehungen, im Selbstbild und in den Affekten sowie von deutlicher Impulsivität. Der Beginn liegt im frühen Erwachsenenalter und manifestiert sich in den verschiedenen Lebensbereichen. Mindestens 5 der folgenden Kriterien müssen erfüllt sein:

- 1 **Verzweifertes Bemühen, tatsächliches oder vermutetes Verlassenwerden zu vermeiden.**
Beachte: Hier werden keine suizidalen oder selbstverletzenden Handlungen berücksichtigt, die in Kriterium 5 enthalten sind.
- 2 **Ein Muster instabiler, aber intensiver zwischen-menschlicher Beziehungen**, das durch einen Wechsel zwischen den Extremen der Idealisierung und Entwertung gekennzeichnet ist.
- 3 **Identitätsstörung**: ausgeprägte und andauernde Instabilität des Selbstbildes oder der Selbstwahrnehmung.
- 4 **Impulsivität in mindestens zwei potentiell selbstschädigenden Bereichen** (Geldausgaben, Sexualität, Substanzmissbrauch, rücksichtsloses Fahren, „Fressanfälle“)
Beachte: Hier werden keine suizidalen oder selbstverletzenden Handlungen berücksichtigt, die in Kriterium 5 enthalten sind.
- 5 **Wiederholte suizidale Handlungen, Selbstmordandeutungen oder -drohungen oder Selbstverletzungsverhalten.**
- 6 **Affektive Instabilität**, die durch eine ausgeprägte Reaktivität der Stimmung gekennzeichnet ist (z.B. hochgradige episodische Dysphorie, Reizbarkeit oder Angst, wobei diese Verstimmung gewöhnlich einige Stunden nur selten mehr als einige Tage andauern).
- 7 **Chronische Gefühle der Leere.**
- 8 **Unangemessene, heftige Wut oder Schwierigkeiten, Wut od. Ärger zu kontrollieren** (z.B. häufige Wutausbrüche, andauernde Wut, wdh. körperliche Auseinandersetzungen).
- 9 Vorübergehende, durch Belastungen ausgelöste **paranoide Vorstellungen oder schwere dissoziative Symptome.**

Materialien zur Eignungsprüfung der Fallbeschreibung

Christian-Albrechts-Universität zu Kiel
INSTITUT FÜR PSYCHOLOGIE

Differentielle & Diagnostische Psychologie

Dr. Katja vom Schemm

Dipl.-Psych. Britta Dreger



Liebe/r Teilnehmer/in,

wir planen derzeit eine Studie zur diagnostischen Urteilsbildung im Kontext von rechtspsychologischen Fragestellungen. Dazu haben wir eine reale Fallgeschichte bzgl. einer Brandstiftung mit Todesfolge komprimiert zusammengestellt und wollen nun vor Studienbeginn die Eigenschaften der verwendeten Informationen überprüfen. Anbei finden Sie alle relevanten Artikel des Strafgesetzbuches, Material über ethische Rechtsprechung und Informationen zum Standardvorgehen der Kriminalpolizei bei Vorliegen eines Brandes.

Wir wären Ihnen sehr dankbar, wenn Sie sich 15-20 Minuten Zeit nehmen würden, um uns bei diesem Vorhaben zu unterstützen und damit einen wertvollen Beitrag für die Forschung zu leisten.

Sie bleiben in dieser Befragung vollkommen anonym, denn wir fragen Sie nirgends nach Ihrem Namen, sondern bitten Sie lediglich um einige allgemeine Angaben zu Ihrer Person:

Wie alt sind Sie?

..... Jahre

Ihr Geschlecht:

- weiblich
 männlich

Sind Sie Studierende/r?

- ja
 nein

Sind Sie bereits Diplom-Psychologe/in?

- ja
 nein
-

Vielen Dank für Ihre bisherigen Antworten! Sie erhalten nun die Fallinformationen. Bitte lesen Sie sich diese aufmerksam durch und beantworten Sie die Fragen auf den folgenden beiden Seiten!

Fallinformationen

In der Tatnacht zum 30. Juni 2006 um 4 Uhr 15 morgens brach im zweiten Stock eines Wohnhauses in Mannheim ein Feuer aus. Die zweijährige Katharina, die sich zu diesem Zeitpunkt alleine in der Wohnung aufgehalten hatte, starb an einer Rauchvergiftung. Der 22-jährige Peter Wagner wurde bald darauf als dringend tatverdächtig festgenommen und wegen Brandstiftung mit Todesfolge in Tateinheit mit Einbruch angeklagt.

Um die Transparenz der nachfolgenden Schilderungen vor, während und nach dem Tatgeschehen zu gewährleisten, soll die folgende Abbildung 1 die Wohnsituation der beteiligten Personen veranschaulichen.



Abb 1: Wohnsituation und beteiligte Personen.

Frau Schröder zog am 15. Juni 2006 in das Gebäude in dem der Brand später stattfand. Zu diesem Zeitpunkt lernte sie Herrn Wagner kennen, mit dem sie sehr schnell eine sexuelle Beziehung einging. Auch mit einem anderen Nachbarn, Herrn Braun, hatte sie Ende Juni eine einmalige sexuelle Begegnung, bevor sie Herrn Lange kennenlernte.

Am Sonntag den 29. Juni, also am Abend der Tatnacht, nahm Frau Schröder ihren neuen Freund Herrn Lange mit zu einer Feier in der nahegelegenen Wohnung von Frau König; einige Nachbarn, darunter auch Herr Wagner und Frau Bauer, nahmen an dieser Party teil.

Während dieser Feier küsste Frau Schröder Herrn Lange öffentlich und erklärte Herrn Wagner, sie wolle nun mit Herrn Lange zusammen sein. Um 1 Uhr morgens verließen Frau Schröder und Herr Lange gemeinsam die Feier und gingen zu Frau Schröders Wohnung.

Drei Zeugen, die ebenfalls auf der Feier waren, berichteten später, Herr Wagner habe die Absicht geäußert, er wolle den Wohnbereich [in dem auch Frau Schröder lebte] niederbrennen. Alle drei Zeugen erklärten, Herr Wagner sei sehr wütend gewesen und mindestens ein Zeuge glaubte, diese Wut liege an Frau Schröders neuer Beziehung. Ein weiterer Zeuge gab an, dass Herr Wagner gesagt habe: „Wenn ich sie [Frau Schröder] nicht haben kann, kann es keiner.“ Frau König, die Gastgeberin dieser Feier, konnte sich erinnern, Herr Wagner habe ihr gegenüber geäußert: „Noch bevor die Nacht vorbei ist, wird eine Wohnung [des betreffenden Gebäudes] niederbrennen“. Dies bestätigten auch zwei weitere Zeuginnen.

Kurz nach 3 Uhr morgens fragte ein männlicher Bekannter Frau Bauer, ob sie nicht noch mit ihm zu ihm wolle. Sie erwiderte, sie habe keinen Babysitter, der auf ihre Tochter Katharina aufpassen könnte. Daraufhin bot Herr Wagner an, auf das Kind Acht zu geben, sodass Frau Bauer schließlich die Einladung ihres Bekannten annahm. Frau Bauer wies Herrn Wagner darauf hin, die Wohnung nicht zu verlassen, da sie keinen Schlüssel dabei hätte, die sie ihm überlassen könnte.

Die Staatsanwaltschaft geht davon aus, dass Herr Wagner auf seine Ex- Freundin Frau Schröder und deren neuen Freund Herrn Lange wütend war, deren Wohnung sich direkt unter der Wohnung von Frau Bauer, in der er Babysitten sollte, befand. Das Feuer habe er somit absichtlich gelegt, in der Hoffnung, dass das Feuer auch auf die Wohnung der Ex-Freundin, die sich zum Zeitpunkt des Brandes mit ihrem neuen Freund dort aufhielt, übergreifen würde.

Der Tathergang wurde folgendermaßen rekonstruiert: Herr Wagner habe die Wohnung von Frau Bauer verlassen und sei in das nahegelegene Gewächshaus eingebrochen, um dort Benzin und Verdünnungsmittel zu stehlen. Da er keinen Schlüssel zur Wohnung gehabt habe, sei er über den Elektrizitätskasten auf den Balkon von Frau Bauer geklettert. In der Wohnung habe er das Benzin und das Verdünnungsmittel verteilt und angezündet und sei anschließend mit den leeren Kanistern über den Balkon geflohen.

Wir möchten, dass Sie sich in die Rolle eines/r Richters/-in versetzen:

Im Folgenden sind die Urteile aufgeführt, die in der Rechtssprechung möglich sind. Bitte geben Sie jeweils an, für wie wahrscheinlich Sie den jeweiligen Urteilsspruch bei dem konkreten Angeklagten Herrn Wagner aufgrund der Fallinformationen halten.

Kreuzen Sie bitte eine Zahl an und machen Sie kein Kreuz zwischen die Ziffern!

<u>Urteile</u>	Sehr unwahrscheinlich						Sehr wahrscheinlich	
Einstellung wegen Geringfügigkeit	0	1	2	3	4	5	6	
Schuld-Urteil	0	1	2	3	4	5	6	
Freispruch aus Mangel an Beweisen	0	1	2	3	4	5	6	
Freispruch wegen erwiesener Unschuld	0	1	2	3	4	5	6	

Für das weitere Vorgehen wäre es nun aus strafprozessualer Sicht wichtig, dass Sie sich für ein mögliches Urteil entscheiden, das Sie zunächst weiter abklären möchten, d.h. dass Sie zum jetzigen Zeitpunkt ein erstes Urteil stellen. Bitte geben Sie daher an welches Urteil Ihrer Meinung nach am wahrscheinlichsten ist. Treffen Sie Ihre Entscheidung aufgrund der vorliegenden Fallinformationen.

Ich komme zu folgendem ersten Urteil: _____

<u>Sicherheit</u>	Gar nicht						Sehr
	sicher						sicher
Wie sicher sind Sie sich bei Ihrer Entscheidung?	0	1	2	3	4	5	6

Bitte beantworten Sie auch die folgenden Fragen zu den Fallinformationen:

<u>Lesbarkeit</u>	Sehr schlecht						Sehr gut
	lesbar						lesbar
Wie schätzen Sie die Lesbarkeit des Textes ein?	0	1	2	3	4	5	6

<u>Plausibilität</u>	Gar nicht						Sehr
	plausibel						plausibel
Wie plausibel finden Sie die Informationen?	0	1	2	3	4	5	6

Unabhängig von Ihrem ersten Urteil: Wie schätzen Sie die Fallinformationen ein?

<u>Schuld (erwiesen)</u>	Gar nicht						Sehr
	charakteristisch						charakteristisch
Wie charakteristisch sind die Informationen für eine erwiesene Schuld?	0	1	2	3	4	5	6

<u>Freispruch (Mangel an Beweisen)</u>	Gar nicht						Sehr
	charakteristisch						charakteristisch
Wie charakteristisch sind die Informationen für einen Freispruch aus Mangel an Beweisen?	0	1	2	3	4	5	6

<u>Freispruch (erwiesene Unschuld)</u>	Gar nicht						Sehr
	charakteristisch						charakteristisch
Wie charakteristisch sind die Informationen für eine erwiesene Unschuld?	0	1	2	3	4	5	6

Ist Ihnen beim Lesen des Textes noch etwas aufgefallen?

Vielen herzlichen Dank für Ihre Mithilfe!

Konstruktion der Zusatzinformationen_Instruktionen

Liebe Rater!

Ich benötige zu meinem Studienmaterial einige Einschätzungen von unabhängigen und hypothesenblinden Ratern zur Validierung. Es existieren einerseits 39 Informationen, die aus einer Studie im Zusammenhang mit Essstörungen stammen und meinen eigentlichen VPn vorgelegt werden, die das Vorliegen einer Bulimie bei der jungen Patientin überprüfen sollen. Andererseits existieren 42 Informationen, die für die Schuld oder Unschuld eines Angeklagten in einem Brandstiftungsfall herangezogen werden. Es liegen daher insgesamt 81 Items vor, die in Bezug auf 5 Merkmale einzuschätzen sind, die im Folgenden erläutert werden.

Konsistenz. Auf einer Skala von *-5=sehr inkonsistent* über *0=neutral/irrelevant* bis *+5=sehr konsistent* soll eingeschätzt werden, wie konsistent, d.h. konform die Information mit einer bestimmten Annahme ist. Im Essstörungsfall geht es um die Frage, wie konsistent die Angaben im Falle einer vorliegenden Bulimie wären, d.h. ob sie zu dem typischen Krankheitsbild passen (konsistent) oder eigentlich nicht erwartet werden, wenn die Störung vorliegt (inkonsistent). Wenn die Informationen in dieser Hinsicht neutral sind oder gar nichts mit der Essstörung zu tun haben, soll die null angekreuzt werden. Im Brandstiftungsfall geht es analog um die Frage, ob man das Vorliegen der Information erwarten würde, wenn der Angeklagte schuldig ist (konsistent) oder gerade nicht (inkonsistent).

Wichtigkeit. Auf einer Skala von *0=gar nicht* bis *100=sehr wichtig* soll die Wichtigkeit des Items eingeschätzt werden. Es geht bei der Beurteilung darum, wie viel Nutzen man aus den einzelnen Informationen ziehen kann, um zu entscheiden, ob jemand unter einer Bulimie leidet bzw. einer Brandstiftung schuldig ist. Wenn eine Information kann dazu keinen Beitrag leisten, ist sie unwichtig, sonst je nach Nutzen unterschiedlich wichtig. Es soll bei dieser Einschätzung unterstellt werden, dass die gegebenen Informationen tatsächlich wahr sind, d.h. eine vollkommene Zuverlässigkeit aufweisen (s.u.)

Zuverlässigkeit. Auf einer Skala von *0=gar nicht* bis *100=vollkommen* soll die Zuverlässigkeit der Information eingeschätzt werden. Damit ist gemeint, wie fehlerfrei und damit reliable ein Item ist. Eine Information kann dabei auf unterschiedliche Weise fehlerbehaftet sein, bspw. weil es sich um eine subjektive Einschätzung handelt, weil man etwas nur durch Hören-Sagen kennt, weil man eine Motivation hat, zu lügen. Es geht daher letztendlich darum, wie viel Glauben man der Information schenken möchte.

Diagnostizität. Auf einer Skala von *0=gar nicht* bis *100= sehr diagnostisch* ist das Ausmaß der Diagnostizität des Items einzuschätzen. Ein Merkmal ist dann diagnostisch, wenn es unter der

angenommenen Hypothesen (entweder Bulimie oder Schuld des Angeklagten) eine andere Auftretenswahrscheinlichkeit hat als unter der Alternative (entweder Anorexie oder Unschuld). Ist es bspw. einerseits wahrscheinlich, dass eine extravertierte Person häufig auf Partys auf dem Tisch tanzt, und gleichzeitig andererseits unwahrscheinlich, dass eine introvertierte Person dies tut, so ist das Merkmal „auf dem Tisch tanzen“ diagnostisch. Das Merkmal „Bücher lesen“ hingegen ist nicht diagnostisch, denn dass ein Introvertierter zu einem Buch greift, ist genauso wahrscheinlich wie bei einem Extravertierten. Man muss sich also immer fragen, ob das genannte Merkmal unter der einen Hypothese genauso wahrscheinlich ist wie unter der anderen. Je unterschiedlicher die Wahrscheinlichkeiten sind, desto diagnostischer sind die Items.

Relevanz. Hinsichtlich des fünften Merkmals ist eine kategoriale Einschätzung abzugeben, d.h. die Information ist einer von 4 Merkmalsklassen zuzuordnen: eindeutiger Beleg, brauchbarer Hinweis, Widerspruch, irrelevante Information.

Ein „**eindeutiger Beleg**“ zeichnet sich dadurch aus, dass die betreffende Angabe vorliegen muss, damit das Urteil gerechtfertigt ist. Der Befund „*Person A erreicht auf der Skala Extraversion ein überdurchschnittliches Ergebnis.*“ kann bspw. einen eindeutigen Beleg darstellen. Im Fall des Essstörungskontexts ist die z.B. immer dann gegeben, wenn die Information ein Merkmal enthält, das in den diagnostischen Kriterien genannt wird.

Ein „**brauchbarer Hinweis**“ zeichnet sich dadurch aus, dass die betreffende Sachlage/ die betreffende Aussage vorliegen kann, wenn das Urteil gerechtfertigt ist, obwohl es nicht ausreicht, dass diese Information vorliegt. Es kommt gehäuft vor, jedoch nicht notwendigerweise. Die Information „*Person A wurde auf einer Party gesehen*“ wäre in diesem Zusammenhang ein Hinweis.

Ein **Widerspruch** zeichnet sich dadurch aus, dass die betreffende Angabe den Vorgaben für das Urteil widerspricht bzw. gewöhnlich nicht vorliegt. Die Beobachtung „*Person A kam nicht zur Party, um zuhause ein Buch zu lesen*“ wäre ein Widerspruch für Extraversion.

Eine **irrelevante Information** bezieht keinerlei Zusammenhang mit dem Urteil – ob sie vorliegt oder nicht, hat keinen Einfluss. *Dass Person A am liebsten Erdbeereis mag, ist irrelevant* für Extraversion.

Zunächst bitte ich darum, die Items im Zusammenhang mit der Essstörung zu bewerten. Dies geschieht pro Merkmal immer blockweise, d.h. erst werden alle Items präsentiert, um sie hinsichtlich der Konsistenz einzuschätzen. Dann werden die Items wiederholt und hinsichtlich des Informationsgehalts bewertet etc. Im Anschluss daran geschieht dasselbe für die

Brandstiftungsinformationen. Aufgrund der häufigen Wiederholungen der Items entsteht die große Anzahl an auszufüllenden Seiten. Ich hoffe jedoch, durch das blockweise Arbeiten das Kriterium, dass ihr jeweils für die Einschätzung anlegt, konstant halten zu können. Haltet euch also bitte unbedingt an die blockweise Darstellung und macht nur zwischen den verschiedenen Merkmalen eine Pause; möglichst nicht innerhalb eines Merkmalsblocks.

Es geht bei den Bewertungen um eure subjektive Einschätzung. Ein richtig oder falsch gibt es nicht, alleine eure Bewertung des Items zählt. Ihr erfüllt den Zweck des Ratings also am besten, indem ihr so spontan wie möglich eure persönliche Einschätzung abgebt. Dennoch ist es im Kontext der Essstörung erforderlich, dass ihr euch vor dem Rating (neben der Fallbeschreibung) den beigefügten Informationstext durchlest und im Zweifelsfall immer wieder zu Rate zieht. Er liefert euch das Fachwissen, das für die Entscheidung notwendig ist. Ein vergleichbares Fachwissen ist im Brandstiftungsfall nicht relevant, allerdings ist auch hier ein genaues Studierendes der Fallbeschreibung wichtig, da viele Personen an der Geschichte beteiligt sind und es essentiell ist, dass ihr bei den Bewertungen immer genau wisst, wer welche Rolle spielt.

Hinweis: Die Items sind nicht der Reihenfolge nach nummeriert, das hat für euch keine Bedeutung!

Zur Bewertung setzt ihr jeweils ein in die Spalte, die eure Meinung am besten wiedergibt!

Konstruktion der Zusatzinformationen_Items und Kennwerte

Tabelle B-1. Überblick über die 42 Items sowie deren Kennwerten aus dem gemeinsamen Rating

Nr.	Item	Konsistenz		Wichtigkeit		Zuverlässigkeit		Diagnostizität		Relevanz	
		M (SD)	t'	p	d	M (SD)	M (SD)	M (SD)	M (SD)		Code ²
1	Frau Bauer gab an, Herr Wagner habe regelmäßig auf Katharina Acht gegeben.	-1.33 (1.58)	-2.53	.02	0.84	13.33 (19.37)	57.78 (23.33)	5.56 (11.30)	I		6/9
2	Eine Probe des Teppichs aus der verbrannten Wohnung enthält Benzin, eine andere Teppichprobe und Holzsplitter aus der Wohnung enthielten Verdünnungsmittel. Zwei unabhängige Spezialisten bestätigten dies.	2.44 (1.51)	4.86	.00	1.62	64.44 (24.04)	95.56 (5.27)	57.78 (21.67)	H		5/9
3	Herr Wagner behauptete anfangs, er könnte sich nicht daran erinnern, zugestimmt zu haben auf Katharina aufzupassen.	0.44 (2.19)	0.61	.28	0.20	8.89 (9.28)	26.67 (18.71)	7.78 (6.67)	I		6/9
4	Nachdem Herr Wagner am 24. Juni erfahren hatte, dass Frau Schröder eine sexuelle Beziehung mit Herrn Braun hatte, konfrontierte er Herrn Lange und bedrohte ihn mit einem Messer.	2.44 (1.24)	5.93	.00	1.97	45.56 (28.33)	65.56 (19.44)	20.00 (18.03)	H		6/9
5	Ein Rauchmelder war nachweislich von der Decke entfernt worden.	1.89 (1.54)	3.69	.00	1.23	46.67 (28.28)	92.22 (16.42)	28.89 (32.19)	H		7/9
6	Ein Zeuge berichtete, er habe in der Nacht des Brandes Herrn Wagner betrunken in einem Gebüsch zusammenbrechen sehen.	0.00 (3.12)	0.00	.50	0.00	42.22 (32.70)	48.89 (25.71)	25.56 (27.44)	W		5/9
7	Brandinspektor Huber bezeugte, dass das Brandmuster in der Wohnung auch Resultat eines natürlich entstandenen Brandes sein könnte. Wenn die Annahme der Staatsanwaltschaft bezüglich des Ortes, an dem der Brand entstand und der Art, wie dies geschehen ist, zutreffen würde, wären überdies 37 Liter Benzin notwendig gewesen.	-3.89 (1.27)	-9.19	.00	3.06	85.56 (13.33)	72.22 (17.87)	58.89 (32.19)	W		9/9
8	Der Vater von Herrn Wagner sagte aus, sein Sohn sei mit 16 Jahren in einer jugendpsychiatrischen Anstalt gewesen.	1.00 (1.12)	2.68	.02	0.89	23.33 (18.71)	68.89 (27.13)	7.78 (10.93)	I		6/9

Anmerkungen.

¹ Die Anzahl der Freiheitsgrade beträgt jeweils $df = 8$ ² Code: I= irrelevante Information; H= brauchbarer Hinweis; W: Widerspruch; B= eindeutiger Beleg; - = keine Mehrheit; Ü = Übereinstimmung von X/9 Ratern

Nr.	Item	Konsistenz			Wichtigkeit		Zuverlässigkeit		Diagnostizität		Relevanz	
		M (SD)	t'	p	d	M (SD)	d	M (SD)	d	M (SD)	d	Code ²
9	Weder auf Herrn Wagners Kleidung noch auf seinen Stiefeln oder dem Verband, den er an diesem Abend getragen hatte, konnten Spuren von leicht entflammaren Substanzen wie z.B. Benzin gefunden werden.	-4.00 (1.12)	-10.73	.00	3.57	83.33 (19.37)		92.22 (8.33)	58.89 (34.08)		W	8/9
10	Herr Wagner gab an, Frau Schröder sei seine vierte feste Beziehung gewesen.	0.00 (0.50)	0.00	.50	0.00	3.33 (5.00)		63.33 (25.98)	2.22 (4.41)		I	9/9
11	In den Wochen vor dem Brand wurde der Feuerwehrhauptmann aufgrund von Rauchbildung aus unbekannter Quelle drei Mal zur Wohnung gerufen. Dies war bereits einmal geschehen, bevor Herr Wagner eine sexuelle Beziehung mit Frau Schröder einging.	-3.00 (0.87)	-1.39	.00	3.45	81.11 (14.53)		93.33 (11.18)	52.22 (33.83)		W	8/9
12	Zeugen sagten aus, dass Herr Wagner, nachdem er von der neuen Beziehung seiner Ex-Freundin erfahren hatte, nach einer Diskussion mit dieser und Herrn Lange die Beziehung akzeptierte. Bei der Beschlagnahme seines Tagebuches fand die Polizei einen Eintrag, der dies bestätigte.	-3.44 (1.33)	-7.75	.00	2.59	58.89 (32.19)		72.22 (28.19)	42.22 (29.49)		W	8/9
13	In der Nacht in der das Feuer ausbrach, trug Herr Wagner seinen Bundeswehr-Tarnanzug und entsprechende Stiefel.	0.78 (0.67)	3.50	.00	1.16	21.11 (14.53)		87.78 (13.02)	7.78 (9.71)		I	6/9
14	Herr Wagner gab an, er habe bevor er zu seinem Vater gegangen sei, nach Katharina gesehen, die geschlafen habe.	-0.78 (1.79)	-1.31	.12	0.44	25.56 (14.24)		46.67 (23.98)	30.00 (36.40)		---	4 W 4 I
15	Während Herr Wagner Herrn Braun, mit dem Frau Schröder kurzfristig eine sexuelle Beziehung hatte, mit einem Messer bedrohte, brach Herr Wagner sich in der Auseinandersetzung am 24. Juni die Hand. In der Nacht des Brandes war seine Hand noch in einem Gipsverband.	-2.56 (1.81)	-4.24	.00	1.41	57.78 (34.92)		90.00 (20.00)	37.78 (39.30)		W	7/9

Anmerkungen.¹ Die Anzahl der Freiheitsgrade beträgt jeweils $df = 8$ ² Code: I = irrelevante Information; H = brauchbarer Hinweis; W = Widerspruch; B = eindeutiger Beleg; - = keine Mehrheit; Ü = Übereinstimmung von X/9 Ratern

Nr.	Item	Konsistenz			Wichtig- keit	Zuver- lässigkeit		Diagnos- tizität		Relevanz
		M (SD)	t'	p		d	M (SD)	M (SD)	M (SD)	
16	Mehrere Psychologen bezeugten, dass Herr Wagner an einer Antisozialen und Borderline-Persönlichkeitsstörung litt. Er beschäftigte sich übermäßig mit Tod, Blut und Gewalt sowie mit Selbstverletzung und habe bereits in der Vergangenheit versucht, sich das Leben zu nehmen. Der Psychiater Dr. Kröger gab an, er schätze Herrn Wagner als extrem impulsiv, selbstzentriert und hedonistisch ein.	2.11 (1.36)	4.64	.00	1.55	55.56 (19.44)	85.56 (13.33)	28.89 (26.19)	H	6/9
17	Auf der Feier von Frau König schrieb Frau Schröder ihren Namen auf den Gips von Herrn Wagner.	0.22 (0.83)	0.80	.23	0.27	7.78 (9.72)	82.22 (23.86)	1.11 (3.33)	I	9/9
18	Während Herr Wagner in Untersuchungshaft saß, schrieb er einem Freund in Schottland: „... Wenn sie mich nur zu einer Gefängnisstrafe verurteilen, dann sollten sie besser hoffen, dass ich dort drin sterbe, denn wenn ich rauskomme, werde ich nicht Halt machen sie alle zu jagen, bis jeder, der in diesen Fall involviert ist, tot ist.“	1.33 (1.41)	2.83	.01	0.94	41.11 (22.61)	81.11 (23.15)	21.11 (20.88)	I	6/9
19	Als Frau Bauer Nachricht von dem Feuer und dem Tod ihrer Tochter bekam, machte sie keine Angaben bezüglich eines Babysitters, auch fragte sie nicht bei der Polizei nach dessen Verbleib. Erst später gab sie an, Herr Wagner hätte eingewilligt, auf Katharina in dieser Nacht aufzupassen.	-1.11 (1.83)	-1.82	.06	0.61	35.56 (25.06)	54.44 (26.03)	15.56 (24.04)	I	5/9
20	Frau Schröder wurde von mehreren Zeugen als promiskuitiv beschrieben.	-0.22 (0.67)	-1.00	.18	0.33	1.11 (3.33)	45.56 (28.77)	0.00 (0.00)	I	9/9
21	Während Herr Wagner in Untersuchungshaft saß, schrieb er einem Freund in Schottland: „Erinnerst du dich an den Tag an dem ich Greg mit deiner Waffe in den Kopf geschossen habe? Ich habe mich fast tot gelacht. Wenn die Polizei [in Schottland] auch nur die Hälfte von den Sachen herausfindet, die wir getan haben, führen sie für uns die Todesstrafe wieder ein.“	1.67 (1.32)	3.78	.00	1.27	40.00 (25.98)	74.44 (28.77)	26.67 (28.28)	I	6/9

Anmerkungen.

¹ Die Anzahl der Freiheitsgrade beträgt jeweils $df = 8$

² Code: I= irrelevante Information; H= brauchbarer Hinweis; W: Widerspruch; B= eindeutiger Beleg; - = keine Mehrheit; Ü = Übereinstimmung von X/9 Ratern

Nr.	Item	Konsistenz			Wichtigkeit		Zuverlässigkeit		Diagnositizität		Relevanz	
		M (SD)	t'	p	d	M (SD)	M (SD)	M (SD)	M (SD)	Code ²	Ü ²	
22	Herr Wagner nahm am Trauergottesdienst für Katharina teil.	-0.44 (0.73)	-1.84	.05	0.60	20.00 (24.45)	97.78 (6.67)	12.22 (14.81)	I	6/9		
23	Frau Bauer war von dem Verlust ihrer Tochter Katharina tief erschüttert und nahm eine Woche später psychologische Betreuung in Anspruch.	-0.22 (0.67)	-1.00	.18	0.33	1.11 (3.33)	86.67 (12.25)	2.22 (4.41)	I	9/9		
24	Herr Wagner gab in einem polizeilichen Interview an, dass er an dem Abend des Brandes in die Wohnung seines Vaters gegangen sei und dort verblieben wäre, bis er zum Einsatzort gekommen sei.	-2.22 (2.11)	-3.16	.01	1.05	71.11 (30.60)	36.67 (18.71)	52.22 (38.98)	W	8/9		
25	Der Eigentümer des Gewächshauses konnte nicht sagen, ob etwas von seinen Vorräten an Benzin und Verdünnungsmittel fehlte.	-2.22 (1.09)	-6.10	.00	2.04	70.00 (20.00)	66.67 (32.79)	17.78 (21.67)	W	5/9		
26	Frau Schröder und Herr Lange gaben beide an, sie seien leichte Schläfer, hätten in den Stunden vor dem Brand aber keinerlei Geräusche gehört. Das Schlafzimmerfenster war offen und weniger als zwei Meter von dem Elektrizitätskasten entfernt, an dem Herr Wagner hochgeklert sein soll.	-3.44 (1.13)	-9.14	.00	3.04	64.44 (26.03)	48.89 (35.16)	50.00 (34.64)	W	8/9		
27	2007 wiesen führende Wissenschaftler der USA darauf hin, dass die ursprünglichen Brandschuttanalysen völlig unzuverlässig waren. Von ihnen durchgeführte Untersuchungen führten zu der Erklärung, dass man keine Aussage darüber treffen könne, ob die Proben Spuren von brennbaren Substanzen enthielten oder nicht.	-3.11 (1.45)	-6.42	.00	2.14	84.44 (13.33)	84.44 (15.90)	43.33 (34.64)	I	5/9		
28	Einer von Herr Wagners früheren Lehrern sagte aus, dass er ein aufmerksamer und fleißiger Schüler gewesen sei.	-0.44 (1.01)	-1.32	.12	0.44	2.22 (4.41)	47.78 (29.49)	1.11 (3.33)	I	9/9		

Anmerkungen.

¹ Die Anzahl der Freiheitsgrade beträgt jeweils $df = 8$

² Code: I= irrelevante Information; H= brauchbarer Hinweis; W: Widerspruch; B= eindeutiger Beleg; - = keine Mehrheit; Ü = Übereinstimmung von X/9 Ratern

Nr:	Item	Konsistenz			Wichtig- keit		Zuver- lässigkeit		Diagnos- tizität		Relevanz	
		M (SD)	t'	p	d	M (SD)	M (SD)	M (SD)	M (SD)	Code ²	Ü ¹	
29	Zwei Experten führten die Untersuchungen der Brandschuttproben durch. Dabei benutzte Herr Roth eine Untersuchungsmethode, die wissenschaftlich nicht anerkannt war und von einer internationalen Autorität in forensischen Fragen kritisiert wurde. Der zweite hinzugezogene Experte war Herr Vogel. Dieser hatte zu diesem Zeitpunkt ein viertägiges Training in der forensischen Untersuchung von Brandstiftung absolviert und wurde angewiesen seine Untersuchung auf zehn Stunden zu begrenzen, um Geld zu sparen. Er wandte keine eigenen Testverfahren an und bestätigte das Urteil von seinem Kollegen Herrn Roth, das für Brandstiftung sprach.	-1.89 (2.62)	-2.16	.03	0.72	77.78 (20.48)	53.33 (38.73)	31.11 (33.33)	I		5/9	
30	Als Herr Wagner an den Einsatzort ankam, versuchte er wiederholt in das brennende Gebäude einzudringen - nach eigener Aussage, um Katharinas Leben zu retten. Die Polizei hielt ihn jedoch davon ab.	-2.33 (1.87)	-3.74	.01	1.25	45.56 (30.87)	70.00 (31.62)	26.67 (30.00)	W		5/9	
31	In einer späteren Vernehmung sagte Herr Wagner, er würde dem Staatsanwalt die Kehle durchschneiden.	0.44 (0.53)	2.53	.02	0.83	31.11 (24.72)	57.78 (25.87)	12.22 (14.81)	I		7/9	
32	Der Feuerwehrhauptmann hielt den Brand vorerst für einen Unfall.	-2.33 (1.23)	-5.72	.00	1.89	56.67 (29.16)	52.22 (25.39)	25.56 (23.51)	W		6/9	
33	Benzin und Verdünnungsmittel lagerten unverschlossen in zwei Fächern des Gewächshauses in das Herr Wagner in der Nacht des Brandes nach eigener Aussage eingebrochen war.	0.78 (2.49)	0.94	.19	0.31	66.67 (28.72)	83.33 (16.58)	31.11 (25.71)	H		7/9	
34	Es gibt Berichte darüber, dass Frau Bauer ihre Tochter regelmäßig alleine ließe und ihr manchmal vorher Schlaftabletten geben würde. Zwei Mal hatte das Jugendamt aufgrund von Benachrichtigungen durch Nachbarn mit ihrer Verbindung aufgenommen	-1.33 (1.30)	-3.02	.01	1.02	30.00 (25.50)	58.89 (27.59)	11.11 (18.33)	I		7/9	

Anmerkungen.

¹ Die Anzahl der Freiheitsgrade beträgt jeweils $df = 8$

² Code: I= irrelevante Information; H= brauchbarer Hinweis; W: Widerspruch; B= eindeutiger Beleg; - = keine Mehrheit; Ü = Übereinstimmung von X/9 Ratern

Nr.	Item	Konsistenz		Wichtigkeit		Zuverlässigkeit		Diagnostizität		Relevanz	
		M (SD)	t'	p	d	M (SD)	d	M (SD)	M (SD)	M (SD)	Code ²
35	Fünf Zeugen sagten aus, dass, als Herr Wagner beim brennenden Gebäude ankam, er wiederholt gerufen habe. „Da ist noch ein Kind im Haus.“	-1.78 (1.64)	-3.25	.01	1.09	40.00 (28.72)		81.11 (9.28)	16.67 (21.21)	I	5/9
36	Am 1. Juli wurde Herr Wagner festgenommen. In einem aufgezeichneten polizeilichen Interview wies er zurück das Feuer gelegt zu haben.	-2.33 (1.87)	-3.74	.01	1.25	60.00 (33.91)		64.44 (25.06)	35.56 (32.82)	W	6/9
37	Als Teil der Untersuchungen wurden sechs Proben aus dem Brandschutt entnommen. Mehrere dieser Proben stammten von dem Teppich, der vorerst im Müll entsorgt worden war. Zwei Tage nach dem Brand wurde er vom Müll geholt und für drei Wochen auf dem Parkplatz eines Polizisten gelagert, keine 13 Meter entfernt von einer Tankstelle, bevor er in ein entsprechendes Labor zur Untersuchung gebracht wurde. Auch eine Probe eines Holzstücks wurde erst circa 3 Wochen nach dem Brand aus der Wohnung zur Untersuchung entfernt.	-2.89 (1.27)	-6.83	.00	2.28	74.44 (29.63)		68.89 (34.80)	27.78 (32.32)	---	---
38	In einem polizeilichen Interview direkt nach seiner Festnahme, gab Herr Wagner an, in der Nacht des Brandes betrunken gewesen zu sein und sich nicht an vieles mehr erinnern zu können.	-0.44 (1.51)	-0.88	.20	0.29	53.33 (30.00)		47.78 (21.67)	8.89 (11.67)	H	6/9
39	Als Herr Wagner an der Brandszene ankam, versuchte er vermehrt in das brennende Gebäude einzudringen, nach eigener Aussage, um Katharinas Leben zu retten. Die Polizei musste ihn bald darauf davon abhalten.	-2.11 (1.76)	-3.59	.01	1.20	46.25 (34.62)		66.67 (25.50)	23.33 (28.28)	I	5/9
40	Frau König, die Gastgeberin der Feier, nahm ihre anfänglich bei der Polizei gemachte Aussage, sie habe gehört, wie Herr Wagner drohte die Wohnung niederzubrennen in einem beglaubigten Schreiben zurück.	-2.67 (1.32)	-6.05	.00	2.02	66.67 (20.62)		63.33 (27.39)	35.56 (32.83)	W	8/9

Anmerkungen.

¹ Die Anzahl der Freiheitsgrade beträgt jeweils $df = 8$

² Code: I= irrelevante Information; H= brauchbarer Hinweis; W: Widerspruch; B= eindeutiger Beleg; - = keine Mehrheit; Ü = Übereinstimmung von X/9 Ratern

Nr.	Item	Konsistenz			Wichtig- keit		Zuver- lässigkeit		Diagnos- tizität		Relevanz	
		M (SD)	t'	p	d	M (SD)	M (SD)	M (SD)	M (SD)	Code ²	Ü ²	
41	Herr Wagner gab zu in das Gewächshaus eingebrochen zu sein, um für Frau Bauer zwei Pflanzen zu stehlen. Benzin und Verdünnungsmittel waren ebenfalls dort gelagert.	-1.11 (1.83)	-1.82	.06	0.61	57.78 (28.63)	48.89 (29.35)	30.00 (23.98)	H	5/9		
42	Da der Feuerwehrhauptmann Herr Peters anfangs vermutet hatte das Feuer sei aufgrund eines Unfalls ausgebrochen, wurden Teile des Brandschuttes vom Wohnungseigentümer entfernt. Unter dem Verdacht auf Brandstiftung wäre die Wohnung für Untersuchungen versiegelt worden.	-1.11 (1.27)	-2.63	.02	0.87	62.22 (25.39)	83.33 (17.32)	17.78 (23.33)	---	---		

Anmerkungen.

¹ Die Anzahl der Freiheitsgrade beträgt jeweils $df = 8$

² Code: I= irrelevante Information; H= brauchbarer Hinweis; W: Widerspruch; B= eindeutiger Beleg; - = keine Mehrheit; Ü = Übereinstimmung von X/9 Ratern
Zusatzanmerkung. Die Items 1-18 wurden im Zuge der AV 4 (Bestätigungstendenz B) eingesetzt. Die Items 19-30 für die AV 5 a+b (Bestätigungstendenz I sowie hypothesenkonstante Interpretation uneindeutiger Evidenz). Die Items 31-42 wurden nicht verwendet.

Anhang C - Studie aus Probandensicht

Einladungsschreiben	275
Bildschirmansichten.....	277

Einladungsschreiben





CAU Kiel – Institut für Psychologie - Olshausenstraße 75 - 24118 Kiel

Liebe Kolleginnen, liebe Kollegen!

Liebe Kommilitoninnen, liebe Kommilitonen!

Institut für Psychologie
*Abteilung für Psychologische Diagnostik,
Rechts- und Persönlichkeitspsychologie*

Prof. Dr. Günter Köhnken
Dipl.-Psych. Britta Dreger

Durchwahl: 0431-880-1201
E-Mail: dreger@psychologie.uni-kiel.de

Einladung zur Teilnahme am Forschungsprojekt „OPUS“

Sehr geehrte Damen und Herren,

am Institut für Psychologie der CAU zu Kiel läuft derzeit das Forschungsprojekt „**OPUS**“ an, das sich mit der **OPTIMIERUNG PSYCHOLOGISCHER URTEILSFINDUNGS-STRATEGIEN** beschäftigt und sich sowohl an Studierende der Psychologie, klinisch tätige Diplom-Psychologen/-innen, approbierte Psychotherapeuten/-innen sowie Fachärzte/-innen mit Zusatzausbildung in Psychotherapie richtet. Wir möchten Sie ganz herzlich bitten, unser Forschungsprojekt „**OPUS**“ zu unterstützen, indem Sie an unserer internetbasierten Befragung teilnehmen. Sie erreichen die Studie unter:

<http://opus.psychologie.uni-kiel.de>

Nachdem Sie sich mithilfe des Passworts **opus2008** eingeloggt haben, präsentieren wir Ihnen dort eine Fallgeschichte, bzgl. derer wir Sie um ein diagnostisches Urteil bitten möchten. Grundlegend ist dabei, dass die Forschung in den vergangenen Jahrzehnten einerseits die mangelnde Übereinstimmung von mehreren Diagnostikern trotz identischer Informationslage angeprangert hat, und dass der (menschliche) Diagnostiker andererseits jedoch aufgrund seiner besonderen Fähigkeiten zur Bildung von Gesetzmäßigkeiten geschätzt wird. Der menschliche Diagnostiker scheint daher *Getriebe der Urteilsbildung* und *Sand* gleichermaßen zu sein. Das vorliegende Forschungsprojekt soll einen Beitrag zur Untersuchung der entsprechenden Einflussgrößen leisten.

Wir wären Ihnen daher sehr verbunden, wenn Sie sich ca. 50 Minuten Zeit für die Studie nähmen, in denen Sie einen sehr wertvollen Beitrag für die Wissenschaft leisten könnten. Im Austausch für die von Ihnen investierte Zeit werden wir für jeden vollständig ausgefüllten und verwertbaren Fragebogen 2,50 € an die Stiftung KinderHerz Deutschland. Selbstverständlich garantieren wir Ihnen für die wissenschaftlich auszuwertenden Daten absolute Anonymität. Nähere Informationen erhalten Sie in der Datenschutzerklärung zu Beginn der Befragung.

Bitte scheuen Sie sich nicht, diesen Aufruf an Kollegen/-innen weiterzuleiten und bei Rückfragen mit uns Kontakt aufzunehmen.

Beste Grüße aus Norddeutschland,



Bildschirmansichten

OPUS
Optimierung Psychologischer
Urteilsfindungs-Strategien

CAU
Christian-Albrechts-Universität zu Kiel

Begrüßung

Sehr geehrte Teilnehmerin, sehr geehrter Teilnehmer,
Wir begrüßen Sie ganz herzlich beim Forschungsprojekt OPUS und danken Ihnen sehr für Ihr Interesse!

Am Institut für Psychologie der Christian-Albrechts-Universität zu Kiel läuft derzeit ein Forschungsprojekt zur **OPTIMIERUNG PSYCHOLOGISCHER URTEILSFINDUNGS-STRATEGIEN** – kurz OPUS. Dabei sollen die Urteilsfindungsstrategien von angehenden und bereits im Berufsleben stehenden Psychologen/-innen und Therapeuten/-innen betrachtet werden.

Auf der nächsten Seite erhalten Sie detaillierte Informationen über Dauer und Aufbau der Studie sowie Angaben zum Datenschutz. Bitte lesen Sie sich die entsprechenden Hinweise genau durch und entscheiden Sie im Anschluss daran über eine Mitarbeit an dem Forschungsprojekt!

Wir möchten zunächst unseren rechtlichen Pflichten nachkommen und geben Ihnen daher an dieser Stelle die Möglichkeit, sich über die obligatorischen Haftungsbeschränkungen und unsere Anbieterkennzeichnung zu informieren. Die erscheinenden Fenster können Sie über (X) wieder schließen und gelangen zurück zu dieser Stelle der Befragung.

Kosten und Nutzen

Die Teilnahme an dieser Befragung ist für Sie völlig risikofrei, da Sie freiwillig teilnehmen und Ihre Anonymität gewahrt bleibt.

Während es für Sie persönlich zunächst keine Vorteile durch die Teilnahme gibt, so hoffen wir doch, durch die Ergebnisse einen Beitrag zur Verbesserung der Qualität in der psychologischen Urteilsbildung leisten zu können. Außerdem spenden wir für jeden komplett(!) ausgefüllten und verwertbaren Fragebogen einen Betrag von 2,50€ an die [Stiftung KinderHerz Deutschland](#). Durch Ihre Teilnahme leisten Sie also sowohl einen wissenschaftlichen, als auch einen karitativen Beitrag!



Die Studie OPUS richtet sich ausdrücklich an eine ausgewählte Zielgruppe. Die eingeladenen Teilnehmer/-innen haben ein Passwort erhalten, welches Sie nun bitte in das folgende Textfeld eingeben. Die Studie kann erst nach Eingabe dieses Passworts begonnen werden. Sollten Sie über kein Passwort verfügen, jedoch Interesse an der Teilnahme an der Studie haben, setzen Sie sich bitte per E-Mail an dreger@psychologie.uni-kiel.de mit uns in Verbindung und teilen Sie uns bitte in kurzen Worten mit, welcher Berufsgruppe Sie angehören bzw. welches Studienfach Sie studieren. Wir werden Ihnen dann umgehend antworten und Ihnen ggf. das Passwort mitteilen.

Passwort:

Da sich das Projekt OPUS ausdrücklich an spezielle Berufsgruppen/Studenten richtet, würden wir gerne schon an dieser Stelle von Ihnen erfahren, welche Aussage auf Sie zutrifft:

- Ich studiere derzeit Psychologie
- Ich befinde mich derzeit in der Ausbildung zum/r Psychologischen Psychotherapeuten/-in bzw. Kinder- und Jugendlichenpsychotherapeuten/-in
- Ich habe mein Psychologie-Studium abgeschlossen und arbeite im klinischen Bereich (ohne Therapieausbildung)
- Ich bin approbierte/r Psychologische/r Psychotherapeut/-in bzw. Kinder- und Jugendlichenpsychotherapeut/-in
- Ich bin Arzt/Ärztin im psychotherapeutischen Bereich (Auch Ärzte/-innen in Weiterbildung)
- Auf mich trifft keine der genannten Beschreibungen zu

Abbildung C-1. Erscheinungsbild Online-Studie

<p>Begrüßung</p> <p>Sehr geehrte Teilnehmerin, sehr geehrter Teilnehmer, Wir begrüßen Sie ganz herzlich beim Forschungsprojekt OPUS und danken Ihnen sehr für Ihr Interesse!</p> <p>Am Institut für Psychologie der Christian-Albrechts-Universität zu Kiel läuft derzeit ein Forschungsprojekt zur OPTIMIERUNG PSYCHOLOGISCHER URTEILSFINDUNGS-STRATEGIEN – kurz OPUS. Dabei sollen die Urteilsfindungsstrategien von angehenden und bereits im Berufsleben stehenden Psychologen/-innen und Therapeuten/-innen betrachtet werden.</p> <p>Auf der nächsten Seite erhalten Sie detaillierte Informationen über Dauer und Aufbau der Studie sowie Angaben zum Datenschutz. Bitte lesen Sie sich die entsprechenden Hinweise genau durch und entscheiden Sie im Anschluss daran über eine Mitarbeit an dem Forschungsprojekt!</p>
<p>Wir möchten zunächst unseren rechtlichen Pflichten nachkommen und geben Ihnen daher an dieser Stelle die Möglichkeit, sich über die obligatorischen Haftungsbeschränkungen und unsere Anbieterkennzeichnung zu informieren. Die erscheinenden Fenster können Sie über (X) wieder schließen und gelangen zurück zu dieser Stelle der Befragung.</p> <p style="text-align: center;"><input type="button" value="Haftungsausschluss"/></p> <p style="text-align: center;"><input type="button" value="Impressum"/></p>
<p>Kosten und Nutzen</p> <p>Die Teilnahme an dieser Befragung ist für Sie völlig risikofrei, da Sie freiwillig teilnehmen und Ihre Anonymität gewahrt bleibt.</p> <p>Während es für Sie persönlich zunächst keine Vorteile durch die Teilnahme gibt, so hoffen wir doch, durch die Ergebnisse einen Beitrag zur Verbesserung der Qualität in der psychologischen Urteilsbildung leisten zu können. Außerdem spenden wir für jeden komplett(!) ausgefüllten und verwertbaren Fragebogen einen Betrag von 2,50€ an die <u>Stiftung KinderHerz Deutschland</u>. Durch Ihre Teilnahme leisten Sie also sowohl einen wissenschaftlichen, als auch einen karitativen Beitrag!</p> <div style="text-align: center;">  </div> <p>Die Studie OPUS richtet sich ausdrücklich an eine ausgewählte Zielgruppe. Die eingeladenen Teilnehmer/-innen haben ein Passwort erhalten, welches Sie nun bitte in das folgende Textfeld eingeben. Die Studie kann erst nach Eingabe dieses Passworts begonnen werden. Sollten Sie über kein Passwort verfügen, jedoch Interesse an der Teilnahme an der Studie haben, setzen Sie sich bitte per E-Mail an dreger@psychologie.uni-kiel.de mit uns in Verbindung und teilen Sie uns bitte in kurzen Worten mit, welcher Berufsgruppe Sie angehören bzw. welches Studienfach Sie studieren. Wir werden Ihnen dann umgehend antworten und Ihnen ggf. das Passwort mitteilen.</p> <p style="text-align: center;">Passwort: <input type="text"/></p>
<p>Da sich das Projekt OPUS ausdrücklich an spezielle Berufsgruppen/Studenten richtet, würden wir gerne schon an dieser Stelle von Ihnen erfahren, welche Aussage auf Sie zutrifft:</p> <p><input type="radio"/> Ich studiere derzeit Psychologie</p> <p><input type="radio"/> Ich befinde mich derzeit in der Ausbildung zum/r Psychologischen Psychotherapeuten/-in bzw. Kinder- und Jugendlichenpsychotherapeuten/-in</p> <p><input type="radio"/> Ich habe mein Psychologie-Studium abgeschlossen und arbeite im klinischen Bereich (ohne Therapieausbildung)</p> <p><input type="radio"/> Ich bin approbierte/r Psychologische/r Psychotherapeut/-in bzw. Kinder- und Jugendlichenpsychotherapeut/-in</p> <p><input type="radio"/> Ich bin Arzt/Ärztin im psychotherapeutischen Bereich (Auch Ärzte/-innen in Weiterbildung)</p> <p><input type="radio"/> Auf mich trifft keine der genannten Beschreibungen zu</p>

Abbildung C-2. Bedingungsinvarianter Begrüßungsbildschirm mit Abfrage des beruflichen Status

<p>Studieninhalt</p> <p>Wir möchten Sie bitten, an einer Studie teilzunehmen, die untersucht, welche Strategien Personen zur Urteilsfindung einsetzen, wenn sie mit einer beruflichen Fragestellung konfrontiert werden. Es ist dabei von Interesse, welche Informationen Ihnen wichtig erscheinen und wie Sie diese bewerten. Außerdem werden u.a. auch soziodemografische Faktoren erhoben, die mit der Urteilsfindung assoziiert sein könnten.</p>
<p>Ablauf</p> <ul style="list-style-type: none"> • Nachdem Sie der Teilnahmeerklärung(s.u.) zugestimmt haben, erhalten Sie einen Informationstext, der Sie über den Kontext der diagnostischen Entscheidungsfindung informieren wird. • Im Anschluss an die Hintergrundinformationen, die Ihnen während der gesamten Befragung zur Verfügung stehen werden, erhalten Sie eine Fallbeschreibung, zu der Sie sich ein Urteil bilden sollen. • Dafür erhalten Sie nach einander verschiedene weitere Informationen, zu denen wir Ihre Einschätzung erfragen wollen. • Im Anschluss an die Bearbeitung der Fallgeschichte präsentieren wir Ihnen noch einen Fragebogen zu persönlichen soziodemografischen Daten, in dem nach Informationen wie z.B. Geschlecht, Alter und Berufserfahrungen gefragt wird. • Die Befragung endet mit weiteren Informationen zu den Hintergründen der Studie.
<p>Vertraulichkeit</p> <p>Alle erhobenen Daten werden selbstverständlich streng vertraulich behandelt und ausschließlich für wissenschaftliche bzw. didaktische Zwecke ausgewertet und präsentiert. Zu keinem Zeitpunkt fragen wir Sie nach Ihrem Namen, so dass Ihre Identität nicht preisgegeben werden kann.</p>
<p>Kontaktmöglichkeiten</p> <p>Sollten Sie Fragen haben, dann scheuen Sie sich nicht, uns zu kontaktieren. Nutzen Sie dafür die angegebene E-Mail-Adresse oder nehmen Sie telefonisch Kontakt auf.</p> <p>Britta Dreger, Dipl.-Psych. dreger@psychologie.uni-kiel.de Tel.: 0431/880-1201</p>
<p>Bitte geben Sie nun durch Anklicken an, ob Sie an dem Forschungsprojekt teilnehmen möchten. Sie bestätigen damit, dass Sie die oben angegebenen Informationen gelesen haben und keine Fragen unbeantwortet geblieben sind.</p>
<p>Ich bin bereit, an dem Forschungsprojekt OPUS teilzunehmen.</p> <p><input type="radio"/> Ja</p> <p><input type="radio"/> Nein</p>

Abbildung C-3. Bedingungsinvariante Erläuterungen zu Inhalt, Ablauf und Vertraulichkeit

<p>Vielen Dank für Ihre Teilnahmebereitschaft!</p> <p>Die folgende Online-Befragung unterstützt Sie technisch beim korrekten Ausfüllen der einzelnen Fragen. In der Regel wird es so sein, dass bei jeder Frage eine Antwort in Form einer Eingabe in einem Eingabefeld bzw. die Auswahl über ein entsprechendes Element erfolgen muss. Andersfalls wird in den Bearbeitungshinweisen explizit auf die Optionalität einer Antwort hingewiesen. Auf folgende Dinge möchten wir Sie vor dem Beginn der Bearbeitung des Fragebogens gerne hinweisen:</p> <ul style="list-style-type: none"> • Die Anwendung beschränkt die Eingabemöglichkeiten auf sinnvolle Zeichen. Bei einem Eingabefeld, welches nach einer Anzahl fragt, wird z.B. auch nur die Eingabe einer Ziffer akzeptiert. • Erst nach dem korrekten Ausfüllen aller erforderlichen Fragen eines Bildschirms wird auch der WEITER-Button aktiv geschaltet und Sie können zum nächsten Bildschirm voranschreiten. Durch diese Technik werden versehentliche Fehleingaben vermieden. • Nachdem Sie den WEITER-Button betätigt haben, ist keine Änderung vorheriger Eingaben mehr möglich. Bitte lesen Sie sich alle Fragen also sorgfältig durch, bevor Sie sie beantworten und zum nächsten Bildschirm wechseln. Bitte betätigen Sie auf keinen Fall den "Zurück"-Button Ihres Browsers. In diesem Fall würden ALLE bereits getätigten Eingaben verloren gehen! • Es ist möglich, dass sich bei Ihrem individuellen Bildschirm Fragen außerhalb des sichtbaren Bildschirmbereiches befinden. Sie erkennen dies an dem Scrolling-Balken am rechten Rand dieser Anwendung. Verschieben Sie diesen wie aus anderen Anwendungen gewohnt. • Im oberen Bereich der Anwendung sehen Sie einen Fortschrittsbalken, der Sie darüber informiert, an welcher Stelle der Bearbeitung Sie sich befinden. • Wenn Sie eine Antwort eingeben möchten, klicken Sie mit der Maus entweder auf das entsprechende Steuerelement oder auf das Texteingabefeld. Sobald Sie dies getan haben, wird das entsprechende Element von einem blauen Rahmen umgeben und Sie können Ihre Antwort verfassen. <p>Und nun wünschen wir Ihnen viel Erfolg bei der Bearbeitung des Fragebogens.</p>

Abbildung C-4. Bedingungsinvariante technischen Erläuterungen zur Bearbeitung der Studie

Klinische Fragestellung
Wir legen Ihnen nun unmittelbar eine klinische Fragestellung vor, bei der es um die Frage des Vorliegens einer psychischen Störung geht, da es sich hierbei um eines der Kerngebiete psychologischer Diagnostik handelt, das zudem mit einer hohen Verantwortung einhergeht.
Eine bereits gewonnene Erkenntnis möchten wir dabei gerne nutzen: da Studien zeigen konnten, dass das Urteilsvermögen positiv beeinflusst wird, wenn der Beurteiler sich in die Situation hinein versetzt und aktiv die Rolle des Diagnostikers einnimmt, möchten wir Sie bitten, sich vorzustellen, dass Sie in einer Klinik tätig sind und dass es sich bei der jungen Frau im Rahmen der Fallgeschichte um eine Ihrer Patientinnen handelt.
Natürlich handelt es sich bei dieser Studie nicht um ein getreues Abbild der Realität, da wir Ihnen Informationen ausschließlich in schriftlicher Form vorlegen und Sie keinen persönlichen Kontakt mit der zu Begutachtenden herstellen können. Dennoch möchten wir Sie bitten, die Fragen mit Ihren erlernten Fähigkeiten und Fertigkeiten nach bestem Wissen und Gewissen zu beantworten.

Abbildung C-5. *Bedingung Fachwissen relevant: Einführung in die zu bearbeitende klinische Fragestellung*

Kriminalistische Fragestellung
Wir legen Ihnen nun unmittelbar eine kriminalistische Fragestellung vor, bei der es um die Frage der Schuld bzw. Unschuld eines fiktiven Angeklagten geht, da es sich hierbei um ein Gebiet handelt, das mit einer hohen Verantwortung einhergeht und bei dem diagnostische Fähigkeiten, wie sie unter anderem in der Ausbildung von Psychologen bzw. Therapeuten vermittelt werden, von Bedeutung sind.
Eine bereits gewonnene Erkenntnis möchten wir dabei gerne nutzen: da Studien zeigen konnten, dass das Urteilsvermögen positiv beeinflusst wird, wenn der Beurteiler sich in die Situation hinein versetzt, und aktiv die Rolle des Diagnostikers einnimmt, möchten wir Sie bitten, sich vorzustellen, dass Sie als Richter beim Strafgericht tätig sind und dass es sich bei dem jungen Mann im Rahmen der Fallgeschichte, um einen Angeklagten handelt, über den Sie ein Urteil fällen sollen.
Natürlich handelt es sich bei dieser Studie nicht um ein getreues Abbild der Realität, da wir Ihnen Informationen ausschließlich in schriftlicher Form vorlegen und Sie keinen persönlichen Kontakt mit dem zu begutachtenden Angeklagten herstellen können. Dennoch möchten wir Sie bitten, die Fragen mit Ihren erlernten Fähigkeiten und Fertigkeiten nach bestem Wissen und Gewissen zu beantworten.

Abbildung C-6. *Bedingung Fachwissen irrelevant: Einführung in die zu bearbeitende kriminalistische Fragestellung*

Hintergrundinformationen
Wir möchten Ihnen nun einige Hintergrundinformationen zur Verfügung stellen. Ab sofort öffnet sich ein Lehrbuchkapitel zum Thema Essstörungen , sobald Sie im unteren Bereich der Anwendung auf den Button mit der Beschriftung INFO klicken.
Bitte öffnen Sie das Kapitel jetzt und lesen Sie die Hintergrundinformationen. Durch das Schließen (X) kehren Sie zum Fragebogen zurück. Sie haben im weiteren Verlauf der Studie jederzeit die Möglichkeit, diese Informationen erneut aufzurufen.

Abbildung C-7. *Bedingung Fachwissen relevant: Hinweis auf die Hintergrundinformationen, die über einen Button zu öffnen sind*

Hintergrundinformationen
Wir möchten Ihnen nun die Hintergrundinformationen zur Verfügung stellen. Ab sofort öffnet sich ein Informationstext zum Thema Brandstiftung , sobald Sie im unteren Bereich der Anwendung auf den Button mit der Beschriftung INFO klicken.
Bitte öffnen Sie den Informationstext jetzt und lesen Sie die Hintergrundinformationen. Durch das Schließen (X) kehren Sie zum Fragebogen zurück. Sie können die Information jederzeit im Verlauf der Befragung erneut aufrufen.

Abbildung C-8. *Bedingung Fachwissen irrelevant: Hinweis auf die Hintergrundinformationen, die über einen Button zu öffnen sind*

Fallgeschichte
Sie erhalten nun die Fallgeschichte, indem Sie ebenfalls am unteren Rand der Anwendung auf den Button FALL klicken. Bitte lesen Sie sich auch diese Informationen aufmerksam durch, bevor Sie über (X) zum Fragebogen zurückkehren. Sie haben im weiteren Verlauf der Studie jederzeit die Möglichkeit, diese Fallgeschichte erneut aufzurufen.

Abbildung C-9. *Bedingungsinvarianter Hinweis auf die Fallbeschreibung, die über einen Button zu öffnen ist*

Nachdem Sie nun die ersten Informationen über den Fall erhalten haben, möchten wir Sie bitten, die folgenden Fragen zu beantworten:

Der Hausarzt hat die Patientin mit dem Verdacht auf eine **Bulimie-Erkrankung** überwiesen. Wie sieht Ihre derzeitige, d.h. vorläufige Einschätzung vor dem Hintergrund der vorliegenden Fallinformationen aus? Für wie wahrscheinlich halten Sie die angesprochene Bulimie-Erkrankung?

Bitte geben Sie Ihre Einschätzung ab, indem Sie eine der Ziffern auswählen. Bitte stufen Sie Ihre Einschätzung zwischen 0=sehr unwahrscheinlich und 6=sehr wahrscheinlich ab.

sehr unwahrscheinlich 0 1 2 3 4 5 6 sehr wahrscheinlich

Abbildung C-10. Bedingung Fachwissen relevant: Abfrage der Überzeugtheit von der Prüfhypothese

Nachdem Sie nun die ersten Informationen über den Fall erhalten haben, möchten wir Sie bitten, die folgenden Fragen zu beantworten:

Die Staatsanwaltschaft geht davon aus, dass Herr Wagner schuldig ist, den Brand in der Wohnung von Frau Bauer, in dessen Folge Katharina Bauer zu Tode kam, vorsätzlich gelegt zu haben. Wie sieht Ihr derzeitiger, d.h. vorläufiger Verdacht vor dem Hintergrund der vorliegenden Fallinformationen aus? Für wie wahrscheinlich halten Sie die von der Staatsanwaltschaft vertretene Annahme, dass Herr Wagner schuldig sei?

Bitte geben Sie Ihre Einschätzung ab, indem Sie eine der Ziffern auswählen. Bitte stufen Sie Ihre Einschätzung zwischen 0=sehr unwahrscheinlich und 6=sehr wahrscheinlich ab.

sehr unwahrscheinlich 0 1 2 3 4 5 6 sehr wahrscheinlich

Abbildung C-11. Bedingung Fachwissen irrelevant: Abfrage der Überzeugtheit von der Prüfhypothese

In der ersten Phase der Urteilsbildung ist es wichtig, sich Gedanken über mögliche Symptomkonstellationen bzw. mögliche Informationsmuster zu machen. Wir möchten nun an dieser Stelle erfahren, welche Bezüge Sie herstellen und bitten Sie, dazu diese erste Aufgabe zu bearbeiten.

Modelle der Informationsverarbeitung - z.B. von Trope und Liberman (1996) - gehen davon aus, dass Beurteiler eingangs so genannte *Evidenzregeln* bilden. Damit ist gemeint, dass Personen bewusst oder unbewusst *Wenn-Dann*-Regeln aufstellen, die Vorhersagen darüber beinhalten, wann eine mögliche Erklärung (in diesem Fall eine Diagnose) bestätigt oder widerlegt wird. Eine solche Evidenzregel beinhaltet demnach die Forderung, dass bestimmte Gegebenheiten vorliegen müssen oder aber gerade nicht gegeben sein dürfen, damit eine vorgegebene Vermutung zutrifft.

Zur Veranschaulichung hier ein alltagspsychologisches Beispiel: Student A möchte herausfinden, ob Student B feindselig ist. Mögliche Evidenzregeln lauten wie folgt:

- Wenn B feindselig ist, dann müsste er schon auf kleine Provokationen sehr ärgerlich und unangemessen reagieren.
- Wenn B feindselig ist, dann dürfte er es nicht kommentarlos hinnehmen, dass jemand anderes ihn versehentlich anrempelt.

Ihre Aufgabe wird es nun sein, 6 Evidenzregeln aufzustellen, die Sie für die Beurteilung des Falles hilfreich finden. Überlegen Sie sich bitte, welche Symptomkonstellation Sie erwarten bzw. nicht erwarten würden. Geben Sie dazu immer an, welches Störungsbild Sie in Betracht ziehen (Wenn-Passage) und welche Datenlage (Dann-Passage) Sie erwarten.

z.B.	Wenn die Person feindselig ist (Wenn-Passage),	dann sollte sie schon bei Kleinigkeiten aggressiv reagieren (Dann-Passage).
1.		
2.		
3.		

Abbildung C-12. Bedingung Fachwissen relevant: Abfrage der AV 3 (Generieren von Evidenzregeln)

In der ersten Phase der Urteilsbildung ist es wichtig, sich Gedanken über mögliche Beweiskonstellationen bzw. mögliche Informationsmuster zu machen. Wir möchten nun an dieser Stelle erfahren, welche Bezüge Sie herstellen und bitten Sie, dazu diese erste Aufgabe zu bearbeiten.

Modelle der Informationsverarbeitung - z.B. von Trope und Liberman (1996) - gehen davon aus, dass Beurteiler eingangs so genannte *Evidenzregeln* bilden. Damit ist gemeint, dass Personen bewusst oder unbewusst *Wenn-Dann-Regeln* aufstellen, die Vorhersagen darüber beinhalten, wann eine mögliche Erklärung (in diesem Fall die Schuld bzw. Unschuld) bestätigt oder widerlegt wird. Eine solche Evidenzregel beinhaltet demnach die Forderung, dass bestimmte Gegebenheiten vorliegen müssen oder aber gerade nicht gegeben sein dürfen, damit eine vorgegebene Vermutung zutrifft.

Zur Veranschaulichung hier ein alltagspsychologisches Beispiel: Student A möchte herausfinden, ob Student B feindselig ist. Mögliche Evidenzregeln lauten wie folgt:

- Wenn B feindselig ist, dann müsste er schon auf kleine Provokationen sehr ärgerlich und unangemessen reagieren.
- Wenn B feindselig ist, dann dürfte er es nicht kommentarlos hinnehmen, dass jemand anderes ihn versehentlich anrempelt.

Ihre Aufgabe wird es nun sein, 6 Evidenzregeln aufzustellen, die Sie für die Beurteilung des Falles hilfreich finden. Überlegen Sie sich bitte, welches Informationsmuster Sie erwarten bzw. nicht erwarten würden. Geben Sie dazu immer an, welche Schuldannahme(Schuldig vs. Unschuldig) Sie in Betracht ziehen (Wenn-Passage) und welche Datenlage (Dann-Passage) Sie erwarten.

z.B.	Wenn die Person schuldig [unschuldig] ist (Wenn-Passage),	dann sollte sie... (Dann-Passage)
1.		
2.		
3.		
4.		

Abbildung C-13. Bedingung Fachwissen irrelevant: Abfrage der AV 3 (Generieren von Evidenzregeln)

Sie erhalten nun weitere Informationen über die Patientin, die im Rahmen der Diagnostik erhoben worden sind. Wir möchten Sie bitten, sich die Angaben genau anzuschauen und dann hinsichtlich ihrer Bedeutsamkeit zu bewerten.

Bitte erinnern Sie sich an die Verdachtsdiagnose „Bulimie“ des Arztes. Im Folgenden präsentieren wir Ihnen weitere Informationen, die Sie bitte vor dem Hintergrund dieser Annahme beurteilen. Zunächst wollen wir von Ihnen wissen, wie wichtig Sie die dargestellte Information für die Beurteilung des Falls halten. Verwenden Sie dafür bitte die Skala von 0=*gar nicht wichtig* bis 6=*sehr wichtig*, indem Sie eine der Ziffern ankreuzen.

Im Anschluss daran sollen Sie bewerten, wie die genannte Information die Wahrscheinlichkeit des Zutreffens der Annahme des Arztes beeinflusst. Sie können Ihre Einschätzung dabei abstufen zwischen -3 = *sehr starke Abschwächung der Annahme des Arztes*, 0 = *kein Einfluss* und +3 = *sehr starke Bekräftigung der Annahme des Arztes*. Bitte gehen Sie dabei immer von der anfänglichen Einschätzung des Arztes aus und bewerten Sie die Informationen isoliert – damit meinen wir, dass Sie jede einzelne Information für sich genommen betrachten und Ihren alleinigen Einfluss angeben sollen.

Abbildung C-14. Bedingung Fachwissen relevant: Instruktionen zur AV 4 (Bestätigungstendenz B)

Sie erhalten nun weitere Informationen über den Angeklagten, die im Rahmen der Ermittlungen erhoben worden sind. Wir möchten Sie bitten, sich die Angaben genau anzuschauen und dann hinsichtlich ihrer Bedeutsamkeit zu bewerten. Bitte erinnern Sie sich an die Schuldannahme seitens der Staatsanwaltschaft. Im Folgenden präsentieren wir Ihnen weitere Informationen, die Sie bitte vor dem Hintergrund dieser Annahme beurteilen. Zunächst wollen wir von Ihnen wissen, wie wichtig Sie die dargestellte Information für die Beurteilung des Falls halten. Verwenden Sie dafür bitte die Skala von 0=*gar nicht wichtig* bis 6=*sehr wichtig*, indem Sie eine der Ziffern ankreuzen.

Im Anschluss daran sollen Sie bewerten, wie die genannte Information die Wahrscheinlichkeit des Zutreffens der Annahme der Staatsanwaltschaft beeinflusst. Sie können Ihre Einschätzung dabei abstufen zwischen -3 = *sehr starke Abschwächung der Annahme der Staatsanwaltschaft*, 0 = *kein Einfluss* und +3 = *sehr starke Bekräftigung der Annahme der Staatsanwaltschaft*. Bitte gehen Sie dabei immer von der anfänglichen Einschätzung der Staatsanwaltschaft aus und bewerten Sie die Informationen isoliert – damit meinen wir, dass Sie jede einzelne Information für sich genommen betrachten und Ihren alleinigen Einfluss angeben sollen.

Abbildung C-15. Bedingung Fachwissen irrelevant: Instruktionen zur AV 4 (Bestätigungstendenz B)

Information							
Die Patientin gab an, dass sie in ihrer Freizeit auf einem Ponyhof mithilfe.							
Wichtigkeit der Information für den Fall							
	gar nicht						sehr
	0	1	2	3	4	5	6
	<input type="radio"/>						
Abschwächung oder Bekräftigung der Bulimie-Annahme des Arztes							
sehr starke Abschwächung			kein Einfluss				sehr starke Bekräftigung
-3	-2	-1	0	+1	+2	+3	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Abbildung C-16. *Bedingung Fachwissen relevant: Exemplarische Bildschirmansicht zur Abfrage der AV 4 (Bestätigungstendenz B), jede der 18 Informationen wurde auf einem eigenen Bildschirm dargeboten*

Information							
Frau Bauer gab an, Herr Wagner habe regelmäßig auf Katharina Acht gegeben.							
Wichtigkeit der Information für den Fall							
	gar nicht						sehr
	0	1	2	3	4	5	6
	<input type="radio"/>						
Abschwächung oder Bekräftigung der Schuld-Annahme der Staatsanwaltschaft							
sehr starke Abschwächung			kein Einfluss				sehr starke Bekräftigung
-3	-2	-1	0	+1	+2	+3	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Abbildung C-17. *Bedingung Fachwissen irrelevant: Exemplarische Bildschirmansicht zur Abfrage der AV 4 (Bestätigungstendenz B), jede der 18 Informationen wurde auf einem eigenen Bildschirm dargeboten*

Wir möchten Sie nun bitten, noch weitere Informationen zu bewerten. An dieser Stelle ist es uns wichtig, dass Sie einen bestimmten Blickwinkel einnehmen. Verschiedene Teilnehmer der Befragung werden dabei gebeten, unterschiedliche Perspektiven einzunehmen. Der Computer wird dies nun zufällig für Sie festlegen. Bitte klicken Sie nun auf "weiter", dann erscheint die Störung, vor deren Hintergrund Sie die nachfolgend genannten Informationen bitte bewerten.

Abbildung C-18. *Bedingungsinvariant: Täuschung darüber, dass der PC nun zufällig einen Blickwinkel auswählte, der im Folgenden zu beachten sei*

Ihr Blickwinkel: Bulimie
Stellen Sie sich für die Bewertung der folgenden Informationen jeweils die Frage: "Welche Bedeutung hat diese Angabe, wenn Marisa unter einer Bulimie leidet?"

Abbildung C-19. *Bedingung Fachwissen relevant: Hinweis, welcher Blickwinkel ausgewählt wurde. Alle Probanden erhielten den Blickwinkel Bulimie.*

Ihr Blickwinkel: **Schuldig**

Stellen Sie sich für die Bewertung der folgenden Informationen jeweils die Frage: "Welche Bedeutung hat diese Angabe, wenn Herr Wagner **schuldig** ist?"

Abbildung C-20. Bedingung Fachwissen irrelevant: Hinweis, welcher Blickwinkel ausgewählt wurde. Alle Probanden erhielten den Blickwinkel schuldig.

Zunächst geben wir Ihnen jedoch noch einige allgemeine Informationen: Für jede angegebene Information wollen wir wissen, ob Sie sie als eindeutigen Beleg, brauchbaren Hinweis oder Widerspruch für die oben angegebene Annahme bewerten. Sie können jedoch auch angeben, dass Sie die Information als völlig irrelevant halten.

Ein eindeutiger Beleg zeichnet sich dadurch aus, dass die betreffende Angabe vorliegen muss, damit die Diagnose gerechtfertigt ist. Soll überprüft werden, ob eine Person extravertiert ist, könnte der Befund „Person A erreicht auf der Skala Extraversion mit einem PR=95 ein überdurchschnittliches Ergebnis“ einen eindeutigen Beleg darstellen.

Ein brauchbarer Hinweis zeichnet sich dadurch aus, dass die betreffende Angabe/ das betreffende Symptom vorliegen kann, wenn die Diagnose gerechtfertigt ist, obwohl es nicht ausreicht, dass dieses Symptom vorliegt. Es kommt gehäuft vor, jedoch nicht notwendigerweise. Die Information „Person A wurde auf einer Party gesehen“ wäre im Zusammenhang mit dem Extraversionsbeispiel ein Hinweis.

Ein Widerspruch zeichnet sich dadurch aus, dass die betreffende Angabe den Vorgaben für die Diagnose widerspricht, bzw. gewöhnlich nicht vorliegt, wenn jemand eine bestimmte Erkrankung hat. Die Beobachtung „Person A kam nicht zur Party, um zuhause ein Buch zu lesen“ wäre ein Widerspruch.

Eine irrelevante Information besitzt keinerlei Zusammenhang zu der Diagnose – ob sie vorliegt oder nicht, hat keinen Einfluss auf die Störung. Dass Person A am liebsten Erdbeereis mag, ist irrelevant.

Abbildung C-21. Bedingung Fachwissen relevant: Instruktionen zur AV 5a + AV 5b.

Zunächst geben wir Ihnen jedoch noch einige allgemeine Informationen: Für jede angegebene Information wollen wir wissen, ob Sie sie als eindeutigen Beleg, brauchbaren Hinweis oder Widerspruch für Ihren Blickwinkel bewerten. Sie können jedoch auch angeben, dass Sie die Information als irrelevant erachten.

Ein eindeutiger Beleg zeichnet sich dadurch aus, dass die betreffende Angabe vorliegen muss, damit das Urteil gerechtfertigt ist. Soll überprüft werden, ob eine Person extravertiert ist, könnte der Befund „Person A erreicht auf der Skala Extraversion ein überdurchschnittliches Ergebnis“ einen eindeutigen Beleg darstellen.

Ein brauchbarer Hinweis zeichnet sich dadurch aus, dass die betreffende Angabe vorliegen kann, wenn das Urteil gerechtfertigt ist, obwohl es nicht ausreicht, dass dieses Merkmal vorliegt. Es kommt gehäuft vor, jedoch nicht notwendigerweise. Die Information „Person A wurde auf einer Party gesehen“ wäre im Zusammenhang mit dem Extraversionsbeispiel ein Hinweis.

Ein Widerspruch zeichnet sich dadurch aus, dass die betreffende Angabe Ihrem Blickwinkel widerspricht, bzw. gewöhnlich nicht vorliegt, wenn die Annahme in Ihrem Blickwinkel zutrifft. Die Beobachtung „Person A kam nicht zur Party, um zuhause ein Buch zu lesen“ wäre ein Widerspruch.

Eine irrelevante Information besitzt keinerlei Zusammenhang zu dem Urteil – ob sie vorliegt oder nicht, hat keinen Einfluss auf die Schuldfrage. Dass Person A am liebsten Erdbeereis mag, ist irrelevant.

Abbildung C-22. Bedingung Fachwissen irrelevant: Instruktionen zur AV 5a + AV 5b.

Zur Erinnerung:
 Beleg: Angabe **muss** vorliegen, Hinweis: Angabe **kann** vorliegen, Widerspruch: Angabe **darf nicht** vorliegen, Irrelevant: Angabe hat **keinen Zusammenhang**

Information

Die Patientin hatte kein Normalgewicht, denn bei einer Körpergröße von 1,75 m und einem Gewicht von 50 kg lag ihr BMI bei 16,3.

Bedeutung

	eindeutiger Beleg	brauchbarer Hinweis	Widerspruch	irrelevant
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Abbildung C-23. Bedingungsinvariant: Exemplarische Bildschirmansicht zur Abfrage der AV 5a + AV 5b, jede der 12 selbstverständlich fallspezifischen Informationen wurde auf einem eigenen Bildschirm dargeboten.

Sie haben nun alle Informationen erhalten, die über die Patientin verfügbar sind.

Geben Sie deshalb bitte jetzt an, welche Störung Sie in diesem Fall diagnostizieren würden und nehmen Sie eine Einschätzung Ihrer subjektiven Sicherheit vor. Stellen Sie die Diagnose mit Hilfe der DSM-Klassifikationen und nennen Sie ggf. den Subtypen.

Ich stelle folgende Diagnose:

Abbildung C-24. *Bedingung Fachwissen relevant: Abfrage der Schlussfolgerung (AV 1)*

Sie haben nun alle Informationen erhalten, die über den Angeklagten verfügbar sind.

Geben Sie deshalb bitte jetzt an, welches Urteil Sie in diesem Fall abgeben würden und nehmen Sie eine Einschätzung Ihrer subjektiven Sicherheit vor.

Ich fälle folgendes Urteil:

Abbildung C-25. *Bedingung Fachwissen irrelevant: Abfrage der Schlussfolgerung (AV 1)*

Bitte geben Sie an, wie sicher Sie sich bei Ihrer Entscheidung sind

sehr unsicher	0	1	2	3	4	5	6	sehr sicher
	<input type="radio"/>							

Abbildung C-26. *Bedingungsinvariante Abfrage der subjektiven Sicherheit (AV 2)*

Bitte überlegen Sie, welches Thema Sie in der ersten Therapiesitzung mit der Patientin besprechen würden:

Bitte geben Sie an, was für eine Hausaufgabe Sie der Patientin geben würden:

Abbildung C-27. *Bedingung Fachwissen relevant: Abfrage der Themen für die erste Therapiesitzung und zu verordnenden Hausaufgaben (AV 6)*

Zum Abschluss benötigen wir noch einige persönlichen Angaben von Ihnen:

Wie alt sind Sie? Jahre

Ihr Geschlecht:

männlich

weiblich

In welchem Bundesland leben Sie?

Baden-Württemberg

Bayern

Berlin

Brandenburg

Bremen

Hamburg

Hessen

Mecklenburg-Vorpommern

Niedersachsen

Nordrhein-Westfalen

Rheinland-Pfalz

Saarland

Sachsen

Sachsen-Anhalt

Schleswig-Holstein

Thüringen

außerhalb Deutschlands

Abbildung C-28. *Bedingungsinvariante Abfrage von Altern, Geschlecht und Wohnort*

<p>Wie ist Ihr Studium organisiert?</p> <p><input type="radio"/> Diplom-Studiengang</p> <p><input type="radio"/> Bachelor-Master-Studiengang</p>
<p>In welchem Fachsemester studieren Sie? (Wenn Sie im Master studieren, zählen Sie bitte auch die Bachelor-Semester hinzu, d. h. bilden Sie die Gesamtzahl!)</p> <p><input type="text"/> Semester</p>
<p>Welche Schwerpunkte haben Sie in Ihrem Studium gewählt? (Mehrfachwahl möglich)</p> <p><input type="checkbox"/> Klinische Psychologie</p> <p><input type="checkbox"/> Pädagogische Psychologie</p> <p><input type="checkbox"/> Arbeits- und Organisationspsychologie</p>
<p>weitere/andere Studienschwerpunkte:</p> <p><input type="text"/></p>
<p>Haben Sie bereits Praktika abgeleistet?</p> <p><input type="radio"/> Ja</p> <p><input type="radio"/> Nein</p> <p>Wenn „Ja“, in welchem Bereich haben sie Ihre Praktika gemacht?</p> <p><input type="text"/></p>
<p>Haben Sie bereits Veranstaltungen im Bereich „Klinische Psychologie“ besucht?</p> <p><input type="radio"/> Ja</p> <p><input type="radio"/> Nein</p> <p>Wenn „Ja“, dann: Welche Veranstaltungen haben Sie in diesem Bereich besucht? (Vorlesungen, Seminare, Fallstudien etc.)</p> <p><input type="text"/></p>
<p>Haben Sie die Abschlussprüfung (Diplom- oder Masterprüfung) im Fach „Klinische Psychologie“ bereits (studienbegleitend) absolviert?</p> <p><input type="radio"/> Ja</p> <p><input type="radio"/> Nein</p>
<p>Haben Sie bereits Veranstaltungen im Bereich „Psychologischer Diagnostik“ besucht?</p> <p><input type="radio"/> Ja</p> <p><input type="radio"/> Nein</p> <p>Wenn „Ja“, dann: Welche Veranstaltungen haben Sie in diesem Bereich besucht? (Allg. Vorlesung, Testtheorie, spezielle Testverfahren, Gutachtenkurse etc.)</p> <p><input type="text"/></p>
<p>Haben Sie die Abschlussprüfung (Diplom- oder Masterprüfung) im Fach „Psychologische Diagnostik“ bereits (studienbegleitend) absolviert?</p> <p><input type="radio"/> Ja</p> <p><input type="radio"/> Nein</p>
<p>Waren Sie schon mal an der Erstellung von Gutachten beteiligt (z.B. als Teil eines Praktikums oder eines Fallseminars)?</p> <p><input type="radio"/> Ja</p> <p><input type="radio"/> Nein</p> <p>Wenn „ja“, dann: In was für einem Bereich haben Sie das Gutachten verfasst? (z.B. Neuropsychologie, Kinder- und Jugendpsychiatrie, Pädagogische Psychologie etc.)</p> <p><input type="text"/></p>

Abbildung C-29. Abfrage der biographischen Daten, wenn auf dem Begrüßungsbildschirm angegeben wurde, Studierender zu sein

Sind Sie Psychologische/r Psychotherapeut/in oder Kinder- und Jugendlichenpsychotherapeut(in)?

Psychologische/r Psychotherapeut/in

Kinder- und Jugendlichenpsychotherapeut/in

Vor wie vielen Jahren haben Sie Ihre Approbation erhalten?

vor Jahren

Vor wie vielen Jahren haben Sie ihr Diplom in Psychologie erhalten?

vor Jahren

Welcher Therapieschule gehören Sie bzw. die Einrichtung, in der Sie arbeiten, an?

Psychoanalyse/ Tiefenpsychologie

Verhaltenstherapie

Anderen:

Abbildung C-30. Abfrage der biographischen Daten, wenn auf dem Begrüßungsbildschirm angegeben wurde, Psychologischer Psychotherapeut und/oder Kinder- und Jugendlichenpsychotherapeut zu sein

In welchem Gebiet haben Sie Ihre Facharztausbildung absolviert?

Facharzt für Kinder- und Jugendpsychiatrie und –psychotherapie

Facharzt für Psychiatrie und Psychotherapie

Facharzt für Psychosomatische Medizin und Psychotherapie

Noch keine Facharztausbildung absolviert

Andere Richtung (bitte angeben):

Welcher Therapieschule gehören Sie bzw. die Einrichtung, in der Sie arbeiten, an?

Psychoanalyse/ Tiefenpsychologie

Verhaltenstherapie

Anderen:

Abbildung C-31. Abfrage der biographischen Daten, wenn auf dem Begrüßungsbildschirm angegeben wurde, Facharzt zu sein

Wann haben Sie Ihr Diplom/Examen erhalten?
Jahr (YYYY): <input type="text"/>
Seit wann sind Sie im Klinischen Bereich tätig?
Jahr (YYYY): <input type="text"/>
Haben Sie einmal eine Therapieausbildung begonnen und nicht beendet?
<input type="radio"/> Ja, zur/zum Psychologische/r Psychotherapeut/in <input type="radio"/> Ja, zur/zum Kinder- und Jugendlichenpsychotherapeut/in <input type="radio"/> Nein
Können Sie das theoretische Fundament Ihre Tätigkeit einer der folgenden Therapieschulen zuordnen?
<input type="radio"/> Psychoanalyse/ Tiefenpsychologie <input type="radio"/> Verhaltenstherapie <input type="radio"/> keiner bestimmten <input type="radio"/> Anderen:
<input type="text"/>

Abbildung C-32. Abfrage der biographischen Daten, wenn auf dem Begrüßungsbildschirm angegeben wurde, als Diplom-Psychologe ohne Therapieausbildung im klinischen Bereich tätig zu sein

Bitte beschreiben Sie die Hauptinhalte Ihrer Arbeit und beantworten Sie dafür die folgenden Fragen!
Arbeiten Sie hauptsächlich mit Erwachsenen oder Kindern und Jugendlichen?
<input type="radio"/> Erwachsene <input type="radio"/> Kinder / Jugendliche <input type="radio"/> beide gleichermaßen
Mit welchen Störungsbildern haben Sie hauptsächlich Kontakt?
<input type="text"/>
Wie viele Stunden arbeiten Sie pro Woche?
<input type="text"/> Stunden
Wie viele Stunden davon arbeiten Sie psychotherapeutisch?
<input type="text"/> Stunden
Wie viele Stunden davon arbeiten Sie psychodiagnostisch?
<input type="text"/> Stunden

Abbildung C-33. Abfrage zur Charakterisierung des Arbeitsumfelds, wenn auf dem Begrüßungsbildschirm angegeben wurde, Psychologischer Psychotherapeut und/oder Kinder- und Jugendlichenpsychotherapeut bzw. Facharzt bzw. Diplom-Psychologe ohne Therapieausbildung im klinischen Bereich zu sein

Bitte geben Sie zum Schluss noch einmal in eigenen Worten an, was das Ziel dieser Studie ist!

Abbildung C-34. *Bedingungsinvariante Abfrage zur SV Bewusstsein hinsichtlich konfirmatorischer Prozesse*

Vielen herzlichen Dank für Ihre Mitarbeit!

Sie haben den Fragebogen bis zum Ende ausgefüllt und wurden als Teilnehmer/-in gezählt. Als symbolisches Dankeschön spenden wir daher weitere 2,50€ an die Stiftung Kinderherz Deutschland, die die Lebenschancen und die Lebensqualität herzkranker Kinder verbessern will.

Auf der nächsten Seite finden Sie noch einige Informationen zu den Studienhintergründen.



Abbildung C-35. *Bedingungsinvariante Danksagung und Hinweis auf die Erhöhung des Spendenzählers*

Informationen über die Studie

Liebe Teilnehmerin, lieber Teilnehmer!

An dieser Stelle möchten wir Ihnen abschließend noch einen kleinen Einblick in die Hintergründe und Ziele unserer Studie geben.

Basis des Forschungsprojekts OPUS bildet die bereits 1986 von Leary und Miller postulierte Annahme, dass Therapeuten (bzw. allgemeiner: Diagnostiker) nicht einfach nur passive Rezipienten der Schilderungen ihrer Patienten sind, sondern aktiv nach Informationen suchen und diese anschließend integrieren. Daraus lässt sich unmittelbar ableiten, dass der Faktor „Mensch“ einen entscheidenden Einfluss auf die psychologische Urteilsbildung ausübt, der sich auch nachteilig auswirken kann. So konnte vielfach gezeigt werden, dass unterschiedliche Beurteiler bei identischer Datenlage zu anderen Schlussfolgerungen gekommen sind. Mithilfe der gewonnenen Daten wollen wir die folgenden Forschungsfragen analysieren:

- Welchen Einfluss hat die Berufserfahrung eines Diagnostikers (in Jahren) auf die Urteilsstrategien und das gefällte Urteil? Gibt es spezifische Vorteile von Berufserfahrenen, die sich in die Ausbildung von Diagnostikern und Therapeuten integrieren ließen? Aus diesem Grund haben wir sowohl Studierende und Psychologen in Ausbildung als auch Psychotherapeuten um ihre Mitwirkung gebeten.
- Welchen Einfluss hat das spezifische Fachwissen auf die Urteilsstrategien und das gefällte Urteil? Gibt es spezifische Fehler, die fachfremde Personen begehen, während diese Spezialisten nicht unterlaufen? Dafür haben wir Fälle aus unterschiedlichen Kontexten eingesetzt, wobei jeder Teilnehmer nur einen einzigen bearbeitet hat. Neben dem klinischen Kontext, in dem angehende sowie berufstätige Psychologen über Spezialwissen verfügen, kam auch ein kriminalistischer Kontext für einen Teil der Befragten vor, der für Psychologen deutlich fachfremd ist. Hierbei ist v.a. von Interesse, ob das Urteilsverhalten einer (prinzipiell diagnostisch geschulten) Teilnehmergruppe sich ändert bzw. ebenso valide ist, wenn kaum auf Fachwissen zurückgegriffen werden kann.
- Welchen Einfluss hat der Eindruck, den ein Diagnostiker sich eingangs von einem Fall macht? Die dritte Frage beschäftigt sich daher mit dem Urteilsprozess innerhalb einer Person über den gesamten Zeitraum der Analyse. Hierbei ist interessant, wie die vorgegebenen Informationen vor den möglicherweise unterschiedlichen Bezugsrahmen interpretiert werden, die aus dem ersten Eindruck entstehen. An einer Stelle der Befragung haben wir jedoch einen gemeinsamen Bezugspunkt innerhalb der jeweiligen Fälle geschaffen: Anders als angekündigt hat es keine echte Zufallsauswahl gegeben, als Sie Informationen hinsichtlich Ihrer Bedeutsamkeit einschätzen sollten. Sondern vielmehr hat der PC für alle Personen mit der selben Fallbeschreibung die selbe Annahme vorgegeben. Für diesen Schritt haben wir uns entschieden, da die standardmäßige Vorgabe eines festen Blickwinkels in der Befragungssituation möglicherweise Misstrauen bzw. Zurückhaltung hervorgerufen hätte, die jedoch nicht aufgetreten wären, wenn der Fall in der Realität beurteilt worden wäre.

Wir möchten Sie an dieser Stelle noch einmal bitten, Ihre Kollegen auf diese Studie aufmerksam zu machen und ihnen die Einladung bzw. den Link und das Passwort zu dieser Seite zu schicken. Allerdings liegt es natürlich in unserem und auch in Ihrem Interesse, dass Sie ihre eigenen Erfahrungen mit dem Fall und Ihr Wissen zu den Hintergründen der Studie nicht weitergeben. Denn in diesem Fall wären die Ergebnisse der Untersuchung nicht reliabel und Ihre eigene Zeit und Arbeit, die Sie in die Bearbeitung der Fallgeschichte eingebracht haben, wären vergebens gewesen. Außerdem werden den Teilnehmern wie oben erwähnt verschiedene Fallgeschichten dargeboten, so dass Kollegen eventuell einen anderen Fall bearbeiten.

Vielen Dank für Ihr Verständnis und Ihre Kooperation!

Abbildung C-36. *Bedingungsinvariante Aufklärung und Verabschiedung der Studienteilnehmer*

Anhang D - Ratingmaterial für das OPUS

Ratingmaterial.....	295
---------------------	-----

RATINGMANUAL FÜR DAS OPUS

DIE EXPERTENSTUDIE ZUR DIAGNOSTISCHEN URTEILSBILDUNG

Das folgende Ratingmanual bezieht sich auf eine Studie am Institut für Psychologie der CAU Kiel zur diagnostischen Urteilsbildung. Die Teilnehmer erhielten eine Fallgeschichte und wurden gebeten bestimmte Schritte der diagnostischen Urteilsbildung auszuführen. Einige dieser Aufgaben beinhalteten das Verfassen von Freitextantworten, die nun beurteilt werden müssen, bevor sie statistisch ausgewertet werden können.

Die folgende Anleitung macht Euch mit den Aufgabenstellungen vertraut, die die Teilnehmer bearbeiten mussten und gibt konkrete Anweisungen mit Beispielen, wie die Antworten zu bewerten sind.

Es gab insgesamt 2 Fallgeschichten (*Bulimie* oder *Brandstiftung*). Einmal ging es um eine Patientin mit Verdacht auf eine Bulimie-Erkrankung und in der anderen Beschreibung um einen Angeklagten in einem Brandstiftungsfall.

Am Ende des Manuals befinden sich die Fall- und Hintergrundinformationen, die den Teilnehmern zu Verfügung standen. DIESE Angaben müssen also zusätzlich zu den vorgegebenen Kodieranweisungen als Informationsgrundlage für die Beurteilung der Antworten der Probanden herangezogen werden.

GÜTE DER EVIDENZREGELN (AV 3)

Nachdem die Teilnehmer ihrerseits die Hintergrundinformationen und die Fallbeschreibung gelesen hatten, begannen die Aufgaben. Die erste Aufgabe bezog sich auf so genannte Evidenzregeln. Die Teilnehmer bekamen zur Erläuterung die folgenden Informationen:

INSTRUKTION FÜR DIE VPn | AUSZUG AUS DEM STUDIENMATERIAL

In der ersten Phase der Urteilsbildung ist es wichtig, sich Gedanken über mögliche Symptomkonstellationen bzw. mögliche Informationsmuster zu machen. Wir möchten nun an dieser Stelle erfahren, welche Bezüge Sie herstellen und bitten Sie, dazu diese erste Aufgabe zu bearbeiten.

Modelle der Informationsverarbeitung, z.B. von Trope und Liberman (1996) gehen davon aus, dass Beurteiler eingangs so genannte Evidenzregeln bilden. Damit ist gemeint, dass Personen bewusst oder unbewusst „Wenn-Dann“-Regeln aufstellen, die Vorhersagen darüber beinhalten, wann eine mögliche Erklärung bestätigt oder widerlegt wird. Eine solche Evidenzregel beinhaltet demnach die Forderung, dass bestimmte Gegebenheiten vorliegen müssen oder aber gerade nicht gegeben sein dürfen, damit eine vorgegebene Vermutung zutrifft.

Zur Veranschaulichung hier ein alltagspsychologisches Beispiel: Student A möchte herausfinden, ob Student B feindselig ist. Mögliche Evidenzregeln lauten wie folgt:

- Wenn B feindselig ist, dann müsste er schon auf kleine Provokationen sehr ärgerlich und unangemessen reagieren.
- Wenn B feindselig ist, dann dürfte er es nicht kommentarlos hinnehmen, dass jemand anderes ihn versehentlich anrempelt.

Ihre Aufgabe wird es nun sein, 6 Evidenzregeln aufzustellen, die Sie für die Beurteilung des Falles hilfreich finden. Überlegen Sie sich bitte, welche Symptomkonstellation Sie erwarten bzw. nicht erwarten würden. Geben Sie dazu immer an, welches Störungsbild/welches Ereignis/ welche Annahme Sie in Betracht ziehen (Wenn-Passage) und welche Datenlage (Dann-Passage) Sie erwarten.

Erinnern Sie sich dabei an das Beispiel: *Wenn die Person feindselig ist (Wenn-Passage), dann sollte sie schon bei Kleinigkeiten aggressiv reagieren (Dann-Passage).*

KODIERUNG ALLGEMEIN

Die Probanden haben also 6 entsprechende Evidenzregeln aufgestellt, die jeweils aus einer *Wenn*- und einer *Dann*-Passage bestehen. Diese Regeln sollen nun in Bezug auf 5 Merkmale beurteilt werden:

1. Instruktionskonformität
2. Positive Teststrategie
3. Diagnostizität der angesprochenen Merkmale
4. Berücksichtigung alternativer Hypothesen
5. Gewichtung alternativer Hypothesen

Diese Merkmale sollen sowohl für die Evidenzregeln zum Thema Bulimie als auch zur Brandstiftung bewertet werden. In der Folge werden für beide Fälle getrennt Beispiele gegeben, um die unterschiedliche Bedeutung in den verschiedenen Kontexten deutlich zu machen.

INSTRUKTIONSKONFORMITÄT

Leider wurden die Vorgaben von den Probanden nicht immer so umgesetzt, wie in der Instruktion beschrieben. Da die inhaltliche Bewertung i.d.R. nur dann sinnvoll ist, wenn die Instruktion gewahrt wurde, soll in einem ersten Schritt dieser Fakt überprüft werden.

Es gibt 5 Kategorien, in die die Evidenzregeln eingeordnet werden müssen:

- instruktionskonform
- Wenn- und Dann-Passage vertauscht
- keine explizite Störung/Vermutung angegeben
- nicht instruktionskonform, Restkategorie
- andere Störung [nur für den Bulimie-Fall relevant]

BEWERTUNG DER KONFORMITÄT IM BULIMIE-FALL

Eine Evidenzregel ist dann **instruktionskonform**, wenn in der WENN-Passage ein Störungsbild genannt wird und in der DANN-Passage ein zugehöriges Symptom/Syndrom oder eine Konsequenz. Beispiel: *Wenn eine Anorexie vorliegt, dann verweigert die Patientin das Essen.*

Eine Evidenzregel weist dann eine **vertauschte Wenn-/Dann-Passage** auf, wenn in der Wenn-Passage das Symptom und in der Dann-Passage das Störungsbild thematisiert wird. Beispiel: *Wenn die Patientin das Essen verweigert, dann liegt eine Anorexie vor.*

Eine Evidenzregel **ohne explizite Störung** könnte wie folgt aussehen: *Wenn die Mutter der Patientin eine Essstörung hatte, dann kann sich das auf die Patientin ausgewirkt haben.* Hier runter fallen also alle fachlichen Zusammenhänge, die keinen expliziten Störungsbezug haben. Außerdem fallen solche Evidenzregeln in diese Kategorie, in denen eine StörungsGRUPPE genannt wird, d.h. zum Beispiel *„Wenn eine Essstörung vorliegt, dann ...“*. Es ist unerheblich, ob die Störungsgruppe durch eine zusätzliche Nennung einer einzelnen oder mehrerer konkreter Störungen spezifiziert wird. Sobald eine Störungsgruppe dort steht, wird *ohne explizite Störung* kodiert. Mit Störungsgruppe ist auch gemeint, wenn mehrere Störungen genannt werden (ohne den übergreifenden Gruppennamen!), d.h. *„Wenn Anorexie oder Bulimie vorliegen, dann ...“*. Bitte auch hier einordnen.

Eine Evidenzregel ist als **nicht instruktionskonform** zu bewerten, wenn sie statt der Aufstellung von Beziehungen zwischen Störungen/Ereignissen einen Kommentar zur Aufgabenstellung an sich abgibt oder andere Informationen über den Probanden selbst enthält. Beispiel: *Wenn so wenige Informationen vorgegeben sind, kann man keine hilfreichen Regeln aufstellen.* Diese Kategorie ist als Restkategorie zu verwenden.

Eine Evidenzregel beinhaltet dann eine andere Störung, wenn nicht von Bulimie oder Anorexie oder einem der zugehörigen Subtypen gesprochen wird, z.B. *„Wenn eine Depression vorliegt, dann ...“*.

BEWERTUNG DER KONFORMITÄT IM BRANDSTIFTUNGS-FALL

Eine Evidenzregel ist dann **instruktionskonform**, wenn in der WENN-Passage die Schuld oder Unschuld des Angeklagten oder einer anderen Person genannt wird und in der DANN-Passage ein zugehöriges Argument. Beispiel: *Wenn Herr Wagner schuldig ist, dann hat er kein Alibi.*

Eine Evidenzregel weist dann eine **vertauschte Wenn-/Dann-Passage** auf, wenn in der Wenn-Passage das Argument und in der Dann-Passage die Schuld/Unschuld thematisiert wird. Beispiel: *Wenn Herr Wagner kein Alibi hat, ist er schuldig.*

Eine Evidenzregel **ohne explizite Vermutung von Schuld/Unschuld** könnte wie folgt aussehen: *Wenn Frau Bauer nicht so verantwortungslos wäre, dann hätte sie das Kind nicht einem Betrunkenen anvertraut.*

Eine Evidenzregel ist als **nicht instruktionskonform** zu bewerten, wenn ... s.o. [Restkategorie]

Gehört eine Evidenzregel der 3., 4. und 5. Kategorie an, so werden die Spalten für „positive Teststrategie“ und „Diagnostizität“ automatisch mit 999= *nicht auswertbar* kodiert.

POSITIVE TESTSTRATEGIE IM RAHMEN DER EVIDENZREGEL

Für die Beurteilung der Teststrategie, die mit der Evidenzregel verfolgt wird, muss beurteilt werden, in welchem Zusammenhang Wenn- und Dann-Passage stehen. Dies ist wiederum nur dann sinnvoll, wenn die Regel entweder *instruktionskonform* oder *„vertauscht“* ist.

Die folgenden Kodierungen sind möglich:

- 1 = positive Teststrategie
- 2 = negative Teststrategie
- 9 = nicht bewertbar [Restkategorie], nur sparsam verwenden

Eine positive Teststrategie zeigen die Probanden dann, wenn die Evidenzregeln so formuliert werden, dass durch sie das Vorliegen von Merkmalen abgefragt wird, die konsistent mit der eigenen Vermutung sind. In diesem Sinne wird nur auf Informationen geschaut, die das Vorliegen belegen könnten, während Merkmale, die das Vorliegen widerlegen könnten, nicht betrachtet werden. Hier zwei Beispiele aus der intuitiven Persönlichkeitspsychologie:

Beispiel 1: Eine Evidenzregel im Sinne einer positiven Teststrategie bei der Vermutung Extraversion könnte demnach wie folgt lauten: *„Wenn eine Person extravertiert ist, dann geht sie abends gerne unter Menschen.“*, da das Vorliegen dieser Information („abends gerne unter Menschen gehen“) mit dem Zutreffen der Vermutung („Extraversion“) assoziiert ist. Die Dann-Passage spricht ein Merkmal an, das man erwarten würde, wenn die Annahme zutrifft.

Keine positive, sondern eine negative Teststrategie würde vorliegen, wenn folgende Regel aufgestellt würde: *„Wenn eine Person extravertiert ist, dann wird sie abends nicht gerne alleine zuhause sein.“*, weil hier ein Merkmal in Betracht gezogen wird, das das Vorliegen von Extraversion widerlegen kann. Die Dann-Passage gibt zu verstehen, dass das angesprochene Merkmal gerade nicht vorliegen darf, wenn die Annahme gilt.

Beispiel 2: Eine Evidenzregel im Sinne einer positiven Teststrategie bei der Vermutung Feindseligkeit könnte demnach wie folgt lauten: *„Wenn eine Person feindselig ist, dann reagiert sie schon bei Kleinigkeiten aggressiv.“*, da das Vorliegen dieser Information mit dem Zutreffen der Vermutung assoziiert ist. Die Dann-Passage spricht ein Merkmal an, das man erwarten würde, wenn die Annahme zutrifft.

Keine positive, sondern eine negative Teststrategie würde vorliegen, wenn folgende Regel aufgestellt würde: *„Wenn eine Person feindselig ist, dann sollte sie bei einer Provokation nicht ruhig bleiben können.“*, weil hier ein Merkmal in Betracht gezogen wird, das das Vorliegen von Feindseligkeit widerlegen kann. Die Dann-Passage gibt zu verstehen, dass das angesprochene Merkmal gerade nicht vorliegen darf, wenn die Annahme gilt.

Bitte gebt nun für die Regeln an, inwieweit bei ihnen eine positive Teststrategie verfolgt wird. Überlegt Euch dafür bitte, inwieweit das Vorliegen des in der Evidenzregel betrachteten Merkmals für das Zutreffen der Hypothese spricht, die im Wenn-Part der Regel genannt wird. Ihr müsst also jeweils die Dann-Passage einer Regel zur Wenn-Passage in Beziehung setzen. Da es unerheblich ist, ob korrekter Weise die Störung/die Schuldvermutung in der Wenn-Passage und das Argument in der Dann-Passage steht, könnt ihr den Unterschied einfach ignorieren. Eine Unterscheidung für den Bulimie- und den Brandstiftungs-Fall ist ebenfalls irrelevant. Die Instruktion gilt gleichermaßen.

DIAGNOSTIZITÄT DER ANGESPROCHENEN MERKMALE IN DER DANN-PASSAGE

Mit der Diagnostizität der angesprochenen Merkmale ist die Möglichkeit gemeint, mit Hilfe der Inhalte der Dann-Passage der Evidenzregeln zwischen verschiedenen Annahmen zu unterscheiden. Abstrakt gesehen ist ein Merkmal genau dann diagnostisch, wenn es zwischen verschiedenen Alternativhypothesen (die in der Wenn-Passage stehen könnten) eine **unterschiedlich** große Auftretenswahrscheinlichkeit aufweist.

Die folgenden Kodierungen sind möglich:

- 0 = non-diagnostisch
- 1= diagnostisch
- 9 = nicht bewertbar [Restkategorie], nur sparsam verwenden

Es folgen 2 Beispiele aus der intuitiven Persönlichkeitspsychologie:

Beispiel 1: Wenn eine Person extravertiert ist, dann steht die betroffene Person vermutlich gerne auf Partys im Mittelpunkt. Ist die Person dagegen introvertiert ist es weniger wahrscheinlich, dass sie gerne im Mittelpunkt steht. Die Tatsache, dass eine Person gerne im Mittelpunkt steht ist demnach diagnostisch für die Unterscheidung zwischen den Merkmalen Introversion und Extraversion. Das in der Dann-Passage angesprochene Merkmal (im Mittelpunkt Stehen) ist sehr wahrscheinlich unter der einen Alternative der Wenn-Passage (Extraversion) und unwahrscheinlich unter der anderen (Introversion).

Spricht eine Evidenzregel hingegen das Merkmal „Lesen“ an, ist sie nondiagnostisch. Denn sowohl Personen, die extravertiert sind, als auch Personen, die introvertiert sind, lesen vielleicht gerne mal ein Buch, so dass das Merkmal nicht für eine Unterscheidung der beiden Charaktereigenschaften geeignet ist. Das in der Dann-Passage angesprochene Merkmal (Lesen) ist genauso wahrscheinlich unter der einen Wenn-Passage (Extraversion) wie unter der Alternative (Introversion).

Beispiel 2: Wenn eine Person feindselig ist, dann reagiert die betroffene Person schon bei geringen Anlässen aggressiv. Ist die Person dagegen freundlich, ist es weniger wahrscheinlich, dass sie bei Kleinigkeiten aggressiv reagiert. Die Tatsache, dass eine Person schnell aggressiv wird, ist demnach diagnostisch für die Unterscheidung zwischen den Merkmalen freundlich und feindselig. Das in der Dann-Passage angesprochene Merkmal (bei Kleinigkeiten aggressiv Reagieren) ist sehr wahrscheinlich unter der einen Alternative der Wenn-Passage (Feindseligkeit) und unwahrscheinlich unter der anderen (Freundlichkeit).

Spricht eine Evidenzregel hingegen das Merkmal an, dass eine Person auf eine gezielte Provokation aggressiv reagiert, ist der Inhalt der Dann-Passage nondiagnostisch. Denn bei einer gezielten Provokation können sowohl freundliche als auch feindselige Personen aggressiv reagieren, so dass das Merkmal nicht für eine Unterscheidung der beiden Charaktereigenschaften geeignet ist. Das in der Dann-Passage angesprochene Merkmal (Reaktion bei gezielter Provokation) ist genauso wahrscheinlich unter der einen Wenn-Passage (Feindseligkeit) wie unter der Alternative (Freundlichkeit).

Praktischer Hinweis: für die Bewertung der Diagnostizität ist bei instruktionskonformen Evidenzregeln folglich v.a. die Dann-Passage relevant. Die dort enthaltene Information muss hinsichtlich ihrer Auftretenswahrscheinlichkeit eingeschätzt werden. Bei „vertauschten Regeln“ ist es dann natürlich die Information in der Wenn-Passage.

BEWERTUNG FÜR DEN BULIMIE-FALL

Im Fall der Essstörungsgeschichte möchten wir Euch nun bitten, für die Evidenzregeln zu überprüfen, inwieweit die in der Dann-Passage angesprochenen Merkmale geeignet sind, zwischen den Diagnosen Bulimie und Anorexie vom Subtyp Binge-Eating/Purging zu unterscheiden. Das bedeutet, ihr sollt euch überlegen, ob sich die Wahrscheinlichkeiten, dass das angesprochene Symptom/Ereignis bei Bulimia nervosa oder Anorexia nervosa vom Subtyp Binge-Eating/ Purging vorliegt, unterscheiden. Die notwendigen Hintergrundinformationen bezüglich der Merkmale erhaltet Ihr durch den Text „Lehrbuchkapitel – Essstörungen“. Es ist von großer Wichtigkeit, dass ihr diesen Text genau studiert und im Zweifelsfall immer wieder zur Hand nehmt, um die Entscheidung zu treffen, ob das angesprochene Merkmal diagnostisch oder non-diagnostisch ist.

Kernaussage: Als diagnostisch sind solche Merkmale zu bewerten, deren Wahrscheinlichkeiten sich zwischen den Alternativen unterscheiden. Sind die Wahrscheinlichkeiten unter beiden Hypothesen gleich, so ist das Merkmal als nondiagnostisch einzuschätzen.

BEWERTUNG FÜR DEN BRANDSTIFTUNGS-FALL

Im Fall der Brandstiftung möchten wir Euch nun bitten, für die Evidenzregeln zu überprüfen, inwieweit die in der Dann-Passage angesprochenen Merkmale geeignet sind, zwischen der Schuld und Unschuld zu unterscheiden. Das bedeutet, ihr sollt einschätzen, ob sich die Wahrscheinlichkeiten, dass das angesprochene Argument bei Schuld oder Unschuld vorliegt, unterscheiden. Bitte lest auch hierfür die Hintergrundinfos aufmerksam durch und berücksichtigt sie bei euren Urteilen.

Kernaussage: Als diagnostisch sind solche Merkmale zu bewerten, deren Wahrscheinlichkeiten sich zwischen den Alternativen unterscheiden. Sind die Wahrscheinlichkeiten unter beiden Hypothesen gleich, so ist das Merkmal als nondiagnostisch einzuschätzen.

BERÜCKSICHTIGUNG VON ALTERNATIVEN HYPOTHESEN

Im folgenden Schritt soll analysiert werden, ob die Probanden sich entweder auf die Bulimie-Diagnose oder die Schuldvermutung fokussiert haben und Alternativen außer Acht gelassen wurden. Im Bulimie-Fall meint dies konkret, dass die Probanden nicht ausschließlich Merkmale bzgl. eines Störungsbildes ansprechen, sondern mehrere Störungsbilder in Betracht ziehen. Im Brandstiftungs-Fall meint dies konkret, dass nicht nur die Schuld sondern auch die Unschuld thematisiert wird.

BEWERTUNG DER BERÜCKSICHTIGUNG VON ALTERNATIVEN IM BULIMIE-FALL

Die Feststellung erfolgt grundsätzlich durch das **Auszählen** von unterschiedlichen Störungsbildern. Mögliche Werte sind **0** (es wird keine einzige Störung angesprochen) bis **6** (jede Evidenzregel thematisiert eine andere Störung)

- wenn die Regel instruktionskonform ist, dann wird die Anzahl der Störungsbilder in der **Wenn-Passage** ermittelt
- wenn Wenn- und Dann-Passage vertauscht sind, wird die Anzahl der Störungen in der Dann-Passage gezählt
- wenn eine Regel eine *andere Störung* (Kodierung = 5 bei Instruktionskonformität) beinhaltet, wird dies an dieser Stelle berücksichtigt.

Hinweis: berücksichtigt werden hier NUR einzeln genannte Störungen. Störungsgruppen werden nicht gewertet, auch nicht wenn die Störungsgruppe noch durch eine spezifische Störung konkretisiert wird (Bsp.: „*Wenn affektive Störung / Depression, dann ...*“), d.h. solche Regel, die als *ohne explizite Störung* oder *nicht konform* gewertet wurden, bleiben unberücksichtigt.

BEWERTUNG DER BERÜCKSICHTIGUNG VON ALTERNATIVEN IM BRANDSTIFTUNGS-FALL

Es wird ausgezählt, ob null, eine oder zwei Alternativen berücksichtigt werden

- 0= in keiner einzigen Evidenzregel überhaupt die Schuld ODER Unschuld thematisiert wird
- 1=nur eine Alternative wird thematisiert
- 2=sowohl Schuld als auch Unschuld thematisiert werden.

Analog zum Bulimie-Fall ist je nach Instruktionskonformität die Wenn- oder Dann-Passage zu betrachten.

GEWICHTUNG DER ALTERNATIVEN HYPOTHESEN

Nachdem festgestellt wurde, ob mehrere Störungsbilder bzw. sowohl Schuld als auch Unschuld berücksichtigt wurden, soll noch die Gewichtung betrachtet werden. Dies ist natürlich auch wieder nur bei den Evidenzregeln möglich, die instruktionskonform oder „vertauscht“ sind.

BEWERTUNG IM BULIMIE-FALL

Bei instruktionskonformen Evidenzregeln wird dafür einfach ausgezählt, wie viele Dann-Passagen es für eine Wenn-Passage gibt. Als Beispiel liege die Tabelle 1 vor.

Tabelle 1

<u>Wenn-Passage</u>	<u>Dann-Passage</u>
Wenn Bulimie,	dann erbricht die Patientin.
Wenn Bulimie,	dann liegt Normalgewicht vor.
Wenn Bulimie,	dann hat sie Fressanfälle.
Wenn die Mutter pathologisch ist,	dann erhöht dass Marisas Anfälligkeit.
Wenn Anorexie,	dann Untergewicht.
Wenn Anorexie,	dann exzessiver Sport.

Nachdem im vorangegangenen Schritt festgestellt wurde, dass wir 2 unterschiedliche Störungsbilder haben (Bulimie und Anorexie), wird nun ausgezählt, dass für Bulimie 3 und für Anorexie 2 Dann-Passagen vorliegen.

Bei „vertauschten Regeln“ würde dies entsprechend gewertet. Ebenfalls bei Evidenzregeln, die andere Störungen thematisieren. Diese dann in der Spalte „Weitere“ vermerken.

Besonderheit: Unterscheidet ein Proband bei seinen Angaben zwischen der Global-Klasse „Bulimia nervosa“ und „Bulimia nervosa, Nicht-Purging-Typ“, werden diese als unterschiedliche Wenn-Passagen bewertet und der entsprechend genannte Subtyp eingetragen. Verwendet der Probanden AUSSCHLIEßLICH Bezeichnungen der Subtypen, dann trägt bei dem bereits eingetragenen Bulimie/Anorexie entsprechend „0“ ein und zählt die spezifischen wie gewohnt aus.

BEWERTUNG FÜR DEN BRANDSTIFTUNGS-FALL

Bei instruktionskonformen Evidenzregeln wird dafür einfach ausgezählt, wie viele Dann-Passagen es für SCHULD von Herrn Wagner und wie viele für UNSCHULD von Herrn Wagner gibt. Bei „vertauschten Regeln“ würde dies entsprechend gewertet. Die Schuld/Unschuld ist also auf Herrn Wagner begrenzt.

Besonderheit: einige Probanden sprechen von „die Person“, statt eine konkrete aus dem Fall zu benennen. Dies ist so zu werten, als wäre Herr Wagner/ der Angeklagte angesprochen, d.h. zählt zur Kategorie dazu.

Spezielle Hinweise zum Kodierschema:

- vorgegeben sind Spalten für „Anzahl Schuld Wagner“ und „Anzahl Unschuld Wagner“. Hier wird die ausgezählte Anzahl eingetragen
- die letzte Spalte enthält „weitere Alternativen“, die gilt als Sammelbecken. Es werden alle übrigen Alternativen hier kurz genannt und die Anzahl dahinter geschrieben. Z.B. würde dort vermerkt, wenn 3 mal die Schuld von Herrn Lange thematisiert wird (Kodierung: Schuld Lange:3) und in derselben Spalte würde auch „Unschuld Katharina:1“ stehen.

Hier ein ausgefülltes Kodierbeispiel für den Bulimie-Fall:

VP-Code	Evidenzregeln		Instruktionskonformität	Positive Teststrategie	Diagnostizität	Berücksichtigung von Alternativen	Gewichtung von Alternativen		
XYZ	Wenn	Dann	1=konform 2=vertauscht 3= ohne Störung 4=nicht konform	1= positiv 2=negativ 9=Rest 999= nicht auswertbar	0= non-diagnostisch 1=diagnostisch 9=Rest 999= nicht auswertbar	Bulimie: Anzahl Störungen (0-6) Brand: 0=weder noch 1=Schuld 2=Unschuld	Bulimie: Anzahl Bulimie Brand: Anzahl Schuld Wagner	Anorexie: Anzahl Anorexie Brand: Anzahl Unschuld Wagner	Weitere Bitte für alle eintragen Alternative: Anzahl
1	... Anorexie vorliegt...	... übergibt M. sich	1	1	0				Depression:1
2	... die Mutter essgestört ist erhöht das die Wahrscheinlichkeit für eine Essstörung	3	999	999				
3	... M. stundenlang joggt...	... spricht das für Anorexie	2	1	0				
4	Bei Bulimie...	... liegt Krankheitseinsicht vor	1	1	1	3	2	2	
5	... Bulimie...	... hängt Selbstwert von Gewicht ab	1	1	0				
6	... Depression vorliegt...	... sollten keine längere Phasen mit hohem Antrieb vorkommen	5	999	999				

Hier ein ausgefülltes Kodierbeispiel für den Brandstiftungs-Fall:

VP-Code	Evidenzregeln		Instruktionskonformität	Positive Teststrategie	Diagnostizität	Berücksichtigung von Alternativen	Gewichtung von Alternativen		
PQR	Wenn	Dann	1=konform 2=vertauscht 3= ohne Störung 4=nicht konform	1= positiv 2=negativ 9=Rest 999= nicht auswertbar	0= non-diagnostisch 1=diagnostisch 9=Rest 999= nicht auswertbar	Bulimie: Anzahl Störungen Brand: 0=weder noch 1= nur Schuld 2=sowohl als auch	Bulimie: Anzahl Bulimie	Anorexie: Anzahl Anorexie	Weitere Bitte eintragen Alternative: Anzahl
1	... die Person schuldig ist...	... hat sie kein Alibi	1	1	0				Lange Schuld:1
2	... Herr Wagner sich kooperativ bei den Ermittlungen verhält...	... spricht das für seine Unschuld	2	1	0				
3	... Katharina älter wäre...	...spräche das gegen ein kindliches Zündeln	3	999	999	2	1	3	
4	... Herr Wagner unschuldig ist...	... sollte er nichts über die Benzinkanister wissen	1	2	1				
5	... die Person unschuldig ist...	... sollte sie geschockt über die Anklage sein	1	1	0				
6	... Herr Lange schuldig ist...	... müsste man sein Motiv genauer untersuchen	4	999	999				

HINWEISE AUF MÖGLICHE UNGEREIMTHEITEN

Immer wenn die Probanden sich nicht strikt an die Instruktion gehalten haben, kann es zu Bewertungsproblemen kommen. Hier sind einige aufgeführt, die bei der Durchsicht der Antworten auf den ersten Blick auffielen.

MEHRERE INFOS IN DER WENN-/DANN-PASSAGE

Es kann sein, dass die Probanden Folgendes schreiben: *Wenn Bulimie vorliegt, dann sollte M. erbrechen, Fressanfälle haben und exzessiv Sport betreiben*. Hier sind in der Dann-Passage 3 Informationen enthalten.

In diesem Fall bitte wie folgt vorgehen: bewertet werden soll ausschließlich die ERSTE Information. Streicht bitte die anderen durch und ignoriert sie im Urteil. Für das Beispiel würde es also bedeuten, dass die Regel als *Wenn Bulimie vorliegt, dann sollte M. erbrechen*, ~~*Fressanfälle haben und exzessiv Sport betreiben*~~ dargestellt und bewertet wird.

Dies kann genauso gut im Brandstiftungsfall passieren, das Vorgehen ist identisch, ein Beispiel überflüssig.

VERHALTENSKONSEQUENZEN (AV 6)

Neben den Wenn-Dann-Regeln haben die Probanden auch an anderer Stelle Freitextantworten gegeben, die im Folgenden zu beurteilen sind.

NUR BULIMIE: THEMEN IN DER THERAPIESITZUNG/ HAUSAUFGABEN

In einem nächsten Schritt sollten die Probanden, die den klinischen Fall betrachtet haben, angeben, welche Themen sie in den ersten Sitzungen mit der Patientin behandeln würden und welche Hausaufgaben sie ihr geben würden.

Die beiden Typen an Freitextantworten sollen nun in eine von 7 Kategorien zugeordnet werden:

- 1=unspezifisch
- 2=eher konsistent mit Bulimie
- 3=eher konsistent mit Anorexie
- 4= mit beiden gleichermaßen konsistent
- 5=untypisch für Essstörungen
- 6=keine/nichts
- 7=nicht bewertbar

Eine **unspezifische Angabe** meint, dass das vorgeschlagene Thema oder die vorgeschlagene Hausaufgabe nicht speziell auf eine Essstörung zugeschnitten sind. Die angesprochenen Dinge würde man auch bei Patienten anderer Störungsbilder vornehmen. Sie sind sinnvoll für das Gesamtkonstrukt der Therapie, aber eben noch keine spezielle, störungsindizierte Maßnahme.

Eher konsistent mit Bulimie sollen solche Themen bzw. Hausaufgaben gewertet werden, die man mit einer größeren Wahrscheinlichkeit bei Bulimie-Patienten einsetzt als bei Anorexie-Patienten. Die Wahrscheinlichkeit muss nicht bei 1 resp. 0 liegen, es soll nur ein deutliches Gefälle in der Anwendungshäufigkeit vorhanden sein. D.h. bspw., dass man sich eigentlich wundern würde, wenn dieses Thema bei der anderen Störung angesprochen bzw. diese Hausaufgabe vergeben würde.

Eher konsistent mit Anorexie ist analog zu bewerten.

Wenn die Wahrscheinlichkeiten unter beiden Störungsbildern gleich bzw. sehr ähnlich sind (z.B. bei einer Relation von 60:40), dann soll mit **beiden gleichermaßen konsistent** kodiert werden.

Untypisch für Essstörungen (also sowohl bei Bulimie als auch Anorexie) sind Angaben, die sich theoretisch nicht aus den (durchaus verschiedenen) Therapieansätzen ableiten lassen und daher eine Überraschung in der Verordnung darstellen. Man würde also nicht erwarten, dass die angesprochenen Merkmale tatsächlich bei Essstörungenbehandlungen eingesetzt werden. Nicht eingeordnet werden sollen hier solche Merkmale, die man aufgrund seiner eigenen theoretischen

Orientierung (VT, Analytisch, Systemisch etc.) für falsch hält, die aber in die jeweiligen Konzepte passen.

Keine/nichts ist zu kodieren, wenn ausdrücklich angegeben wird, dass seitens des Therapeuten keine Hausaufgaben aufgegeben werden bzw. keine Themen vorzuschlagen sind.

Schließlich gibt es eine Restkategorie (**nicht bewertbar**), in die alle Aussagen kommen, die nicht den bisherigen 5 Kategorien zugeordnet werden konnten. Z.B. weil die Probanden eher die Aufgabe kommentiert haben statt eine Hausaufgabe zu nennen o.Ä.

Weitere Hinweise

- Jede Angabe des Probanden wird nur einer einzigen Kategorie zugeordnet.
- Die freitextantworten enthalten ein paar Besonderheiten, weil sie im PC eingegeben wurden
 - Es werden keine Umlaute angezeigt, sondern immer durch ae, oe etc. umschrieben, das erschwert das Lesen leider ein wenig
 - Es findet sich desöfteren folgende Kombination im Text
.

Sie kann einfach ignoriert werden, da sie technisch bedingt ist.

- Manche Probanden schreiben kurze Antworten, was die Kodierung vereinfacht. Andere Probanden schreiben sehr viele Aspekte in eine Antwort, was die Einordnung sehr problematisch macht. Sollten die Teilantworten von unterschiedlicher Qualität sein, d.h. würden sie verschiedenen Kategorien zugeordnet, wenn ihr sie einzeln einschätzen solltet, dann gebt bitte an, dass die Aussage **nicht bewertbar** ist. Wenn alle genannten Punkte in dieselbe Kategorie fallen würden, dann wählt ihr diese bitte.

ZIEL DER BEFRAGUNG IN EIGENEN WORTEN | BEIDE FÄLLE | SUSPICION CHECK

Ziel der vorliegenden Befragung war es, zu zeigen, dass es auch im Bereich der klinischen Diagnostik zu konfirmatorischem Hypothesentesten kommt. Damit ist gemeint, dass der Ausgang der Hypothesenprüfung durch die Auswahl und/oder Verarbeitung der Informationen zur Prüfung einer Hypothese so beeinflusst wird, dass die Wahrscheinlichkeit ansteigt, die Hypothese zu bestätigen (Hager & Weißmann, 1991, S. 20).

Da Probanden, die sich darüber im Klaren sind, dass wir diese Prozesse untersuchen wollen, eventuell versuchen könnten, diesen entgegen zu wirken, war es nötig, die Probanden über das eigentliche Ziel der Studie im Unklaren zu lassen. Ob uns dieses gelungen ist, möchten wir nun mit Eurer Hilfe herausfinden.

Euch möchten wir nun bitten, zu bewerten, ob die Angaben, die zum Ziel der Studie von den Probanden gemacht wurden, erkennen lassen, dass diese den wahren Grund der Untersuchung durchschaut haben. **Wenn die Probanden äußern, dass wir sie durch die Vorinformationen gezielt beeinflusst/manipuliert haben, die [Bulimie]/[Schuld] Annahme zu übernehmen, um zu überprüfen, ob sie die Voreinstellung „blind“ bestätigen**, dann gilt das Studienziel als durchschaut.

Formuliert der Proband hingegen eine Zusammenfassung wie „*Testen der Fähigkeiten zur integrativen Verarbeitung von Informationen*“, so geht die Vermutung in eine andere Richtung als in die des konfirmatorischen Hypothesentestens. Auch unbedenklich ist es, wenn der Proband den Titel wiederholt (z.B. „*Übung zur diagnostischen Urteilsbildung*“/ Optimierung psychologischer Urteilsfindungs-Strategie etc.) oder von der Erfassung diagnostischer Fähigkeiten bzw. des Wissenstandes von Studierenden/Praktikern bezüglich der Diagnostik spricht. Auch Vermutungen, die nicht mit dem von uns genannten Ziel zusammenhängen, aber nichts mit dem konfirmatorischen Hypothesentesten zu tun haben, sind als „Vermutung in eine andere Richtung“ einzuordnen. So z.B. die Vermutung, wir untersuchten, ob die Reihenfolge, in der Informationen vorgegeben werden, einen Einfluss hat. Oder der Verdacht, in der Studie ginge es um Probleme des DSM.

Bitte vermerkt für jeden Probanden, ob er „**0**“ **unkritisch**, d.h. tatsächlich unkritisch oder aber in eine andere Richtung vermutet oder „**1**“ **kritisch** im Sinne des konfirmatorischen Hypothesentestens ist. Falls ein Proband keine Angaben zum Ziel der Studie gemacht hat, bzw. nur eine sinnlose Buchstabenfolge eingegeben hat, so gebt bitte eine „**9**“=**sinnlos** ein.

FALLMATERIALIEN

Anmerkung: Die bedingungsspezifischen Fallbeschreibungen und Hintergrundinformationen sind mit denjenigen im Anhang A und B identisch, so dass auf einen doppelten Abdruck verzichtet wurde.

Lebenslauf

Persönliche Daten

Name	Britta Dreger, geb. Schäfer
Geburtsdaten	23.02.1982 in Preetz, Schleswig-Holstein
Nationalität	deutsch
Familienstand	Verheiratet, Mutter eines Sohnes, geb. 2010

Schulischer und beruflicher Werdegang

11.07.2012	Promotion im Fach Psychologie
Seit 2008	Wissenschaftliche Mitarbeiterin am Lehrstuhl für Rechtspsychologie, Psychologische Diagnostik und Persönlichkeitspsychologie, Prof. Dr. Günter Köhnken, Christian-Albrechts-Universität zu Kiel
2007-2009	Wissenschaftliche Mitarbeiterin im Rahmen eines DFG Projekts am Lehrstuhl für Rechtspsychologie, Psychologische Diagnostik und Persönlichkeitspsychologie, Prof. Dr. Günter Köhnken, Christian-Albrechts-Universität zu Kiel
2007	Beginn des Promotionsvorhabens
Seit 2006	Freie Mitarbeiterin in der Beratungsstelle MIND, Beratungsstelle für Hochbegabungsdiagnostik der Arbeitsgruppe „Psychologie für Pädagogen“, Prof. Dr. Jens Möller, Christian-Albrechts-Universität zu Kiel
2001-2006	Studium der Psychologie an der Christian-Albrechts-Universität zu Kiel, Abschluss als <i>Diplom-Psychologin</i>
1992-2001	Friedrich-Schiller-Gymnasium in Preetz, Schleswig-Holstein <i>Hochschulreife</i>
1988-1992	Grundschule Schellhorn

