

What works best?

Vergleich dreier objektiver Verfahren zur Messung der pädophilen Orientierung bei Männern

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

vorgelegt von
Dipl.-Psych. Alexander Pohl

Kiel
(2015)

Erstgutachter: Prof. Dr. Günther Köhnken

Zweitgutachter: Priv.-Doz. Dr. Jorge Ponseti

Tag der mündlichen Prüfung: 29.01.2016

Durch den zweiten Prodekan Prof. Dr. John Peterson zum Druck genehmigt am: 01.02.2016

Danksagung

Zuallererst bedanke ich mich auf diesem Weg ganz herzlich bei Herrn PD Dr. phil. Jorge Ponseti – du hast mir durch das Vertrauen, das du in meine Arbeitsweise gesetzt hast, stets das Gefühl vermittelt, den Aufgaben gewachsen zu sein und hast Hannah und mir viel Spielraum in der Ausgestaltung unserer Ideen gelassen. Auch Hannah Gerwinn gilt mein Dank für die großartige und entspannte Zusammenarbeit. Die Arbeit im Team war stets konstruktiv und von gegenseitigem Vertrauen und Respekt geprägt. Besonderer Dank gilt auch Oliver Granert, der die Auswertung z.B. durch die Bereitstellung seiner Toolboxen und die gemeinsamen Besprechungen erheblich erleichterte. Besonderer Dank gilt auch Herrn Prof. Dr. Günther Köhnken, der – trotz Ruhestand – immer mit Rat und Tat zur Seite stand, und allen Versuchspersonen, die so bereitwillig die langen Testungen über sich ergehen lassen haben.

Nicht zuletzt bedanke ich mich bei meinen Eltern für die moralische und finanzielle Unterstützung während meiner langen Ausbildung, ohne welche die Aufnahme und Anfertigung dieser Arbeit kaum möglich gewesen wäre. Auch bei meinem Mann, Florian Pohl, möchte ich mich für seine stets lieben und aufmunternden Worte bedanken. Bei der Korrektur und Fertigstellung habe mich zudem tatkräftig gute Freunde und Weggefährten begleitet: Ina Tikalsky, Sara Jahnke, Philip Brysch und Sophus Damm.

Vorbemerkungen

Alle Teilnehmer der Studie waren Männer. Somit beziehen sich sämtliche Aussagen, Ergebnisse und Interpretationen dieser Arbeit ausschließlich auf Männer.

Die vorliegende Dissertationsschrift basiert auf den im Rahmen des vom Bundesministerium für Bildung und Forschung (BMBF) geförderten Forschungsverbundes NeMUP (Neural mechanisms underlying pedophilia and sexual offending against children: origins, assessment, and therapies, Förderkennzeichen: 01KR1205D) erhobenen Daten. Im Folgenden zeige ich an, welche Forschungsaufgaben ich alleine und welche in Kooperation mit anderen Mitarbeitern des Projekts durchgeführt habe. Im Rahmen des NeMUP-Projektes wurden pädophile und nicht-pädophile Männer untersucht. Die verwendeten Methoden unterschieden sich dabei zwischen den Standorten (Berlin, Duisburg-Essen, Hannover, Kiel und Magdeburg), da neben den allgemeinen Instrumenten an jedem Standort noch standortspezifische Projekte umgesetzt worden sind. Förderbeginn des Projektes war der 01.05.2012. Ich war vom 01.07.2012 bis 30.06.2015 in diesem Projekt angestellt. Somit war ich an der Beantragung des Projektes nicht beteiligt, habe aber u.a. die penisplethysmographische Messung (PPG) für alle Probanden entworfen und eingeführt sowie an der Konzeptualisierung des verwendeten Eingangsinterviews mitgewirkt, um die vorliegende Arbeit zu ermöglichen. Neben dem erwähnten Eingangsinterview fand in Kiel die Erhebung mehrerer Fragebögen, eine neuropsychologische Testung, eine PPG, zwei Messtermine im Magnetresonanztomographen (MRT; u.a. eine strukturelle und drei funktionelle Messungen zur Erfassung der sexuellen Orientierung), eine Blutentnahme, ein Reaktionszeitexperiment (Viewing Time Verfahren), ein computerbasiertes Verfahren zur Erfassung von Empathie und eine Intelligenzmessung statt (für eine genauere Übersicht siehe 2.2.1). Die Datenerhebung wurde in Kiel von mir und meiner Kollegin Frau Gerwin durchgeföhrt. In den meisten Fällen hat ein Versuchsleiter die gesamte Testung eines Probanden durchgeföhrt, sodass Frau Gerwin und ich die gesamte Testbatterie mehrfach selbstständig durchgeföhrt haben. Die vorliegende Arbeit nutzt allein die in Kiel erhobenen Daten des Interviews (zur Gruppeneinteilung), des Viewing-Time-Verfahrens (VT), der PPG und Teile der im MRT erhobenen Daten (das strukturelle Bild und die erste Messung zur Erfassung der sexuellen Orientierung). Die ersten Analysen zur Bestimmung der Klassifikationsgenauigkeit basierend auf den fMRT-Daten (der ersten Messung zur Erfassung der sexuellen Orientierung im MRT) habe ich gemeinsam mit Frau Gerwin durchgeföhrt. Bei technischen Fragen stand uns Oliver Granert zur Seite. Wir haben untersucht, welche Stimuli am besten zur Klassifikation der pädo- und hebephilen Probanden geeignet sind. Anhand dieser

Voranalysen habe ich die Auswahl der in meiner Arbeit verwendeten Stimuli, statistischen Kontraste sowie die Zusammensetzung der Probandengruppen (gemeinsame Betrachtung der pädo- und hebephilen Probanden als eine Gruppe) bestimmt. Die Unterschiede der einzelnen Klassifikationsmodelle sollen in dieser Arbeit nicht weiter vorgestellt werden. Die Auswahl der Probanden für meine Analysen habe ich genauso selbstständig vorgenommen wie die endgültige Auswertung und Aufbereitung der hier dargestellten Ergebnisse. Die Auswertung der PPG- und VT-Daten habe ich eigenständig durchgeführt. Darauf aufbauend habe ich die Klassifikationsgenauigkeiten der drei Verfahren (VT, PPG, fMRT) verglichen und die Ergebnisse diskutiert. Da somit z.B. die Datenerhebung nicht von mir alleine durchgeführt worden ist, werde ich für Teile, in welchen ich durch unser Team unterstützt wurde, das Personalpronomen „Wir“ verwenden. Wann immer ich das Personalpronomen „Ich“ verwende, zeigt dies an, dass ich die entsprechenden Teile selbstständig und alleine geplant, durchgeführt und interpretiert habe.

1	THEORETISCHER HINTERGRUND.....	12
1.1	Begriffsdefinitionen	12
1.2	Nutzen objektiver Verfahren.....	14
1.2.1	Zu den wichtigsten psychometrischen Kennwerten	18
1.2.2	Überblick über das Viewing Time Verfahren	22
1.2.2.1	Zur Entstehung und Methodik.....	22
1.2.2.2	Theoretische Grundlage	23
1.2.2.3	Psychometrische Kennwerte	25
1.2.2.4	Das Abel Assessment for Sexual Interest (AASI).....	25
1.2.3	Überblick über die Penisplethysmographie	29
1.2.3.1	Zur Entstehung und Methodik.....	29
1.2.3.2	Theoretische Grundlage	34
1.2.3.3	Psychometrische Kennwerte	34
1.2.4	Überblick über die funktionelle Magnetresonanztomographie (fMRT)	37
1.2.5	Vergleichende Arbeiten	40
1.3	Ableitung der Fragestellung	47
2	METHODEN.....	48
2.1	Probanden	48
2.2	Untersuchungsablauf und Methodik.....	51
2.2.1	Untersuchungsablauf im Projekt NeMUP	51
2.2.2	Untersuchungsablauf und Methodik VT	53
2.2.2.1	Stimuli und Methodik beim VT	53
2.2.2.2	Untersuchungsablauf VT	53
2.2.2.3	Auswertung VT	54
2.2.3	Untersuchungsablauf und Methodik PPG.....	54
2.2.3.1	Stimuli und Methodik der PPG	54
2.2.3.2	Untersuchungsablauf PPG	56
2.2.3.3	Auswertung PPG	57
2.2.4	Untersuchungsablauf und Methodik fMRT.....	58
2.2.4.1	Stimuli und Methodik beim fMRT	58
2.2.4.2	Untersuchungsablauf fMRT	60

2.2.4.3	Auswertung fMRT.....	61
2.2.5	Berechnung der internen Konsistenz und Vergleich der verschiedenen Verfahren.....	65
3	ERGEBNISSE.....	66
3.1	Vorbemerkungen und Reihenfolge der Testverfahren.....	66
3.2	Reliabilität und Validität des VT.....	66
3.3	Reliabilität und Validität der PPG.....	68
3.4	Validität der fMRT.....	72
3.5	Vergleich der Validität der einzelnen Verfahren.....	74
3.6	Konvergente und inkrementelle Validität.....	75
4	DISKUSSION.....	77
4.1	Vorbemerkungen.....	77
4.2	Diskussion VT.....	78
4.3	Diskussion PPG.....	80
4.4	Diskussion fMRT.....	82
4.5	Vergleich der Verfahren.....	84
4.6	Konvergente und inkrementelle Validität.....	86
4.7	Problematische Fälle.....	88
4.8	Verwendung objektiver Verfahren in der Therapie und Prognostik.....	89
4.9	Verwendung objektiver Verfahren als Screening-Instrumente.....	89
4.10	Stigmatisierung.....	92
5	ZUSAMMENFASSUNG.....	93

6	LITERATURVERZEICHNIS.....	96
7	ANHANG.....	108
7.1	Anhang 1: Probandeninformation für NeMUP	108
7.2	Anhang 2: Einverständniserklärung NeMUP.....	111
7.3	Anhang 3: Interviewleitfaden NeMUP.....	113
7.4	Anhang 4: Instruktion VT.....	129
7.5	Anhang 5: Instruktionen PPG – Probedurchgang	130
7.6	Anhang 6: Instruktionen PPG – Hauptdurchgang	131
7.7	Anhang 7: Verteilung der Probanden in Relation der Altersindizes bei VT, PPG und fMRT	132
7.8	Anhang 8: Interkorrelationen der Stimuluskategorien zwischen VT, PPG und fMRT	134
7.9	Anhang 9: Karikatur zum Thema Pädophilie.....	135
8	LEBENS LAUF.....	136

Abkürzungsverzeichnis

APA	American Psychiatric Association
BMBF	Bundesministerium für Bildung und Forschung
DSM	Diagnostisches und statistisches Manual psychischer Störungen (herausgegeben von der APA)
fMRT	funktionelle Magnetresonanztomographie
ICD	Internationale Klassifikation der Krankheiten (von der WHO herausgegeben)
M	Mittelwert
(s)MRT	(strukturelle) Magnetresonanztomographie
NeMUP	Forschungsverbund NeMUP (Neural mechanisms underlying pedophilia and sexual offending against children: origins, assessment, and therapies)
PPG	Penisplethysmographie
SD	Standardabweichung
SSPI	Screening Scale for Pedophilic Interest
STT	Semantic tracking task
VT	Viewing-Time-Verfahren
WHO	Weltgesundheitsorganisation

Abbildungsverzeichnis

Abbildung 1: Beispiele für eine ROC Analyse.	21
Abbildung 2: Schematische Darstellung einer BOLD Antwort.	37
Abbildung 3: Schematische Darstellung der fMRT-Auswertung.	64
Abbildung 4: Mittlere Betrachtungszeit der zehn Stimulusgruppen beim VT.	67
Abbildung 5: μ -Volt-Schwankungen in Reaktion auf die Stimulus-Gruppen in der PPG... ..	69
Abbildung 6: <i>T</i> -Karten bei den Kontrasten Männer > Jungen bzw. Frauen > Mädchen	72
Abbildung 7: Altersbedingte Atrophien zweier älterer Probanden.	74
Abbildung 8: Verteilung der Probanden entlang der zwei Altersindizes des VT	132
Abbildung 9: Verteilung der Probanden entlang der zwei Altersindizes der PPG (vor Ausschluss vor Low-Respondern)	132
Abbildung 10: Verteilung der Probanden entlang der zwei Altersindizes der PPG (nach dem Ausschluss von Low-Respondern).....	133
Abbildung 11: Verteilung der Probanden entlang der zwei Altersindizes der fMRT.....	133

Tabellenverzeichnis

Tabelle 1: Zuordnung der sexuellen Alters- und Geschlechtsorientierungen zu den jeweils präferierten Reizen.	34
Tabelle 2: Übersicht über bisher veröffentlichte Studien, die PPG und VT miteinander verglichen haben.	46
Tabelle 3: Stichprobenbeschreibung	49
Tabelle 4: Werte für Cronbachs Alpha für die einzelnen Stimuluskategorien des VT	67
Tabelle 5: Klassifikationsleistung des VT.	68
Tabelle 6: Split-Half-Reliabilitätskoeffizienten der PPG.....	70
Tabelle 7: Klassifikationsleistung der PPG.....	71
Tabelle 8: Korrekt klassifizierten Probanden bei der fMRT.....	73
Tabelle 9: Klassifikationsleistungen der drei Verfahren in der Übersicht	75
Tabelle 10: Korrelationen zwischen den Alters-Indizes der drei Verfahren.....	75
Tabelle 11: Korrelationen zwischen den Alters-Indizes der drei Verfahren (aufgeteilt nach Geschlecht).....	76
Tabelle 12: Partialkorrelation (bereinigt um den Einfluss der sexuellen Geschlechtsorientierung) zwischen den Alters-Indizes der drei Verfahren.	76
Tabelle 13: Spezifität, Sensitivität sowie mittlere Klassifikationsgenauigkeit bei der Kombination von mehreren Prädiktoren.	77
Tabelle 14: Interkorrelationen der Indizes der einzelnen Verfahren.....	134

1 Theoretischer Hintergrund

1.1 Begriffsdefinitionen

Nach LeVay und Baldwin (2009) bezeichnet die sexuelle Orientierung die Ausrichtung der sexuellen Anziehung auf das weibliche oder das männliche (oder auf beide) Geschlecht(er). Bailey (2009) grenzte das Konzept der sexuellen Orientierung von dem privaten und öffentlichen sexuellen Selbstkonzept, dem sexuellen Verhalten und der sexuellen Präferenz ab. Mit sexueller Präferenz meint er dabei, ob eine Person es bevorzugt, sexuelle Kontakte mit Männern und/oder Frauen zu unterhalten. Diese verschiedenen Ebenen können voneinander abweichen, sodass z.B. aus dem sexuellen Verhalten oder der Selbstdefinition nicht ohne weiteres auf die sexuelle Orientierung geschlossen werden kann. Nach Bailey (2009) bestimmt die sexuelle Orientierung die Ausrichtung unserer sexuellen Wünsche, sexuellen Erregung, Fantasien und Anziehungen. Die sexuelle Orientierung gilt dabei als stabiles Persönlichkeitsmerkmal (Heitzman, Lew-Starowicz, Pacholski & Lew-Starowicz, 2014) und resistent gegenüber sogenannten Konversionstherapien (American Psychological Association, 2009; Beckstead, 2012; Haldeman, 1994). Sie ist vergleichbar mit einem Kompass, der unsere Ausrichtung auf sexuelle Reize angibt, nicht aber zwangsläufig unser sexuelles Verhalten.

Dem diagnostischen und statistischen Manual psychischer Störungen zufolge (DSM-5; American Psychiatric Association, 2013) ist die Pädophilie durch wiederkehrende, intensive sexuelle Fantasien, Bedürfnisse oder Verhaltensweisen geprägt, die sich auf Kinder vor dem 14. Lebensjahr beziehen. Da die körperliche Reifung einer erheblichen Variabilität unterliegt (zwischen den Geschlechtern, Individuen und zeitlichen Epochen; Beccuti & Ghizzoni, 2000; Kryst, Kowal, Woronkowicz, Sobiecki & Cichocka, 2012), eignet sich die Einteilung in Reifestadien besser zur Charakterisierung sexueller Alterspräferenzen als das chronologische Alter. Tanner (1978) beschreibt die Entwicklung der primären und sekundären Geschlechtsmerkmale und unterteilt die Entwicklung in fünf Stadien. Das Tanner Stadium 1 beschreibt einen präpuberalen Entwicklungsstand (fehlende Schambehaarung, kein Brustwachstum bei Mädchen), die Tanner Stadien 2 und 3 beschreiben einen frühen puberalen Entwicklungsstand (Beginn der Schambehaarung, erste Brustknospung und beginnendes Wachstum der Brust; Peniswachstum und farbliche Veränderung des Skrotums) und Stadium 4

einen spätpuberalen Entwicklungsstand. Das fünfte und letzte Stadium beschreibt eines der vollständigen sexuellen Reife.

Während sich die sexuellen Fantasien bei pädophil veranlagten Personen auf vorpubertäre Kinder des ersten Tannerstadiums beziehen (Freund, 1967; Tanner, 1978), präferieren hebephile Männer jugendliche Sexualpartner, die sich im Beginn der Pubertät befinden (Tannerstadium 2-3; Blanchard, Lykins, et al., 2009; Glueck, 1955). Als grobe Altersorientierung werden meist die Altersbereiche bis zum 10. Lebensjahr (pädophile Präferenz) und zwischen dem 11. und 14. Lebensjahr (hebephile Präferenz) angegeben. Von diesen beiden in der Literatur pathologisierten Begriffen ist die Orientierung auf spätpubertierende und junge Erwachsene in einem Altersbereich von ca. 15 bis 19 Jahren abzugrenzen (Blanchard et al., 2008). Krafft-Ebbing prägte für diese Orientierung die Begriffe der Parthenophilie (sexuelles Interesse an jugendlichen Frauen) und Ephebophilie (sexuelles Interesse an jungen Männern; Krafft-Ebbing, 1924). Blanchard et al. (2008) haben den bis dahin wenig geläufigen Begriff der Ephebophilie wieder in der Literatur eingeführt, ohne ihn auf ein bestimmtes Zielgeschlecht zu begrenzen. Wie Blanchard et al. (2008) plädierten auch Ahlers, Schaefer und Beier (2005) dafür, eine sexuelle Orientierung auf Personen in diesem Altersbereich bzw. Entwicklungsstadium nicht zu pathologisieren, sondern als erwartbares Phänomen anzusehen.

Die Wünsche und Vorlieben pädophiler und hebephiler Männer beziehen sich – wie bei Teleiophilen auch – nicht ausschließlich auf sexuelle Vorstellungen, sondern schließen alle Aspekte von Liebe, Zuneigung und Beziehungsgestaltung ein (Vogt, 2006). Zur Diagnostik pädophiler Präferenzen eignet sich allerdings vor allem der sexuelle Aspekt, da sich dieser sowohl im Interview als auch in objektiven Messverfahren am einfachsten operationalisieren lässt. Die weiteren Betrachtungen zum Thema Pädophilie stützen sich daher rein auf die sexuelle Natur dieser Orientierung. Unter anderem aufgrund der zeitlichen Konstanz pädophiler Neigungen plädierte Seto (2012) dafür, die Pädophilie als eine sexuelle Orientierung aufzufassen. Muller et al. (2014) hingegen gingen von einer Änderbarkeit pädophiler Präferenzen aus. Sie konnten zeigen, dass ca. die Hälfte ihrer Probanden nach einer Therapie eine verminderte genitale Reaktion in Bezug auf Kinder und gleichzeitig eine erhöhte Reaktion auf Erwachsene aufwies. Ihre Studie führte allerdings zu einer kontroversen Diskussion in der Literatur und eine Replikation der Befunde steht noch aus (Bailey, 2014; Cantor, 2015; Fedoroff et al., 2015; Lalumiere, 2015). Auch wenn im deutschen Sprachraum oft auf eine andere Terminologie zurückgegriffen wird (siehe z.B. Ahlers & Schaefer, 2010), werde ich daher im Folgenden in Anlehnung an Seto (2012) sowohl in Bezug

auf die Alters- als auch die Geschlechtsorientierung von der sexuellen Orientierung sprechen. Dies erlaubt mir, von der Messung der sexuellen Orientierung zu sprechen, unabhängig davon, ob eine Alters- oder Geschlechtspräferenz bestimmt werden soll. In diesem Sinne werde ich das Wort Pädophilie für die sexuelle Orientierung auf das vorpubertäre Körperschema (Tanner 1), den Begriff Hebephilie für die Orientierung auf ein jugendliches Körperbild (Tanner 2-3) und Teleiophilie (Blanchard & Barbaree, 2005) für die Orientierung auf das erwachsene, postpubertäre Körperschema (Tannerstadium 4-5; Beier et al., 2013) verwenden. Die Geschlechtsorientierung ist von der Altersorientierung unabhängig, sodass es sowohl hetero-, homo- als auch bisexuelle Personen der verschiedenen Altersorientierungen gibt¹. Freund und Watson (1992) wiesen allerdings bereits darauf hin, dass der Anteil homosexueller Probanden (zumindest in Straftäterpopulationen) bei pädo- und hebephilen Männern größer als bei teleiophilen Männern ist.

Der überwiegende Teil der Forschung zu Pädophilie und sexuellen Missbrauchsverhalten an Kindern bezieht sich auf Männer, wenngleich es zumindest Einzelfallstudien zu weiblichen Pädophilen gibt (Chow & Choy, 2002; Fedoroff, Fishell & Fedoroff, 1999). Da Frauen aber nur eine Minderheit der Täter von sexuellen Kindesmissbrauch ausmachen (Bundeskriminalamt, 2012) und die Nutzung indirekter Messverfahren bei Frauen generell erschwert ist (Chivers, Rieger, Latty & Bailey, 2004; Lippa, Patterson & Marelich, 2010), befasst sich die vorliegende Arbeit ausschließlich mit Männern.

1.2 Nutzen objektiver Verfahren

Die korrekte Bestimmung der sexuellen Orientierung ist in mehreren Bereichen von außerordentlicher Bedeutung. Das sexuelle Verhalten ist nicht in allen Fällen deckungsgleich mit der sexuellen Orientierung, was eine Erhebung dieses Konstruktes prinzipiell lohnenswert macht. So sind bei weitem nicht alle Personen, die sexuellen Kindesmissbrauch begehen pädophil (Hall & Hall, 2007; Seto, 2007) und nicht alle pädophilen Männer begehen sexuellen Kindesmissbrauch (Beier et al., 2009). Da sexuelle Handlungen deshalb kein perfekter Indikator für die sexuelle Orientierung ist, haben Psychologen und Mediziner weitere Indikatoren für eine objektive Bestimmung pädo- oder hebephiler Orientierungen gesucht und verschiedene Messverfahren entwickelt (Thornton & Laws, 2009). Die objektive Bestimmung der sexuellen Orientierung ist

¹ Mit einem homopädophilen Mann ist demzufolge z.B. ein Mann gemeint, der sexuell auf Jungen vor der Pubertät orientiert ist.

z.B. bei der Prognosebegutachtung hilfreich. Pädophile Täter werden häufiger rückfällig als sogenannte Ersatz- bzw. Gelegenheitstäter (Beier, 1998; Beier, Bosinski & Loewit, 2005; Malcolm, Andrews & Quinsey, 1993). Darüber hinaus erwies sich ein Index für Pädophilie in einer Metanalyse von Hanson und Morton-Bourgon (2005) als der beste einzelne Prädiktor für erneute Sexualdelikte. Für die Therapieplanung – sowohl im ambulanten Rahmen als auch im Strafvollzug – ist die objektive Messung der sexuellen Orientierung ebenfalls von Bedeutung, da sogenannte Neigungstäter von anderen Therapiekonzepten profitieren als Ersatztäter. Auch in Bezug auf die Frage der verminderten Schuldfähigkeit nach § 20 bzw. 21 StGB ist die sexuelle Orientierung von erheblichem Interesse. Eine ausgeprägte Pädophilie kann im Sinne einer schweren anderen seelischen Abartigkeit zu einer verminderten Schuldfähigkeit und – in Kombination mit einer hohen Rückfallgefahr – zu einer Unterbringung in einem psychiatrischen Krankenhaus nach § 63 StGB führen. Die Einschätzung, ob eine Pädophilie im Sinne einer schweren anderen seelischen Abartigkeit und/oder eine erhebliche Gefahr für erneute Straftaten vorliegt, obliegt der Einschätzung eines Gutachters. Die Auswahl der zur Urteilsfindung genutzten Methoden obliegt dem Gutachter, wobei dieser ausschließlich wissenschaftlich fundierte Methoden verwenden soll (Boetticher, Nedopil, Bosinski & Saß, 2007).

In Deutschland findet zur Feststellung einer pädophilen Orientierung vor allem das klinische Interview – meist in einer nicht standardisierten Form – Anwendung. Geprüft wird dabei das Vorliegen einer Diagnose nach DSM oder ICD. Zudem entwickelten Seto, Harris, Rice und Barbaree (2004) eine Kurzform der Fremdeinschätzung: Die „Screening Scale for Pedophilic Interest“ (SSPI). Diese soll als Indikator für das Vorliegen einer sexuellen Orientierung auf Kinder dienen und basiert auf vier Items, welche sich ausschließlich auf bereits stattgefundenen sexuelle Übergriffe beziehen. Laut der Autoren eignet sich diese Checkliste nicht zur Diagnostik im Einzelfall und kann aufgrund der Item-Struktur auch nur bei bereits übergriffig gewordenen Patienten angewandt werden. Der MSI (Multiphasic Sex Inventory; Nichols & Molinder, 1984) erfasst neben einer Reihe anderer Konstrukte ebenfalls sexuelles Interesse an Kindern. Auch dieses Verfahren eignet sich den Autoren zufolge aber nicht zur Diagnosestellung sondern lediglich zu einer Schweregradeinschätzung (Mackaronis, Strassberg & Marcus, 2011).

Kingston, Firestone, Moulden und Bradford (2007) wiesen in ihrer Arbeit zudem auf eine geringe Übereinstimmung zwischen verschiedenen Diagnoseinstrumenten hin. Sie stellten eine Diagnose anhand von DSM-IV Kriterien, SSPI oder PPG. Die Autoren fanden keinen statistisch signifikanten Zusammenhang zwischen einer Diagnose nach DSM-IV, SSPI bzw. PPG. Die

Autoren führen dies unter anderem auf den geringen Nutzen der diagnostischen Kriterien im DSM-IV zurück. Sie stützen sich dabei auf mehrere Studien, die die Kriterien nach DSM-IV kritisieren. Problematisch ist bei der Diagnosestellung zum Beispiel die geringe Übereinstimmung zwischen mehreren Ratern. So ermittelte Levenson (2004) eine Inter-Rater-Reliabilität bei Pädophilie von 64 %.

Wenn sich der Gutachter bzw. Therapeut ausschließlich auf die Selbstausskünfte der zu untersuchenden Person verlässt, kann der Patient das Stellen einer korrekten Diagnose durch mangelnde Kooperationsbereitschaft erheblich erschweren. Die meisten Sexualstraftäter leugnen eine pädophile Motivation oder sogar die ihnen vorgeworfenen Taten (Blanchard, Klassen, Dickey, Kuban & Blak, 2001). Studien konnten belegen, dass Angeklagte in Strafverfahren Testergebnisse häufig bewusst beeinflussen (Laws & Holmen, 1978; Wilson, 1998). Selbst Patienten, die an einer Sexualstraftätertherapie teilnehmen, bagatellisieren bezüglich ihrer Opferzahlen (Ahlmeyer, Heil, McKee & English, 2000). Bei sexuell unerfahrenen Jugendlichen ist die Erhebung einer Sexualanamnese weiterhin erschwert, da bei Jugendlichen sexuelle Kontakte mit Kindern einerseits normal, andererseits aber auch ein Indikator für eine sich entwickelnde Pädophilie sein können. Zudem stützt sich die Sexualanamnese unter anderem erheblich auf frühere sexuelle Erfahrungen. Wenn diese nicht vorhanden sind, fehlen wichtige Informationen zur Einschätzung der sexuellen Orientierung. Zusätzlich erschwert wird das Problem, wenn der Proband verdächtigt wird, die Erhebung durch eine bewusste Täuschungsabsicht zu behindern. Zudem scheint gerade bei jugendlichen Kindesmissbrauchern die Rate an nicht-pädophilen Tätern (zumindest bei Taten zum Nachteil von Mädchen) relativ hoch zu sein (Seto, Lalumière & Blanchard, 2000). Bei sexualmedizinischen Untersuchungen kann auch soziale Erwünschtheit zum Verschweigen von als unpassend oder angenehm erlebten Fantasien führen (Looman, Abracen, Maillet & DiFazio, 1998), was auch bei ambulanten Therapien eine Rolle spielt. Des Weiteren können Sprachbarrieren oder eine eingeschränkte kognitive Leistungsfähigkeit des Probanden die Exploration erschweren. Den therapeutischen Nutzen von objektiven Verfahren beschreiben z.B. Travin, Cullen und Melella (1988). Sie konnten zeigen, dass nach einer Konfrontation mit den erhobenen Messdaten 55 % der Personen, die vorher jegliche pädophile Motivation der Straftaten abstritten (im Folgenden: Non-Admitter), ihre Taten und eine pädophile Motivation einräumten. Dies stellte einen entscheidenden Schritt dar, da erst anschließend eine therapeutische Bearbeitung möglich war.

In der Hoffnung täuschungsresistente und valide Möglichkeiten zu schaffen, die sexuelle Orientierung von Personen zu messen, wurden sogenannte objektive Verfahren entwickelt. Die

Begriffe indirekte (Banse, Schmidt und Clarbour, 2010) bzw. objektive Messverfahren werden von unterschiedlichen Autorengruppen verwendet, sind aber als Synonyme zu betrachten. Damit sind Verfahren gemeint, die sich nicht auf die Selbstauskünfte der zu untersuchenden Person stützen, wie dies z.B. bei Fragebögen- oder Interviewdaten der Fall ist. Mittlerweile gibt es eine ganze Reihe solcher Methoden, die in einer Vielzahl von Studien untersucht worden sind (siehe z.B. Blanchard et al., 2001; Harris, Rice, Quinsey & Chaplin, 1996; Imhoff, Schmidt, Weiss, Young & Banse, 2012; Kalmus & Beech, 2005; Ponseti et al., 2012).

Eine wichtige Voraussetzung für die Anwendung objektiver Verfahren ist die Freiwilligkeit der Teilnahme seitens des Probanden bzw. Patienten. Aktuell gelten objektive Verfahren nach der Strafprozessordnung in Deutschland als verbotene Beweismittel (siehe Urteil des OLG Köln vom 09.03.2004, 2 WS 32/04; Pohl, 2009). Nach § 136a StPO wäre im Strafverfahren – auch wenn objektive Verfahren als Beweismittel zugelassen wären – die Freiwilligkeit der Teilnahme eine Voraussetzung für die Verwendung von Beweismitteln. Zwang darf nach §136a StPO nur dann angewandt werden, „soweit das Strafverfahrensrecht dies zulässt“. 2004 entschied das OLG Köln z.B., dass ein 81jähriger Angeklagter gegen seinen Willen einer phallometrischen Messung zur Bestimmung seiner Erektionsfähigkeit unterzogen werden konnte (OLG Köln, 09.03.2004, 2 WS 32/04). Der Angeklagte hatte in dem konkreten Fall zu seiner Verteidigung vorgebracht, dass er aufgrund fehlender Erektionsfähigkeit keine Vergewaltigung hätte begehen können. Das Bundesverfassungsgericht widersprach dem Urteil des OLG Köln (21.05.2004, 2 BvR 715/04). Zur Begründung wurde ausgeführt, dass die Anwendung in diesem Fall unverhältnismäßig gewesen sei (nach § 81 StPO).

Unabhängig von ethischen Überlegungen ist auch aus messtechnischen Gründen die Freiwilligkeit der Teilnahme eine zwingende Voraussetzung. Sofern der Proband bzw. Patient nicht an einer Messung teilnehmen will, kann er diese durch entsprechendes Einwirken unbrauchbar machen. Bei der PPG kann die manuelle Manipulation des Messgerätes die Auswertung der Messung unmöglich machen; beim fMRT würden schon starke Kopfbewegungen oder das Schließen der Augen eine Messung unbrauchbar machen. Eine Aufdeckung der inneren Wünsche ohne das Einverständnis des Probanden ist somit streng genommen nicht möglich.

Zudem kann mittels objektiver Verfahren immer nur eine Aussage über die sexuelle Orientierung des Probanden getroffen werden. Da sexuelles Verhalten und sexuelle Orientierung zwei distinkte Konstrukte sind (Ahlers & Schaefer, 2010; Seto, 2007) können die Ergebnisse solcher Messungen niemals genutzt werden, um Schuld oder Unschuld von angeklagten Personen

zu beweisen. Es gibt zwar im Hellfeld einen statistischen Zusammenhang zwischen der sexuellen Orientierungen und Rückfälligkeit in Bezug auf Sexualstraftaten (Hanson & Bussiere, 1998; Hanson & Morton-Bourgon, 2005), wie jedoch im Dunkelfeld die Zusammenhänge zwischen sexuellen Fantasien und Straftaten sind, ist allerdings noch weitestgehend ungeklärt (Dombert et al., 2013; Dombert et al., 2015). Zudem sind dies nur statistische Zusammenhänge, die nicht geeignet sind, um die Schuldfrage im Einzelfall zu klären.

Im Mittelpunkt der vorliegenden Arbeit stehen die psychometrischen Kennwerte der sowie die Beziehungen zwischen den folgenden drei Verfahren: Das Viewing-Time-Verfahren (VT), die Penisplethysmographie (PPG) und die Messung der sexuellen Orientierung mittels funktioneller Magnetresonanztomographie (fMRT). Für einen Überblick zu weiteren Verfahren verweise ich auf Thornton und Laws (2009).

1.2.1 Zu den wichtigsten psychometrischen Kennwerten

Ich werde kurz auf die für die Beurteilung der vorgestellten Verfahren wichtigsten psychometrischen Kennwerte eingehen. Die Güte psychologischer Testverfahren wird anhand der Objektivität, Reliabilität und Validität beurteilt. Die Objektivität (unterteilt in Durchführungs-, Auswertungs- und Interpretationsobjektivität) gibt an, wie unabhängig die Messergebnisse vom Testleiter sind. Sie ist z.B. geringer, wenn es bei der Auswertung der Testergebnisse einen großen Spielraum gibt, wie dies z.B. bei projektiven Tests der Fall ist. Da die Datenerfassung und -auswertung bei allen vorgestellten Verfahren automatisiert erfolgen kann – und in der Literatur bisher keine Versuchsleitereffekte berichtet wurden – soll auf dieses Konstrukt nicht weiter eingegangen werden. Die Objektivität könnte höchstens bedroht sein, wenn der Versuchsleiter sich nicht an die notwendigen Instruktionen hält (z.B. Abel, Blanchard & Barlow, 1981), weswegen die Erstellung und Einhaltung von standardisierten Vorgehensweisen zu empfehlen ist.

Die Reliabilität gibt die Zuverlässigkeit von Tests an. Sie wird über korrelative Analysen bestimmt und rangiert deshalb zwischen -1 und 1. Bei der Paralleltestreliabilität werden die Korrelationen zwischen zwei ähnlichen Testformen berechnet; sie ist für unsere Betrachtung von untergeordneter Bedeutung. Die *interne Konsistenz* gibt an, ob die einzelnen Items eines Tests dasselbe Konstrukt messen. So wird von einem Test mit verschiedenen Items gefordert, dass diese hoch miteinander korrelieren, sofern sie dasselbe Konstrukt erfassen. Üblicherweise erfolgt die Messung über Cronbachs Alpha (Korrelation der einzelnen Items untereinander; Cronbach, 1951)

oder die Testhalbierungsreliabilität (Korrelation von 2 Testhälften). Die Retest-Reliabilität gibt die Korrelation bei wiederholter Messung mit demselben Test an. Werte unter .7 werden als unzureichend bewertet, Werte ab .8 werden als gut bzw. hoch eingeschätzt (Cohen, 1992; Schmitt, 1996).

Bei der Validität (Gültigkeit der Messung) ist bei der Beurteilung der Güte objektiver Verfahren zur Messung der sexuellen Orientierung besonders die *Kriteriumsvalidität* entscheidend. Sie gibt an, wie gut die Messergebnisse mit einem empirischen Maß (anerkannter Test oder Expertenrating) übereinstimmen. Beurteilt wird die Kriteriumsvalidität bei objektiven Verfahren meist, indem Gruppen untersucht werden, bei denen die sexuelle Orientierung a priori bekannt ist. So werden bei den ersten Validierungsbestrebungen neuer Verfahren die Probanden meist bezüglich ihrer sexuellen Orientierung befragt und die Übereinstimmung zwischen den Selbstauskünften und den Messergebnissen ermittelt (Freund, 1957; Lippa, 2013; Ponseti et al., 2009). Um die Gültigkeit des Tests in anderen Populationen zu bestätigen, empfiehlt sich bei objektiven Tests zur Messung der sexuellen Orientierung die Kriteriumsvalidität auch bei Non-Admittern zu prüfen. Um dies zu erreichen, werden Personen untersucht, die mit hoher Wahrscheinlichkeit pädophil sind (als Indikator dienen hier meist mehrmalig erfolgte Missbrauchshandlungen), entsprechende Orientierungen aber abstreiten (Blanchard et al., 2001). Umso fehlerbehafteter das Kriterium ist, an dem der Test validiert wird, umso geringer wird die Schätzung der Validität ausfallen (Blanchard et al., 2001).

Die Validität kann mittels verschiedener Herangehensweisen bestimmt werden. Oft wird ein willkürlicher Grenzwert festgelegt und anschließend bestimmt, wie viele Probanden hierdurch korrekt den ursprünglichen Gruppen zugeordnet werden (siehe z.B. Blanchard et al., 2001). Die Sensitivität gibt dabei an, wie viel Prozent der Merkmalsträger (hier: wie viel Prozent der Pädophilen) korrekt als solche erkannt werden (Prozentsatz richtig positiver Testergebnisse). Die Spezifität bezeichnet den Prozentsatz der nicht-Merkmalsträger (hier: Prozentsatz der Teleiophilen), die korrekt als solche erkannt werden (korrekt negative Testergebnisse). Werden Spezifität und Sensitivität gemittelt ergibt sich ein Wert für die Kriteriumsvalidität (mittlere Klassifikationsgenauigkeit).

Neben der Verwendung von Grenzwerten können auch automatische Klassifikationsalgorithmen genutzt werden, um die Güte der Tests zu bestimmen (siehe z.B. Ponseti et al., 2012). Hierzu eignen sich wiederum mehrere Verfahren. Die Diskriminanzanalyse ist ein multivariates Testverfahren. Mit ihrer Hilfe können Personen anhand von einem oder

mehreren Merkmalen in zwei (oder mehr) Gruppen eingeteilt werden (Steven, 2002). Die Zuordnung geschieht automatisiert und kann mittels Kreuzvalidierung abgesichert werden. Ein ähnliches Verfahren stellt die binäre, logistische Regression dar. Hier ist allerdings nur eine Zuordnung in eine von zwei Gruppen möglich. Teilweise werden keine Werte für Spezifität und Sensitivität genannt, sondern Ergebnisse sogenannter ROC (Receiver Operating Characteristic; Zou, O'Malley & Mauri, 2007) Analysen berichtet (siehe z.B. Abel, Jordan, Hand, Holland & Phipps, 2001). Dabei wird das Verhältnis von Sensitivität und Spezifität-1 in einem Diagramm abgetragen (siehe Abbildung 1). Sofern höhere Testwerte ein Indikator für Pädophilie sind, bedeutet z.B. ein AUC (Area Under the Curve) Wert von .95, dass ein zufällig bestimmter Pädophiler mit einer Wahrscheinlichkeit von 95 % einen höheren Wert aufweist als ein zufällig bestimmter Teleiophiler. Ein Verlauf der Vorhersagegrade auf der Diagonalen entspricht der Ratewahrscheinlichkeit; umso näher die Kurve am äußeren linken bzw. oberen Rand verläuft, umso besser ist der Test.

Neben der oben dargestellten *diagnostischen Validität*, welche die Übereinstimmung des Tests mit einem zur selben Zeit erhobenen Außenkriterium (hier: Selbstauskünfte oder Expertenratings zur sexuellen Orientierung der Probanden) angibt, ist auch die *prognostische Validität* von Bedeutung. Zur Bestimmung der prognostischen Validität werden die Messdaten des zu untersuchenden Testverfahrens vor dem Außenkriterium erhoben. Indirekte Verfahren zur Messung der sexuellen Orientierung werden im Kontext von Pädophilie z.B. zur Vorhersage erneuter Straftaten genutzt (Hanson & Morton-Bourgon, 2005).

Die *Konstruktvalidität* ist ein weiteres entscheidendes Merkmal von Testverfahren. Tests, die dasselbe Konstrukt messen, sollten hoch miteinander korreliert sein (*konvergente Validität*), während Tests, die voneinander unabhängig Konstrukte erfassen, nur gering miteinander korrelieren sollten (*diskriminante Validität*).

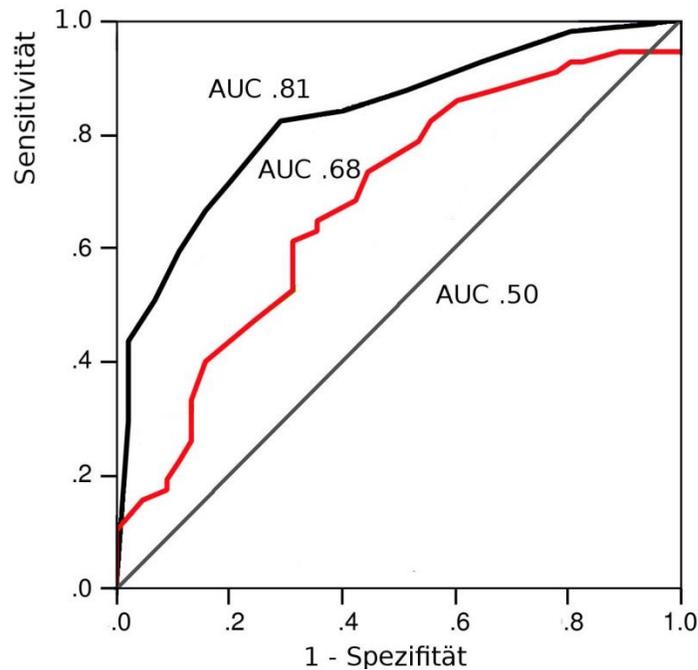


Abbildung 1: Beispiele für eine ROC Analyse mit AUC von .50 (Ratewahrscheinlichkeit), .68 und .81.

Die Kennwerte für die Reliabilität und die Validität können zum einen erhöht werden, indem einem Messinstrument vergleichbare Items hinzugefügt werden. Zur Messung komplexer Konstrukte – wie z.B. bei Persönlichkeitsvariablen – ist es zudem sinnvoll, sowohl prototypische (in Bezug auf die zu messende Eigenschaft) als auch weniger typische Items zu verwenden. Auch die Verwendung von verschiedenen Messungen kann die Vorhersagegenauigkeit erhöhen (Epstein, 1983). Dabei ist es sogar vorteilhaft, wenn die Items nur eine mäßige Korrelation miteinander aufweisen, so lange sichergestellt ist, dass sie dasselbe Konstrukt erfassen. Die Erfassung eines Konstruktes mittels mehrerer Items, Rater, Verfahren usw. nennt man Aggregation (Rushton, Brainerd & Pressley, 1983).

1.2.2 Überblick über das Viewing Time Verfahren

1.2.2.1 Zur Entstehung und Methodik

Das VT geht zurück auf Rosenzweig (1942). Er beschrieb, dass psychiatrische Patienten, die sich exzessiv mit sexuellen Themen beschäftigten und eine Art „hypersexuelles“ Verhalten zeigten, sexuelle Bilder längere Zeit betrachteten, als Patienten, die sich nicht exzessiv mit sexuellen Themen beschäftigten. Zamansky (1956) nutzte diese Beobachtung, um die sexuelle Geschlechtsorientierung seiner Probanden zu messen. Er zeigte den Männern hierzu gleichzeitig zwei Bilder: Eines, das eine Frau, und eines, das einen Mann zeigte. Er erfasste, wie lange die Probanden die Bilder betrachteten und fand, dass homosexuelle Probanden die Männer signifikant länger als die Frauen betrachteten. Aktuell wird das VT – in abgewandelter Form – zur Diagnose pädophiler Orientierungen genutzt. Neben der Messung der Altersorientierung nutzten z.B. Imhoff et al. (2010), Israel und Strassberg (2009) sowie Lippa (2013) das VT auch zur Messung der Geschlechtspräferenz.

Das VT ist ein aufmerksamkeits- bzw. reaktionszeitbasiertes Instrument, bei dem der Proband Bilder von Männern und Frauen in verschiedenen Altersstufen (meist in den Tannerstufen 1-5) betrachtet. In den meisten Studien waren die Probanden während der Bildpräsentation aufgefordert, die Attraktivität der abgebildeten Personen auf einer mehrstufigen Likert-Skala zu bewerten. Als Prädiktor für die sexuelle Orientierung wird die Zeit genutzt, die ein Proband die einzelnen Bilder betrachtet. Das von den Probanden abgegebene Attraktivitätsrating spielt für die eigentliche Auswertung keine Rolle. Andere Autoren (z.B. Bourke & Gormley, 2012) nutzen das VT ohne ihren Probanden während des Betrachtens der Bilder eine spezielle Aufgabe zu geben und erhielten ähnliche Ergebnisse wie Forscher, welche die klassische VT-Variante nutzten.

Als Stimuli werden verschiedene Bildertypen verwendet; ein einheitliches Bilderset liegt bisher nicht vor. Manche Forschergruppen verwendeten digital erzeugte Bilder (wie z.B. beim Programm Affinity; Dombert et al., 2013; Mokros, Butz, Dombert, Santtila, Baeuml, et al., 2011), andere frei verfügbare Bilder aus dem Internet (z.B. Lippa et al., 2010). Meist wurden Ganzkörperbilder verwendet. Imhoff et al. (2010) gelang mit Bildern, die nur die Köpfe von Personen zeigten, typische VT-Effekte zu erzeugen. Ist die Messung der Geschlechtsorientierung das Ziel, werden meist nur Bilder von Erwachsenen genutzt (z.B. Lippa, 2013). Wenn die Altersorientierung gemessen werden soll, kommen zwangsläufig Bilder von Erwachsenen und Kindern beider Geschlechter zur Anwendung (z.B. Schmidt, Gykiere, Vanhoeck, Mann & Banse,

2014), da nur über die relative Reaktion auf kindliche im Vergleich zu erwachsenen Reizen ein Rückschluss auf die Orientierung gezogen werden kann. Diese Variabilität in den genutzten Stimuli, Zielen (Messung der Alters und/oder Geschlechtsorientierung) und Instruktionen (Attraktivitätsrating oder keine bestimmte Aufgabe) erschwert eine Vergleichbarkeit der einzelnen Studien. Für die vorliegende Arbeit sind einzig die psychometrischen Kriterien des VT zur Erfassung der sexuellen Altersorientierung von Bedeutung.

Nachdem ggfs. eine Transformation der Rohdaten vorgenommen wurde (Thornton & Laws, 2009) werden meist Indizes berechnet, welche die relative Präferenz von weiblich zu männlichen bzw. erwachsenen zu kindlichen Reizen angeben. Steht z.B. die Frage im Raum, ob ein Proband pädo- bzw. hebephil ist, werden meist sogenannte Altersindizes gebildet. Dabei werden die mittleren Reaktionszeiten bzgl. der Bilder der Erwachsenen (Tanner 5) genommen und davon die Reaktionszeiten beim Betrachten der Kinderbilder abgezogen (meist Tanner 1-3, teils auch Tanner 1-4; Banse et al., 2010). Werte über Null sprechen dann für eine Teleiophilie, Werte kleiner als Null für eine Pädo- oder Hebephilie. Diese relative Auswertung ist entscheidend, da aus der absoluten Reaktion auf eine bestimmte Stimulus-Kategorie keine validen Schlüsse gezogen werden können. Zum einen erschweren interindividuelle Unterschiede die Interpretation absoluter Reaktionen, zum anderen ist eine erhöhte Latenz bei Reizen, die dem präferierten Reiz ähneln, normal. So nehmen die Reaktionszeiten bei teleiophilen Probanden mit zunehmender körperlicher Reife der abgebildeten Personen zu. Dies geschieht bei beiden Geschlechtern, allerdings ist die Zunahme beim präferierten Geschlecht deutlich ausgeprägter (Imhoff et al., 2010; Imhoff et al., 2012). Diese relative Auswertung wird auch bei anderen objektiven Verfahren empfohlen (Blanchard et al. 2009), allerdings nicht von allen Forschern genutzt (siehe z.B. Abel et al., 2001). Mittlerweile gibt es erste käuflich erwerbbar Testbatterien, die u.a. das VT nutzen: Das Abel Assessment for Sexual Interest (AASI; Abel et al., 2001), das Affinity Tool (Mokros et al., 2013) sowie das Explicit and Implicit Sexual Interest Profile (EISIP; Banse et al., 2010). Aufgrund der Besonderheiten des AASI werde ich im Folgenden zunächst auf die psychometrischen Kennwerte des VT eingehen und anschließend dem AASI einen gesonderten Abschnitt widmen.

1.2.2.2 Theoretische Grundlage

Das VT wird klassischerweise der „aesthetic response“ aus Singer’s Modell (1984) zugeordnet (Flak, Beech & Fisher, 2007). Singer (1984) unterteilte die sexuelle Reaktion in drei sequentiell aufeinanderfolgende Stufen: In der Ersten („aesthetic response“) kommt es zu einer

erhöhten Aufmerksamkeitszuwendung bei sexuell präferierten Reizen, da diese aufgrund ihrer hohen Bedürfnis-Relevanz salienter sind als andere Reize. Im zweiten Schritt („approach step“) kommt es zu einer Hinwendung und ggfs. einer Annäherung zu dem sexuellen Reiz. In der dritten Stufe erfolgt schließlich eine genitale Reaktion („genital response“). Ursprünglich gingen Forscher davon aus, dass der VT Effekt durch eine erhöhte Aufmerksamkeitszuwendung bei präferierten Reizen verursacht wird oder auf dem sogenannten „Sexual Content Induced Delay“ (Geer & Bellard, 1996) beruht. Die Forscher gingen somit von stimulusbasierten Effekten aus. Die Arbeiten von Imhoff et al. (2010) legten allerdings eher kognitive Prozesse als Grundlage des VT Effektes nahe. Die Autoren vermuteten, dass Probanden (beim klassischen VT, bei dem die Attraktivität der präsentierten Reize eingeschätzt werden soll) sequentiell verschiedene Fragen beantworten, um eine Attraktivitätseinschätzung zu treffen. Hierzu muss nach Meinung der Autoren das Geschlecht, das Alter und die körperliche Attraktivität eingeschätzt werden. Diese Hypothese konnten Pohl, Wolters und Ponseti (in press) in ihrer Arbeit bestätigen. Die Probanden wurden gebeten, während sie Bilder von Männern und Frauen betrachteten, entweder anzugeben, (1) ob sie die abgebildete Person sexuell attraktiv fanden (ja/nein) oder (2) welchen Geschlechts die abgebildete Person war (männlich/weiblich). Unter der Annahme, dass der VT-Effekt auf stimulusbasierten Prozessen beruht, wären unter beiden Bedingungen typische Reaktionszeitdifferenzen zwischen präferierten und nicht-präferierten Reizen zu erwarten. Tatsächlich ergaben sich aber nur unter der ersten Bedingung typische VT-Effekte.

Da der VT-Effekt diesen Arbeiten zu Folge auf kognitiven Bewertungsprozessen zu basieren scheint, lässt sich vermuten, dass eine willentliche Beeinflussung der Ergebnisse möglich ist. Diese Vermutung wird durch die Arbeiten von Imhoff et al. (2012) gestützt. In ihren Studien baten sie die Probanden, die Bilder aus vier verschiedenen Perspektiven zu bewerten: Aus der Perspektive eines hetero- oder homosexuellen Mannes und der einer hetero- oder homosexuellen Frau. Dabei ergaben sich je nach vorgegebener Perspektive typische Reaktionsprofile. Es war den Probanden also durch die Instruktion, die Bilder aus einer anderen Perspektive zu bewerten, möglich, ein für diese Gruppe typisches Reaktionszeitmuster zu erzeugen. Sollten die Probanden z.B. die Bilder aus der Perspektive einer heterosexuellen Frau bewerten, brauchten sie für die Bewertung der männlichen Bilder am längsten – weitgehend unabhängig von der eigenen sexuellen Orientierung bzw. des eigenen Geschlechts. Diesen Ergebnissen zufolge basiert der VT-Effekt vermutlich nicht auf der „aesthetic response“ nach Singer (1984). Es scheint vielmehr so, als erfasse das VT das sexuelle Selbstkonzept (Bailey, 2009), welches durch eine simple Instruktion

vorgegeben werden kann (Imhoff et al., 2012) und nicht mit der sexuellen Orientierung übereinstimmen muss. Dies könnte das VT-Verfahren wiederum anfällig gegenüber Verfälschungstendenzen seitens der Probanden machen.

1.2.2.3 Psychometrische Kennwerte

Letourneau (2002), Mokros et al. (2013) und Banse et al. (2010) berichteten für das VT Reliabilitätskennwerte (Cronbachs Alpha) zwischen .72 und .90. Silva (2011) referierte zwar auch sehr gute Werte für die interne Konsistenz ($r > .97$), wies aber auf eine fehlende Retest-Reliabilität des VT hin. Sie konnte keine signifikanten Korrelationen zwischen zwei Testzeitpunkten (in Bezug auf die einzelnen Stimulus-Kategorien) errechnen. Zudem waren die Reaktionszeiten beim zweiten Messtermin geringer als beim Ersten. Silva (2011) berichtete allerdings nicht, wie hoch die Übereinstimmung in Bezug auf die Diagnose zwischen den beiden Messzeitpunkten war. Es ist durchaus denkbar, dass an beiden Testzeitpunkten dasselbe Ergebnis zustande kommt, auch wenn sich die Reaktionszeiten bzgl. der einzelnen Kategorien unterschieden.

Gress (2005) gab in ihrer Studie eine Sensitivität von 84 % bei einer Spezifität von 57 % an. Problematisch dabei ist, dass aus der Stichprobenbeschreibung nicht hervorgeht, wie hoch der Prozentsatz der Pädophilen in der Gruppe der Sexualstraftäter war. Mokros et al. (2013), nach deren Angaben alle getesteten Sexualstraftäter pädophil waren, erreichten auch nur eine ungenügende Spezifität von etwa 50 % bei einer hohen Sensitivität von ca. 90 % für das von ihnen entwickelte Programm Affinity. Worling (2006) berichtete bei jugendlichen Sexualstraftätern ebenfalls von relativ schlechten Validitätswerten (AUC zwischen .42 und .69) bei ausreichender bis guter Reliabilität (.72 - .82). Das VT führte in vielen Studien bei homosexuellen Pädophilen (bzw. Männern, die Kindesmissbrauch zum Nachteil von Jungen begangen habe) zu besseren Validitätskennwerten als bei heterosexuellen Pädophilen (Abel, Huffman, Warberg & Holland, 1998; Banse et al., 2010; Letourneau, 2002; Worling, 2006).

1.2.2.4 Das Abel Assessment for Sexual Interest (AASI)

Abel nutzt seine Technik, um Personen zu identifizieren, die “sexual behavior problems” haben; nicht, um pädophile von teleiophilen Personen zu unterscheiden (siehe <http://diana.public.abelscreen.com>, abgerufen am 22.04.2015). Seit der Gründung von Abel Screen Inc. im Jahre 1995 entwickelte er das AASI für Männer, Frauen sowie Jungen und Mädchen (im Alter von 12-17 Jahren). Es gibt zudem ein gesondertes Instrument für Personen mit geistiger

Behinderung. Es basiert auf einem VT-Verfahren, dem Multiphasic Sex Inventory (Simkins, Ward, Bowman & Rinck, 1989) sowie einem selbst entwickelten Fragebogen. Es lässt sich somit streng genommen nicht zu den reinen VT-Verfahren rechnen. Nach Angaben von Abel et al. (2001) ist es verfälschungsresistent. Veröffentlichte Studien dazu sind allerdings selten und wenig transparent, weswegen es teils stark kritisiert wird (Sachsenmaier & Greess, 2009).

Abel et al. (1998) berichteten relativ gute Validitätskennwerte (zwischen 73 % und 86 %) für die erste Version des AASI, welches ausschließlich auf VT-Werten beruhte. In einer früheren Studie der Autorengruppe (in der die Autoren bereits mehrere Prädiktoren für die sexuelle Orientierung nutzten; Abel, Lawry, Karstrom, Osborn & Gillespie, 1994) wurden ebenfalls gute Validitätskennwerte erhoben (77 bis 98 %; im Mittel: 86.6 %), ohne dass beschrieben wurde, wie die Datenaufzeichnung und -auswertung funktionierte. Die Autoren verglichen die Reaktionen bzw. Antworten von Straftätern (die ihre Taten einräumen) und teleiophilen Kontrollprobanden aus der Allgemeinbevölkerung (die eine Pädophilie bestreiten). Das Ziel der Autoren schien es deshalb zu sein, Personen zu identifizieren, welche potentiell sexuellen Kindesmissbrauch begehen könnten. Im nächsten Satz erwähnen die Autoren allerdings, dass sie Pädophile von „Normalen“ unterscheiden konnten. Somit scheint hier eine Konfundierung der beiden distinkten Konstrukte (Pädophilie und sexueller Kindesmissbrauch) vorzuliegen. Die berichteten Kennwerte lagen bei den Personen höher, die Missbrauch an Jungen begangen haben (bei präpubertären Opfern: Sensitivität 76 %, Spezifität 98 %; bei intrapuberalen Opfern: Sensitivität 90 %, Spezifität 98 %) als bei Personen, die Missbrauch zum Nachteil von Mädchen begangen haben (bei präpubertären Opfern: Sensitivität 91 %, Spezifität 77 %; bei intrapuberalen Opfern: Sensitivität 86 %, Spezifität 77 %). Auch in einer aktuelleren Studie von Abel et al. (2001) zum AASI ist die Zielsetzung des Verfahrens für mich nicht abschließend klar. Auch hier soll den Autoren zufolge eine Differenzierung zwischen Tätern und nicht-Tätern gelingen, wenngleich die Autoren zwischenzeitlich immer wieder von Pädophilen sprechen. So heißt es auf S. 705f.:

Special attention was paid to the selection of admitted child molesters. Those who had only fantasies or urges towards children but had not actually touched a child were excluded to ensure that the child molester group accurately reflected individuals of most concern to the culture, child molesters that actually touch the child. Second, child molesters who had been involved in incest only were excluded. The authors believe that incest sometimes occurs not because of specific sexual interest in children, but secondary to family dynamics, parental absence, availability of the child, and other circumstances. The goal was to identify

a class of child molestation driven by sexual interest in children by only including child molesters who had molested outside of the home.

Durch den oben beschriebenen Ausschluss von Pädophilen, die keine Taten begangen haben, sowie von Personen, die Kindesmissbrauch begangen haben, aber vermutlich nicht pädophil waren (Inzesttätern), ist der wissenschaftliche Nutzen der Studie nur schwer einzuschätzen. Zwar mag es sein, dass es den Autoren gelang, Kontrollprobanden von pädophilen Tätern zu unterscheiden, der Nutzen dieses Tests erschließt sich mir allerdings nicht. Wenn die Autoren ein Screening Tool entwickeln wollen, müsste das Ziel sein, Personen zu identifizieren, die mit hoher Wahrscheinlichkeit Kindesmissbrauch begehen. Ob eine Differenzierung von pädophilen Tätern und pädophilen nicht-Tätern überhaupt gelingen kann, bleibt offen, da eine der beiden Gruppen (pädophile Nicht-Täter) aus der Studie ausgeschlossen worden ist. Ebenso wichtig ist es zu wissen, ob ein Risikoprognosetool Personen identifizieren kann, die Kindesmissbrauch begehen auch wenn sie nicht pädophil sind. Auch diese Frage muss auf Grundlage der von Abel et al. (2001) durchgeführten Studie unbeantwortet bleiben.

Um die Verfälschbarkeit des AASI zu untersuchen haben die Autoren zudem sogenannte Liar-Denier in die Studie aufgenommen. Dabei handelt es sich um Personen, die ihre *Taten* abstritten, bei denen die Therapeuten aber davon ausgingen, dass die Probanden Missbrauch begangen haben (ob sie auch von einem sexuellen Interesse an Kindern ausgingen, bleibt unklar). Manche davon könnten den Autoren zufolge auch Kinder im Alter von 14-17 missbraucht haben. Dies ist sicherlich kein Marker für eine Pädophilie und in einigen Fällen vermutlich nicht einmal für eine Hebephile. 374 Probanden wurden zur Modellerstellung und 373 zur Testung des Modells genutzt (Validierung). 616 Probanden hatten sexuellen Kindesmissbrauch begangen: 77 begingen Missbrauch zum Nachteil von Jungen, 261 zum Nachteil von Mädchen (31 dieser Personen hatten sowohl Jungen als auch Mädchen sexuell missbraucht). Wie die Differenz (324) zwischen den 616 Missbrauchstätern und der Summe aus den beiden Teilkategorien zu erklären ist, lassen die Autoren leider offen. Der Rest der Stichprobe bestand aus Sexualstraftätern oder in Sexualtherapie befindlichen Patienten, die keinen Missbrauch gegenüber Kindern begangen hatten.

Für jede Gruppe wurden andere Prädiktoren genutzt. Aus der VT-Messung wurden jeweils nur die Reaktionszeiten der vermeintlich präferierten Kategorie genutzt (bei Klassifikation von Probanden, die Missbrauch zum Nachteil von Mädchen begangen hatten z.B. die Reaktionszeiten bei Mädchenbildern). Zudem wurden Items aus dem von den Autoren entwickelten Fragebogen genutzt. Bei der Klassifikation von Probanden, die Missbrauch zum Nachteil von Mädchen

begangen hatten, wurden Fragen zur Messung der kognitiven Verzerrungen sowie eine Reihe weiterer Items als Prädiktoren genutzt (z.B. das Item „I feel that I am someone children can look up to“, die selbst berichtete Anziehung durch Mädchen im Alter von 8–10 Jahren, die selbst berichtete Anziehung durch erwachsene Männer, und die Anzahlen an eingegangenen Ehen). Zur Berechnung des Vorhersagemodells wurden dann nur die Kontrollgruppe und die fragliche Vergleichsgruppe einbezogen (z.B. Probanden, die Missbrauch zum Nachteil von Mädchen begangen hatten). Somit ist zur Verwendung dieses Tool ein a priori Verdacht notwendig, zu welcher Gruppe der zu testende Proband gehört. Die Sensitivität dieses Modells lag bei 74 %, die Spezifität bei 73 %. Das Modell für Kindesmissbraucher gegenüber Jungen hatte eine Sensitivität von 86 % und eine Spezifität von 86 %. Bei Liar-Deniers betrug die Sensitivität 75 % und die Spezifität 76 % (in dieses Modell ging überhaupt kein VT-Wert ein). Die Autoren schlussfolgerten auf S. 714, dass:

The clinician could use these results in evaluating patients charged with incest, to identify if they match those who have also molested outside of the home. The clinician could also use these results in dealing with a denying patient since the probability values may identify those who match individuals who have molested outside of the home.

Es scheint den Autoren demzufolge tatsächlich vordergründig um die Identifikation von potentiellen Missbrauchstätern gegangen zu sein, nicht um die Diagnose einer sexuellen Orientierung. Ob dieses Tool unter diesen Bedingungen überhaupt geeignet ist, potentielle Täter zu identifizieren, scheint mir fraglich. Eine unabhängige Replikation steht noch aus. Nichtsdestotrotz wird dieses Instrument bereits vielfach in den USA eingesetzt, wie Abel et al. (2001) auf S. 716 erwähnten:

The Abel Assessment *for sexual interest*TM identical to that described above has been used over 20 000 times at over 270 sites throughout North America by approximately 800 licensed professionals, according to Don Jones of Abel Screening, Inc. (personal communication, February 8, 2000).

Aufgrund der mäßigen Validitätswerte und der intransparenten Darstellung des AASI kommen auch Sachsenmaier und Greess (2009) zu dem Schluss, dass das VT fehleranfälliger als die PPG zu sein scheint. Sie schlugen deshalb vor, das VT eher als Unterstützung im klinischen Interview zu nutzen und nicht als diagnostisches Instrument oder zur Risikoeinschätzung.

1.2.3 Überblick über die Penisplethysmographie

1.2.3.1 Zur Entstehung und Methodik

Freund (1957) entwickelte die PPG ursprünglich zur Messung homosexueller Orientierungen bei Männern. Bei der PPG handelt es sich um ein psychophysiologisches Messverfahren, bei dem die genitale Erregung in Reaktion auf verschiedene sexuelle Reize gemessen wird. Auch die PPG wird heutzutage v.a. zur Messung pädophiler Orientierungen verwendet (Howes, 1995). Entscheidend für den Erfolg einer PPG Messung ist, ob eine ausreichende Erregung bei den Probanden induziert werden kann. Ist die genitale Reaktion zu gering, kann dies zu Problemen bei der Messung der sexuellen Orientierung führen (Kuban, Barbaree & Blanchard, 1999).

Die PPG gilt als valides und reliables Messinstrument, wird allerdings aufgrund einer mangelhaften Standardisierung zu Recht kritisiert (Howes, 1995). O'Donohue und Letourneau (1992) listeten insgesamt 17 Punkte auf, in denen sich Testdurchführung und -auswertung in verschiedenen Kliniken unterscheiden. Aktuell gibt es erneut Bestrebungen, eine Standardisierung zu erreichen (Murphy, Ranger, Stewart, Dwyer & Fedoroff, 2015). Im Folgenden gehe ich auf einige der Faktoren ein, bezüglich derer eine fehlende Standardisierung vorliegt und die die Genauigkeit einer Messung mitbestimmen können.

Zur Messung genitaler Reaktionen stehen zwei Methoden zur Verfügung: (1) Die Volumen-Plethysmographie, bei der sich der Penis in einem luftdichten Zylinder befindet und das durch die Erektion verdrängte Volumen gemessen wird und (2) die Zirkumferenz-Plethysmographie, bei der mittels dehnbare Messschlaufen der Umfang des Penis gemessen wird (Bancroft & Pullan, 1966; Barlow, Becker, Leitenberg & Agras, 1970; Freund, Langevin & Barlow, 1974). Die Volumenmessung wird v.a. von einer Forschergruppe vom Centre for Addiction and Mental Health in Kanada verwendet (z.B. Blanchard & Barbaree, 2005). Alle anderen Forscher und Therapeuten nutzen die weniger genaue aber viel einfacher anwendbare Umfangsmessung (Laws, 2009). Kuban et al. (1999) zeigten, dass die Übereinstimmung zwischen den beiden Methoden bei ausreichender Reaktion hoch ist ($r = .98$). Bei geringen Reaktionen (<2.5 mm) ergab sich hingegen eine mittlere Korrelation von $-.15$. Dies liegt daran, dass der Umfang bei einer einsetzenden Erektion zunächst leicht abnehmen kann, während das Volumen bereits ansteigt (Earls & Marshall, 1983; McConaghy, 1974). Mit der Volumenmessung konnten Kuban et al. (1999) auch bei Low-Respondern (Personen mit sehr geringen Reaktionen während der gesamten

Messung) signifikante Unterschiede in der Reaktion zwischen der präferierten Kategorie und allen anderen Kategorien feststellen. Bei der Umfangsmessung gelang dies bei Low-Respondern nicht. Deshalb schlugen Kuban et al. (1999) als minimales Reaktionskriterium bei Umfangsmessungen 2,5 mm vor. Harris, Rice, Quinsey, Chaplin und Earls (1992) kamen zwar zu dem Schluss, dass durch einen Ausschluss von Low-Respondern (bei der Umfangsmessung) keine Verbesserung der psychometrischen Kennwerte der PPG erreicht werden kann, die meisten Forscher schließen Low-Responder allerdings aus ihren Analysen aus (Howes, 1995). Marshall und Eccles (1991) sowie Malcolm et al. (1993) berichteten, dass sie nur Reaktionsprofile interpretieren, die mindestens 10 % einer vollständigen Erektion aufweisen. Dies entspricht im Mittel etwa einer Umfangsänderung von 4.6 mm (Howes, 2003). Viele Praktiker nutzen als minimales Reaktionskriterium allerdings 20 % einer vollständigen Erektion, was im Mittel einer Umfangsänderung von 9.4 mm entspricht (Howes, 2003). Bei Volumenmessung wird 1 cm³ oft als minimales Reaktionskriterium genutzt (z.B. Blanchard et al., 2001). Der Ausschluss von Low-Respondern führte bei verschiedenen Studien zu einem drop-out zwischen 20 % und 75 % (Looman et al., 1998).

Geringe Reaktionen innerhalb einer Messung können dabei nicht nur auf absichtliche Verfälschung der Messung (Mahoney & Strassberg, 1991) und soziale Erwünschtheit (Barbaree, Marshall, Yates & Lightfoot, 1983) hindeuten, sondern z.B. auch durch das Alter des Probanden (Blanchard & Barbaree, 2005; Hall, 1991) oder eingenommene Medikation (Kolla, Klassen, Kuban, Blak & Blanchard, 2010; Lykins, Robinson, LeBlanc & Cantor, 2014) beeinflusst werden. Auch Angst kann die genitale Reaktion hemmen (Hale & Strassberg, 1990). Dies könnte zumindest partiell erklären, warum Personen, die wegen einer Sexualstraftat angeklagt sind, weniger stark in einer PPG Messung reagieren als Personen, die bereits verurteilt worden sind (Blanchard et al., 2007; Kalmus & Beech, 2005) und muss nicht *ausschließlich* auf bewussten Verfälschungsabsichten der Probanden beruhen.

Eine Variable, welche die Reaktionsstärke und die Diskriminationsleistung der PPG beeinflusst, ist die Zeit, die ein Stimulus dargeboten wird. Lalumiere und Earls (1992) fanden heraus, dass die Umfangszunahme bei ihren Probanden bei präferierten Stimuli positiv mit der Präsentationszeit korreliert war. Um eine möglichst hohe Reaktion zu gewährleisten, scheint eine Präsentationszeit im Rahmen von 2.4 bis 3.4 Minuten besonders geeignet zu sein (Earls, Morales & Marshall, 1988). Auch Launay (1999) weist darauf hin, dass eine blockweise Darbietung von Stimuli – was zu einer längeren Präsentation einzelner Kategorien führt – die Diskriminationsleistung der PPG erhöht.

Auch bezüglich der verwendeten Stimuli gibt es eine erhebliche Variabilität. Die liegt zum einen daran, dass einige Autoren (siehe z.B. Abel et al., 1998) die Darstellung von nackten Kindern für besonders problematisch halten. Solange dies im Rahmen der Erfüllung von „dienstlichen oder beruflichen Pflichten“ (§ 184b StGB, Absatz 5) geschieht, drohen dem Testleiter in Deutschland keine rechtlichen Konsequenzen. In den USA ist die Präsentation nackter Kinder als Stimuli juristisch untersagt, weshalb hier auf auditives Material zurückgegriffen werden muss (Murphy et al., 2015). Bezüglich der Potenz auditiver und visueller Stimuli fanden Abel et al. (1981) keinen Unterschied in der Stärke der Reaktion zwischen den beiden Stimulus-Modalitäten. Golde, Strassberg und Turner (2000) hingegen berichten, dass audiovisuelle Stimuli – zumindest bei wiederholter Darbietung – rein auditiven Stimuli überlegen sind. Auch Lykins et al. (2010) zeigten, dass eine audiovisuelle Präsentation zu höheren Reaktionen führt als eine rein visuelle. Zudem führten höhere Reaktionen zu höherer Übereinstimmung zwischen zwei Messzeitpunkten. Die Konsistenz der Diagnosen war unabhängig von dem Zeitintervall (im Mittel 48 Monate) und die Retest-Reliabilität betrug (bei einem minimalen Reaktionskriterium von 1 mm³) 81 %. Bei einem Test der sexuellen Altersorientierung wird die Verwendung von visuellem Material empfohlen (Lalumière & Harris, 1998), da hier der körperliche Reifegrad besser abgebildet werden kann als durch Altersangaben, wie sie bei auditivem Material verwendet werden. Harris et al. (1992) empfahlen zudem, neben kindlichen Stimuli (Tanner 1) auch intrapubereale Stimuli zu verwenden (Tanner 2-3), da dies die Diskriminationsleistung der PPG erhöht.

Eine Bedrohung für die Validität der PPG stellt die Verfälschbarkeit der Messergebnisse durch den Probanden dar (Laws & Holmen, 1978), was bereits Freund (1961) zeigte. Dabei scheint es leichter, eine andere Altersorientierung als eine andere Geschlechtsorientierung zu simulieren (Freund, Watson & Rienzo, 1988). Probanden können mittels kognitiver und behavioraler Techniken versuchen, die Messergebnisse zu beeinflussen. Das willentliche Herbeiführen von Erektion (Freund et al., 1988) sowie Messartefakte können durch die visuelle Inspektion der Reaktionsprofile teilweise entdeckt werden. In neueren Studien gaben Probanden allerdings meist an, kognitive Methoden zur Beeinflussung ihrer Erektion zu nutzen; muskuläre Anspannung scheint also kaum eine Rolle zu spielen (Golde et al., 2000).

Um die Verfälschung mittels kognitiver Techniken zu erschweren, eignet sich unter anderem die Nutzung auditiver Stimuli. So konnten in einer Untersuchung von Quinsey und Chaplin (1988) 15 von 16 teleiophilen Probanden ihre Reaktion ausreichend beeinflussen. Dies gelang aber nur 2 Probanden, wenn sie parallel zur Stimuluspräsentation eine sogenannte

„Semantic Tracking Task“ (STT) ausführen sollten. Hierbei wurden die Probanden gebeten, eine Taste zu drücken, wenn sie ein sexuelles Wort gehört haben und eine andere Taste, wenn sie ein aggressives Wort gehört haben. Proulx, Côte und Achille (1993) konnten die Eignung dieses Verfahren bei homopädophilen Männern und Harris, Rice, Chaplin und Quinsey (1999) bei Vergewaltigern belegen. McAnulty und Adams (1991) gelang es jedoch nicht, diese vielversprechenden Ergebnisse zu replizieren. Den Probanden war es in dieser Studie weiterhin möglich, ihre genitale Erregung bei präferierten Reizen zu reduzieren. Zudem führte die Verwendung der STT generell zu einer Reduktion der genitalen und subjektiven Erregung. Auch Golde et al. (2000) berichteten, dass ihre studentischen Probanden auch bei Zuhilfenahme der STT in der Lage waren, ihr Erregung bei präferierten Reizen substantiell zu unterdrücken. In dieser Studie sollten allerdings alle Probanden diese Aufgabe ausführen. Ob es den Probanden ohne diese Aufgabe noch besser gelungen wäre, ihre Erregung zu hemmen, ist somit nicht klar. Andere Methoden zur Kontrolle von Verfälschungstendenzen waren hingegen nicht erfolgreich. Laws und Rubin (1969) zeigten ihren Probanden sexuelle Videos. Die Versuchspersonen sollten eine Taste drücken, wenn auf dem Bildschirm ein kleines Licht aufleuchtete. Trotz dieser Aufgabe – die sicherstellen sollte, dass die Probanden die Filme anschauten – konnten sie ihre Erregung unterdrücken, wenn sie dazu aufgefordert wurden. Sie nutzten dazu meist kognitive Techniken (z.B. Rechenaufgaben), um sich abzulenken. Die Probanden hatten aber große Schwierigkeiten, in Abwesenheit sexuell ansprechender Reize willentlich eine Erektion herbeizuführen. Selbst, wenn die Probanden gebeten wurden, das Gesehene kontinuierlich zu beschreiben, waren sie in der Lage ihre Erregung substantiell zu minimieren (Henson & Rubin, 1971). Dies lag vermutlich daran, dass diese Aufgabe zu viel kognitive Kapazität erforderte und somit die Erektionsstärke negativ beeinflusste (Geer & Fuhr, 1976).

Des Weiteren gibt es verschiedene Arten der Datenaufbereitung und -auswertung (Earls, Quinsey & Castonguay, 1976). Während Earls, Quinsey und Castonguay (1987) zeigten, dass z -transformierte Rohwerte einen höheren Varianzanteil aufklärten als z.B. Rohwerte, sprechen sich andere Forscher gegen die Verwendung von z -Transformationen aus (Hall, Proctor & Nelson, 1988; Murphy & Barbaree, 1994). Die inkonsistenten Befunde zur Verwendung von z -Werten könnten in Teilen durch das Ausgangswertgesetz von Wilder erklärbar sein (Polak & Knobloch, 1957). Dieses Gesetz besagt, dass die Stärke einer physiologischen Reaktion negativ mit dem Ausgangsniveau korreliert ist. Bei penilen Reaktionen scheint das Ausgangsniveau allerdings nur unsystematisch mit der maximalen Reaktion korreliert zu sein (Julien & Over, 1981; Masters &

Johnson, 1963). Daraus schlussfolgerten Julien und Over (1981), dass eine Korrektur um den Ausgangswert bei der PPG nicht notwendig ist. Wenn nun in der Studie von Earls et al. (1987) die Korrelation zwischen Ausgangswert und Reaktionsstärke zufälligerweise stark negativ gewesen sein sollte, könnte dies erklären, warum eine z -Transformation (welche über die Subtraktion des personenzentrierten Mittelwertes eine Korrektur um das Ausgangsniveau enthält) der Verwendung von Rohwerten überlegen war. Bei ausreichend großen Stichproben sollte dies allerdings keinen Effekt haben. Julien und Over (1981) schlugen deshalb vor, den Zusammenhang zwischen Ausgangsniveau und Reaktionsstärke zu prüfen.

Teilweise wird eine z -Transformation der Daten auch vorgeschlagen, um Low-Responder nicht ausschließen zu müssen. Howes (2003) ging allerdings davon aus, dass es in Bezug auf Low-Responder nicht ausreicht, eine z -Transformation durchzuführen. Er empfahl daher, alle Personen mit einer maximalen Umfangszunahme von weniger als 9,4 mm im Verlauf der gesamten Messung (entspricht ca. 20 % einer vollständigen Erektion) von der weiteren Auswertung auszuschließen.

Wie beim VT ist auch bei der PPG die Verwendung von Präferenzindizes empfohlen (Blanchard, Kuban, et al., 2009). Laut Lalumière und Harris (1998) eignen sich besonders subtraktiv gebildete Indizes. Bei der Bildung solcher Altersindizes wird z.B. die mittlere Reaktion auf die kindlichen Reize von der mittleren Reaktion auf die erwachsenen Reize (beider Geschlechter) abgezogen. Dieser Index gibt somit ein Ausmaß der Erregung bei erwachsenen *im Verhältnis* zu kindlichen Reizen an. Die kanadische Forschergruppe um Blanchard (siehe z.B. Blanchard et al., 2001) verwendet allerdings einen komplizierteren, aus mehreren Werten zusammengesetzten, Index zur Analyse von PPG-Daten. Wormith (1986) berichtete hingegen, dass eine einfache Messung der maximalen Reaktion ausreichend und anderen Auswertungsvarianten nicht unterlegen ist.

Howes (2009) schlug neben der herkömmlichen Messung der sexuellen Orientierung mittels PPG noch eine weitere Nutzungsmöglichkeit im forensischen Kontext vor. In dieser bat er die Probanden, seine Erregung auf deviante Reize, auf die der Proband im ersten Durchgang substantiell reagiert hat (d.h. pädophile Reize sowie Darstellungen sexueller Gewalt), zu reduzieren. Gelang ihnen diese Inhibition nicht, so schlussfolgerte Howes (2009), dass ihm dies auch in Alltagssituationen nicht gelingen wird, was seiner Ansicht zufolge die Gefahr für (erneute) Sexualstraftaten erhöht. Für diese These fehlen aber noch empirische Belege. Bisher geschieht die Vorhersage von erneuten Sexualstraftaten über die beschriebenen Alters- bzw. Präferenzindizes (Hanson & Morton-Bourgon, 2005).

1.2.3.2 Theoretische Grundlage

Die Messung der genitalen Reaktion mittels PPG lässt sich der dritten Phase aus Singers (1984) Modell zuordnen. Nach Bailey (2009) misst die PPG die sexuelle Erregung und ist somit der beste Indikator für die sexuelle Orientierung. Dabei lassen sich nach Blanchard et al. (2010) Alters- und Geschlechtsorientierung auf einer Achse abtragen. Die Autoren konnten zeigen, dass die Präsentation sexuell präferierter Reize (z.B. vorpubertäre Mädchen bei heteropädophilen Männern) die stärkste genitale Reaktion hervorruft. Bei den sogenannten Nachbarkategorien (in dem Beispiel: intrapubertäre Mädchen und vorpubertäre Jungen; siehe Tabelle 1) ist ebenfalls noch eine substantielle genitale Reaktion in Bezug auf sexuelle Reize zu erwarten. Die Stärke der Reaktion nimmt mit zunehmender morphologischer Unähnlichkeit in Bezug auf den präferierten Reiz ab. Blanchard et al. (2010) konnten ebenfalls zeigen, dass heteropädophile Probanden stärker bei der Präsentation erwachsener Frauen reagierten als homopädophile bei der Präsentation erwachsener Männer. Zudem konnten Freund und Blanchard (1989) zeigen, dass homoteleioophile Männer nicht stärker auf vorpubertäre Jungen reagieren als heteroteleioophile Männer auf vorpubertäre Mädchen. Dies legt die Vermutung nahe, dass die PPG homopädophile Männer besser identifizieren kann als heteropädophile Männer und passt zu den Befunden, dass homopädophile Männer mittels VT besser diagnostiziert werden können als heteropädophile Männer (Abel et al., 1998; Banse et al., 2010; Letourneau, 2002; Worling, 2006) und zu der Tatsache, dass homopädosexuelle Handlungen einen stärkeren Indikator für eine Pädophilie liefern als heteropädosexuelle Handlungen (Seto & Lalumiere, 2001).

Tabelle 1: Zuordnung der sexuellen Alters- und Geschlechtsorientierungen zu den jeweils präferierten Reizen (für Männer). Es ist nach Blanchard et al. (2010) davon auszugehen, dass es bei benachbarten Kategorien noch zu einer substantiellen Erregung kommen kann, die mit zunehmender Distanz von der präferierten Kategorie abnimmt.

Präferierte Reize	Erwachsene Frauen	Intrapubertäre Mädchen	vorpubertäre Mädchen	Vorpubertäre Jungen	Intrapubertäre Jungen	Erwachsene Männer
Sexuelle Orientierung	Hetero-teleiophil	Hetero-hebephil	Hetero-pädophil	Homo-pädophil	Homo-hebephil	Homo-teleiophil

1.2.3.3 Psychometrische Kennwerte

Die Messung des Umfangs mittels PPG an sich ist als reliabel einzuschätzen. Farkas et al. (1979) berichteten bei wiederholter Messung ohne sexuelle Stimulation eine Retest-Reliabilität von .97 und von .75, wenn eine vollständige Erektion gemessen wurde. Becker, Hunter, Goodwin,

Kaplan und Martinez (1992) berechneten Retest-Reliabilitäten von auditiven Stimuli (gemessen mit der Umfangs-PPG) bei jugendlichen Straftätern. Die Stimuli enthielten u.a. Geschichten über einvernehmliche sexuelle Kontakte, Vergewaltigungsszenarien, pädosexuelle Kontakte und Exhibitionismus. Sie berichtete die Ergebnisse über alle Probanden, die sich in den begangenen Straftaten (und somit vermutlich auch in ihren sexuellen Orientierungen) unterschieden. Becker et al. (1992) erreichten Retest-Werte zwischen .21 und .82 (im Mittel: 0,61). Die Reliabilitätswerte waren bei den Stimuli am höchsten, die den von den Tätern begangenen Taten entsprachen. Frenzel und Lang (1989) untersuchten die interne Konsistenz und die Retest-Reliabilitäten der Volumenmessung bei erwachsenen Sexualstraftätern (die meisten hatten Kindesmissbrauch zum Nachteil von Mädchen begangen) und heterosexuellen Kontrollprobanden mittels kurzer Filmclips. Sie referierten eine interne Konsistenz von .93 und Retest-Werte (Messung am selben Tag) von .13 - .87 (im Mittel: .65; ohne neutrale Stimuli: .71; über alle weiblichen Stimuli: .82). Auch hier wurden die geringsten Kennwerte bei nicht-präferierten Stimuli erreicht. Day, Miner, Sturgeon und Murphy (1989) berichteten ebenfalls gute Kennwerte für die interne Konsistenz (im Mittel .89). Auch Fernandez (2000) erzielte ausreichende Werte für die interne Konsistenz bei Inzesttätern und Kindesmissbrauchern gegenüber extrafamiliären Opfern von .84 bis .95. In Bezug auf die Retest-Reliabilität erhielt sie geringere Werte (.13 - .75; im Mittel: .45). Für die über Subtraktion gebildeten Devianz-Indizes erhielt Fernandez (2000) Werte zwischen .36 und .79 (im Mittel: .61).

Geringe Reaktionen sind auch deshalb ein Problem für die PPG (Merdian & Jones, 2011), da die Kennwerte für die Reliabilität bei geringen Reaktionen niedriger als bei höheren Reaktionen ausfallen (Barbaree, Baxter & Marshall, 1989; Becker et al., 1992; Frenzel & Lang, 1989; Hunter, 1994). Dies ist aufgrund der eingeschränkten Varianz bei Low-Respondern zu erwarten (O'Donohue & Letourneau, 1992). Die Konsistenz in Bezug auf die Diagnose bei zwei Messzeitpunkten lag in einer Studie bei 81 % (Lykins et al., 2010).

In einer frühen Arbeit von Freund, Chan und Coulthard (1979) gelang es den Autoren mittels Volumenmessung, 96 % der päd- oder hebephilen Täter korrekt zu diagnostizieren. Bei Tätern, die ihre Taten bzw. eine sexuelle Präferenz für Kinder oder intrapubertäre Jugendliche abstritten, lag die Klassifikationsgenauigkeit bei 66 %. Werte für die Spezifität wurden von Freund et al. (1979) nicht berichtet. Sollte neben der Alterspräferenz gleichzeitig auch die Geschlechtspräferenz vorhergesagt werden, sank die Klassifikationsgenauigkeit ab (auf 66 % bzw. 33 %). Malcolm et al. (1993) untersuchten u.a. die Validität ihrer Umfangs-PPG und kamen dabei auf eine mittlere Klassifikationsgenauigkeit von 76 % (Spezifität 95 % und Sensitivität von 41 %).

Problematisch dabei ist, dass in dieser Studie die Gruppeneinteilung rein anhand von Straftaten vorgenommen wurde. Wie hoch der Anteil von Admittern und/oder Pädophilen in der Straftäterpopulation war, gaben die Autoren nicht an. Eine außergewöhnlich hohe Klassifikationsgenauigkeit erreichten Chaplin, Rice und Harris (1995). Sie nutzten auditive Stimuli, in denen das Leiden der kindlichen Opfer explizit dargestellt wurde und erhielten eine Klassifikationsgenauigkeit von 100 %. Die Ergebnisse basieren allerdings auf einer relativ kleinen Stichprobe ($N = 30$). In einem viel zitierten Referenzpaper gaben Blanchard et al. (2001) für ihre Methode der Volumen-PPG Werte für die Spezifität und Sensitivität im Bereich von 96 % bzw. 61 % an. Dabei bestand die gesamte Experimentalgruppe aus Non-Admittern. Nach Blanchard et al. (2001) stellen die berichteten Kennwerte somit einen unteren Grenzwert für die Validität der PPG dar. O'Donohue und Letourneau (1992) kommen in ihrem Review zu dem Schluss, dass die prädiktive Validität bei etwa 80 bis 96 % liegt, wenn Admitter untersucht werden und, dass dieser Wert geringer ist, wenn Non-Admitter in die Analysen aufgenommen werden. Diese Ergebnisse sprechen für eine ausreichende Validität, die allerdings durch Verfälschungstendenzen seitens der Probanden bedroht ist (Blanchard et al., 2007). Allerdings ist es durchaus möglich, dass einige Leugner tatsächlich keine pädophile Orientierung aufweisen, weshalb die Bestimmung der Validität in Stichproben mit Leugnern immer einen minimalen Schätzer für die Validität ergibt (Blanchard et al., 2001). Keine der aufgeführten Studien, die Non-Admitter eingeschlossen hat, nutzte die STT (Quinsey & Chaplin, 1988) oder ähnliche Methoden zur Minimierung von Verfälschung.

1.2.4 Überblick über die funktionelle Magnetresonanztomographie (fMRT)

Während VT und PPG speziell zur Messung der sexuellen Orientierung entwickelt worden sind, ist diese Art der Nutzung der fMRT nur eine von vielen Anwendungsmöglichkeiten (Ponseti et al., 2009). Die Magnetresonanztomographie geht zurück auf die Forschung von Lauterbur (1973) und wurde entwickelt, um non-invasive, strukturelle Aufnahmen von verschiedenen Organen machen zu können. Zunächst wurde die Technik daher ausschließlich zur strukturellen Aufnahme verschiedener Gewebetypen genutzt (strukturelle Magnetresonanztomographie; sMRT). Später wurde diese Technik weiterentwickelt, um die Aktivität des Gehirns messen zu können (fMRT; Buxton, 2009). Bei erhöhter Aktivität in einzelnen Hirnarealen – die wiederum mit bestimmten Aufgaben assoziiert sein kann – kommt es zu einer vermehrten Durchblutung dieser Areale, um eine ausreichende Versorgung mit Sauerstoff sicherstellen zu können. Hämoglobin, der Stoff, der für den Transport von Sauerstoff im Körper verantwortlich ist, besitzt dabei im oxygenierten Zustand andere magnetische Eigenschaften als im deoxygenierten Zustand. Diese Tatsache wird als BOLD-Effekt (BOLD: blood oxygenation-level dependent) bezeichnet. Bei ansteigender

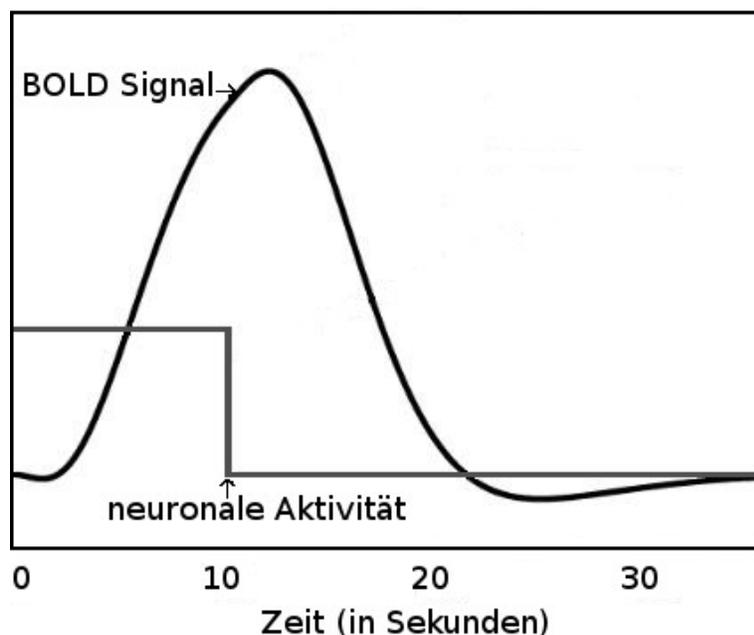


Abbildung 2: Schematische Darstellung einer BOLD Antwort (hämodynamische Antwortfunktion, HRF) auf einen 10-sekündigen Reiz mit einhergehender, erhöhter neuronaler Aktivität. Dabei ist klassischerweise zuerst ein kurzzeitiger „undershoot“ (Abnahme des oxygenierten Hämoglobins) zu verzeichnen, gefolgt von einer Zunahme der Sauerstoffsättigung mit einem Maximum nach ca. 6 Sekunden nach Ende der erhöhten neuronalen Aktivität.

neuronaler Aktivität kommt es aufgrund des erhöhten Stoffwechsels zunächst zu einer Abnahme des oxygenierten Hämoglobins. Im Anschluss nimmt die Durchblutung in den aktiven Arealen zu, was zu einer erhöhten Konzentration von oxygenierten Hämoglobin führt. Das Maximum an

oxygenierten Hämoglobin liegt etwa sechs Sekunden nach Ende der erhöhten neuronalen Aktivität (also z.B. Beendigung der Aufgabe bzw. des Stimulus) vor und fällt über ca. 10-25 Sekunden wieder ab. Die Verteilung von oxygenierten zu deoxygenierten Hämoglobin ist die Grundlage für die Messung der neuronalen Aktivität (Ashby, 2015). Für eine schematische Abbildung eines BOLD-Signals (die hämodynamische Antwortfunktion, kurz: HRF) siehe Abbildung 2. Die fMRT ist demzufolge ebenfalls ein psychophysiologisches Messverfahren, welches allerdings zentralnervöse (und nicht wie die PPG periphere) Prozesse misst.

MRT-Arbeiten haben sowohl strukturelle als auch funktionelle Unterschiede zwischen Patienten mit einer Depression (für einen Überblick siehe: Goodkind et al., 2015; Groenewold, Opmeer, de Jonge, Aleman & Costafreda, 2013) oder Angst- bzw. Zwangsstörungen (für einen Überblick siehe: Damsa, Kosel & Moussally, 2009; Pujol et al., 2004) identifizieren können. Da die Inter-Rater-Reliabilität mancher psychiatrischen Erkrankungen nicht zufriedenstellend ist (Maj, Pirozzi, Formicola, Bartoli & Bucci, 2000; McGorry et al., 1995) und es an Biomarkern eben solcher Erkrankungen fehlt, haben Wissenschaftler zunehmend versucht, auch in diesem Bereich den diagnostischen Prozess durch objektive Marker zu unterstützen. Die vormals entdeckten strukturellen und anatomischen Unterschiede zwischen Kontrollprobanden und Patienten sollten als Indikator für psychische Erkrankungen dienen. Mittels automatischer Klassifikationsalgorithmen gelang es Kloppel et al. (2008) bzw. Duchesne et al. (2008) mit einer Trefferquote von über 90 % auf Grundlage von sMRT-Messungen vorherzusagen, ob der Proband an Alzheimer erkrankt war oder nicht. Soriano-Mas et al. (2007) wendeten dieselbe Technik bei einer Stichprobe von Patienten mit Zwangsstörungen und einer Gruppe von gesunden Kontrollprobanden an. Es gelang ihnen anhand der sMRT-Daten 93.1 % ihrer Probanden korrekt zu klassifizieren (nach Kreuzvalidierung). Als die Autoren ihren Algorithmus auf eine neue Stichprobe anwandten, sank die Klassifikationsgenauigkeit allerdings auf 76.6 %. Fu et al. (2008) belegten, dass sich die automatischen Klassifikationsalgorithmen auch auf fMRT-Daten anwenden lässt. Die Autoren zeigten Patienten, die an einer Depression litten sowie gesunden Kontrollpersonen Bilder von traurigen Gesichtern in drei verschiedenen Intensitäten. Auf Basis der unterschiedlichen Hirnaktivität zwischen den Gruppen gelang es den Autoren das Vorliegen einer Depression mit einer mittleren Genauigkeit von 86 % vorherzusagen.

Parallel dazu zeigten sowohl Kranz und Ishai (2006) als auch Ponseti et al. (2006), dass die zentralnervöse Verarbeitung sexueller Reize abhängig ist von der sexuellen Orientierung der Probanden. Für einen Überblick zur zentralnervösen Verarbeitung sexueller Reize verweise ich auf

die Übersichtsarbeit von Stoleru, Fonteille, Cornelis, Joyal und Moulier (2012). Ponseti et al. (2009) nutzten diese Befunde, um die sexuelle Geschlechtsorientierung ihrer hetero- und homosexuellen Probanden mittels entsprechender automatischer Klassifikationsalgorithmen vorherzusagen. Dazu zeigten sie ihren männlichen Probanden je 30 Bilder von männlichen bzw. weiblichen Genitalien. Jeder Stimulus wurde für 300 ms präsentiert (bei einem Interstimulusintervall von 2.7 Sekunden). Die Stimuluspräsentation erfolgte in einer pseudorandomisierten Reihenfolge. Die Reaktionen der Probanden auf die zwei Stimuluskategorien wurden an einen automatischen Klassifikator übermittelt und führten (nach Kreuzvalidierung) zu einer Klassifikationsgenauigkeit von 90 %.

Während sich herausstellte, dass sich Pädophile und Teleiophile besser anhand der sexuellen Reizverarbeitung als anhand struktureller Unterschiede trennen lassen (Gerwinn et al., 2015; Schiffer, Krueger, et al., 2008; Schiffer, Paul, et al., 2008; Walter et al., 2007), gelang es Ponseti et al. (2012), pädophile und teleiophile Probanden anhand ihrer Hirnantworten voneinander zu unterscheiden. Sie nutzen dabei ein sehr ähnliches Paradigma wie Ponseti et al. (2009). Insgesamt nahmen 24 pädophile und 32 teleiophile Männer an der Studie teil. Die Forscher präsentierten den Probanden für eine Sekunde Bilder von Erwachsenen und von Kindern beider Geschlechter (Gesichter-, Ganzkörper- und Genitalaufnahmen). Ponseti et al. (2012) gelang es, die sexuelle Altersorientierung der Probanden mit einer mittleren Genauigkeit von 95 % vorherzusagen (Spezifität: 100 %, Sensitivität 88 %). Dies erreichten sie, indem sie eine Gruppenkarte anfertigten, die die erhöhte Aktivierung bei erwachsenen Reizen (im Vergleich zu kindlichen Reizen) bei Pädophilie und Teleiophilie verglich; d.h. dass in dieser Karte die Hirnarale abgebildet waren, in denen Teleiophile bei erwachsenen Stimuli (im Vergleich zu kindlichen Stimuli) eine erhöhte Aktivierung aufwiesen als Pädophile. Die Aktivierungen der einzelnen Probanden beim Betrachten von erwachsenen im Vergleich zu kindlichen Stimuli wurden anschließend mit dieser Gruppenkarte multipliziert. Positive Werte deuteten dabei auf eine Teleiophilie, negative auf eine Pädophilie hin (für eine detailliertere Beschreibung des Auswertungsalgorithmus siehe 2.2.4.3).

Zur Klassifikation nutzten Ponseti et al. (2012) die Reaktion der Probanden auf Ganzkörper- und Genitalabbildung. In einer weiteren Auswertung zeigte sich, dass die Klassifikation mit annähernd gleicher Genauigkeit erreicht werden konnte, wenn nur die Gesichtsbilder als Stimuli genutzt wurden (Ponseti et al., 2014; Ponseti et al., 2015). Bei einer Reanalyse der Daten stellte sich heraus, dass – sofern nur eine Stimuluskategorie zur Klassifikation

genutzt werden soll – die Genitalbilder tatsächlich die schlechteste und die Gesichtsbilder die beste Klassifikationsgenauigkeit erzielten (Ponseti, persönliche Mitteilung vom 07.03.2014).

Die vorläufigen Ergebnisse dieser Studie lassen hoffen, dass eine Klassifikation der sexuellen Orientierung auch anhand von Gesichtsbildern möglich ist. Bisher steht allerdings eine Replikation der Ergebnisse anhand einer neueren und im besten Fall größeren Stichprobe aus. Zudem ist bisher unklar, wie reliabel und verfälschungsresistent diese Methode ist. Mit dieser Frage beschäftigt sich die Doktorarbeit meiner Kollegin Frau Gerwinn (vorläufiger Titel: Validity and Reliability of an fMRI-based Classification Tool for the Assessment of Pedophilia). Des Weiteren gelang Ponseti et al. (2012) die Klassifikation pädophiler und teleiophiler Probanden. Ob auch hebephile Probanden mittels dieser Methode ausreichend genau von teleiophilen Probanden abgegrenzt werden können, ist noch fraglich.

Gemessen – und zur Klassifikation der Probanden genutzt – wurde also die präferenzspezifische Aktivierung beim Betrachten verschiedener Bildergruppen. Es ist durchaus denkbar, dass mit diesem Verfahren alle drei Bereiche aus Singer (1984) erfasst werden: die aesthetic response (über Aufmerksamkeitsprozesse sowie erhöhte Aktivierung des dopaminergen Systems), die approach response (mentale Vorbereitung von Handlungen, Fantasietätigkeit, Aktivierung von Spiegelneuronen) und die genital response (zentralnervöse Vorbereitung genitaler Reaktionen). Auch im fMRT scheinen sich heteropädophile und heteroteleiophile Männer ähnlicher zu sein als homopädophile und homoteleiophile Männer. In der Arbeit von Ponseti et al. (2012) wies die männliche Gruppenkarte höhere *t*-Werte und mehr signifikante Voxel auf als der weibliche Kontrast. Dies deutet darauf hin, dass sich homopädophile und homoteleiophile Männer in der Aktivierung verschiedener Hirnregionen beim Betrachten Erwachsener (im Vergleich zu kindlichen Reizen) stärker unterscheiden als heteropädophile und heteroteleiophile Probanden.

1.2.5 Vergleichende Arbeiten

Bisher gibt es eine Handvoll von Arbeiten, die verschiedene objektive Verfahren miteinander verglichen haben. Keine dieser Arbeiten hat bisher die Klassifikationsgenauigkeit von fMRT im Vergleich zu PPG bzw. VT untersucht. Gray und Plaud (2005) testeten ihre 39 pädophilen Personen, die sexuellen Kindesmissbrauch begangen hatten, mittels PPG und der VT-Methode aus dem AASI (Abel et al., 1998) – den Anteil von Non-Admittern in der Stichprobe gaben die Autoren nicht an. Es wurden nur Personen in die Studie aufgenommen, die mindestens

10 % einer vollständigen Erektion erreichten. Sie ermittelten eine Klassifikationsgenauigkeit für VT von 79 % und für PPG von 64 %. Anschließend teilten sie ihre Stichprobe in „reflexive“ (Personen, bei denen angenommen wurde, dass sie ihre sexuelle Orientierung verfälscht darstellten) und „non-reflexive“ Probanden. Um als „reflexive“ Proband klassifiziert zu werden, mussten zwei Bedingungen erfüllt sein (Gray und Plaud, 2005, S. 5):

(1) [...] the participant must have responded on average with a consistent subjective rating to all classes of stimuli that were not adolescent or adult males and females (depending on their sexual orientation towards males/females). For example, if the participant was a heterosexual male, to be classified as a reflexive responder, he might on average score a subjective rating of one [which was anchored as highly sexually disgusted; Anm. des Verfassers] for all the non-adolescent, non-adult stimulus categories, as well as a one for the adolescent and adult male categories, and score any other number on average for the adolescent and adult female stimulus categories.

(2) Der Unterschied zwischen dem größten und kleinsten Standardwert² durfte nicht mehr als 0,75 Standardeinheiten betragen. Daraus folgte, dass 28 der 39 Probanden (72 %) als „non-reflexive“ identifiziert wurden. In dieser Substichprobe wurden durch VT nun 96 % und durch PPG 68 % korrekt klassifiziert (signifikant unterschiedlich). Bei den reflexiven Probanden erreichten VT und PPG nur eine Genauigkeit von 36 bzw. 55 % (kein signifikanter Unterschied). Neben der fehlenden Kontrollgruppe ist die Art der PPG Auswertung ein großer Schwachpunkt der Untersuchung. Statt einen Differenzindex zu nutzen, definierten die Forscher einen Grenzwert von 75 % einer vollständigen Erektion. Jeder Proband, der diesen Grenzwert bei „devianten“ Stimuli überschritt, wurde als pädophil diagnostiziert. Dies widerspricht den üblichen Auswertungsempfehlungen (Blanchard, Kuban, et al., 2009; Harris et al., 1992).

Mackaronis, Byrne und Strassberg (2014) verglichen die Ergebnisse von PPG, VT (Affinity) und der „Screening Scale for Pedophilic Interest“ (SSPI; Seto et al., 2004; Seto & Lalumiere, 2001) bei 103 jugendlichen Sexualstraftätern (Alter 15-20). Eine Kontrollgruppe war nicht in der Studie enthalten, weswegen keine Werte für Spezifität und Sensitivität berichtet werden konnten. Von den 103 Probanden haben nur 16 das VT durchgeführt. Low-Responder wurden anhand von zwei Kriterien bestimmt: Einer maximalen Reaktion von 2,5 mm oder 4,7 mm (was zu einem Ausschluss von 4.9 % bzw. 19.4 % der Probanden führte). In der Substichprobe (der

² Die Entstehung dieser Standardwerte wurde in den von Abel publizierten Arbeiten bisher nicht transparent beschrieben. Die Auswertung der mittels AASI erhobenen Daten geschieht durch Abel Inc. und ist nicht transparent.

Probanden, die das VT durchgeführt haben), wurde kein Proband wegen zu geringer Reaktionen in der PPG ausgeschlossen. Die Autoren berichteten die Auswertung mittels z -transformierter Werte und mittels Rohwerten. Die Korrelationen zwischen PPG- und VT-Werten rangierten zwischen .09 und .27. Nur die Rohwerte der beiden Verfahren waren signifikant miteinander korreliert ($r = .27$), für die z -transformierten Werte ergab sich keine signifikante Korrelation. Die Pädophilie-Indizes der beiden Verfahren (Reaktion auf erwachsene im Vergleich zu kindlichen Kategorien) waren ebenfalls nicht signifikant miteinander korreliert. Die Autoren empfahlen deshalb bei einem Vergleich der Verfahren Roh- und z -Werte zu nutzen.

Harris et al. (1996) untersuchten 26 Personen, die sexuellen Kindesmissbrauch begangen hatten. Wie hoch der Anteil der pädophilen Täter in der Stichprobe war, teilten die Autoren nicht mit. Zudem hatten sechs Probanden auch erwachsene Opfer und zwei hatten ausschließlich Inzesttaten begangen, was die Vermutung nahelegt, dass einige Probanden eventuell nicht pädophil waren. Zehn Probanden aus der Experimentalgruppe hatten Straftaten zum Nachteil von Jungen begangen. Die Kontrollgruppe bestand aus 25 heterosexuellen Probanden aus der Allgemeinbevölkerung, welche jegliche pädophilen Interessen verneinten. Von allen Probanden durchliefen nur 30 (15 Kontrollen und 15 Täter³) die PPG- und VT-Testung. Bei der VT-Messung betrachteten die Probanden die Bilder zwei Mal: Beim ersten Mal sollten sie die Bilder nur anschauen (hier erfolgte die Erfassung der Reaktionszeiten), erst beim zweiten Durchgang gaben die Probanden ihr subjektives Rating ab. Die Stimuli bei PPG waren rein visuell. Mittels PPG konnten 27 von 30 Probanden (90%) korrekt klassifiziert werden. Die Effektstärke (d , welche sich aus der Differenz bzgl. des Pädophilie-Index zwischen den Gruppen berechnete) betrug bei der PPG 2.1 und beim VT 1.0. Die Attraktivitätsratings unterschieden sich zwischen den Gruppen nicht signifikant. Zudem waren die Korrelationen zwischen dem Rating und der VT in der Kontrollgruppe größer als in der Experimentalgruppe. Außerdem betrachteten die Täter die Bilder im Mittel generell kürzer als die Kontrollprobanden. Diese Beobachtungen legen die Vermutung nahe, dass zumindest ein Teil der Experimentalgruppe ihre pädophile Orientierung nicht offen legen wollte. Die Autoren schlussfolgerten, dass sich objektive Verfahren zur Erfassung der sexuellen Orientierung eignen und, dass die PPG bessere Effekte als das VT erbrachte.

³ Die restlichen 11 Sexualstraftäter durchliefen nach Angaben der Autoren im Rahmen der klinischen Untersuchung eine andere Art der PPG Testung und sind nicht noch einmal mittels der in der Studie genutzten PPG Form getestet worden.

Abel et al. (1998) testeten 157 Probanden, die von ihnen begangene Sexualstraftaten einräumten. Wie hoch der Anteil an Personen war, die sexuellen Kindesmissbrauch begangen haben (und wie viele davon pädophil waren), berichten die Autoren nicht. Von diesen Probanden durchliefen allerdings nur 56 eine PPG Testung, da der Erstautor im Laufe der Studie aufhörte, die PPG einzusetzen. Als Stimuli wurden auch hier Bilder genutzt. Leider sind auch in dieser Studie von Abel und Kollegen einige Vorgehensweisen nur vage beschrieben. So wird zu den VT-Daten gesagt, dass "Visual reaction time data were screened for disparate responses and these were removed prior to data analysis." Was genau unter diesen "disparate responses" zu verstehen ist, spezifizierten die Autoren nicht weiter. Die mittlere Reliabilität (interne Konsistenz) betrug für beide Verfahren 87,3 (Range bei der PPG von .66 bis .97, bei der VT von .86 bis .90). Zudem erfolgte eine Einteilung in Personen, die Kindesmissbrauch begangen haben (was als sexueller Kontakt zu Personen vor Vollendung des 18. Lebensjahres operationalisiert wurde) und andere Sexualstraftäter (z.B. Voyeurismus und Exhibitionismus). Als Prädiktoren wurden die Reaktionen in den verschiedenen Stimuluskategorien aufgenommen. Die PPG erreichte eine mittlere Klassifikationsgenauigkeit von 81.1 %, das VT von 82.2 %. Die beiden Verfahren unterschieden sich nicht signifikant hinsichtlich ihrer Klassifikationsleistung.

Letourneau (2002) untersuchte 57 wegen Sexualstraftaten inhaftierte Soldaten. 16.3 % der Probanden hatten erwachsene Opfer und 47.6 % der Täter waren mit dem Opfer verwandt. Über mögliche vorherige Opfer oder Verurteilungen lagen den Autoren keine Angaben vor. Sechs der Probanden bestritten die ihnen zur Last gelegten Taten, einer hatte bereits eine PPG und 22 bereits eine VT-Messung absolviert. 45.6 % befanden sich zum Zeitpunkt der Testung in Therapie. Eine Einteilung in pädophile und teleiophile Probanden war nicht möglich. Die Gruppeneinteilung geschah deshalb ausschließlich anhand des letzten Opfer der Probanden, sodass die objektiven Verfahren quasi genutzt wurden, um das letzte Opfer vorherzusagen. Besonders problematisch dabei ist, dass manche Probanden gleichzeitig in verschiedenen Gruppen aufgeführt wurden, wenn sie Opfer aus verschiedenen Gruppen hatten. Zudem war die Anzahl von Probanden in den verschiedenen Gruppen eher klein ($8 \leq n \leq 34$). Als Stimuli wurden bei der PPG Audiotapes genutzt, wobei das Alter der Interaktionspartner bei den kindlichen Stimuli nicht näher spezifiziert wurde („just the age you prefer“). Das verwendete VT-Paradigma war das VT des AASI (Abel et al., 1998). Vor der Auswertung der VT-Daten wurde die höchste Reaktionszeit jeder Stimuluskategorie gelöscht. Die Ergebnisse wurden für unstandardisierte und z-standardisierte VT-Werte getrennt berichtet. Die Reliabilität (interne Konsistenz) der PPG lag bei den verschiedenen

Stimuluskategorien zwischen .82 und .91 (im Mittel: .87). Beim VT lagen die entsprechenden Werte zwischen .60 und .90 (im Mittel: .82) für unstandardisierte und zwischen .56 und .93 für die z-standardisierten Werte (im Mittel: .85). Die Korrelationen der beiden Verfahren (konvergente Validität) betrug im Mittel .36 und schwankte zwischen .18 und .61 (die Korrelation zwischen den beiden Verfahren war in Bezug auf die hetero-hebephilen Stimuli nicht signifikant). Für die z-standardisierten VT-Werte ergaben sich nur in zwei Kategorien signifikante Korrelationen zwischen VT und PPG Werten: männliche und weibliche Kinder (.32 bzw. .49). Die Korrelationen in den Kategorien weibliche und männliche Jugendliche waren nicht signifikant (.18 bzw. .24). Bei der Bestimmung der Klassifikationsgenauigkeit geschah die Einteilung der Probanden, wie oben beschrieben, anhand der letzten Opfer. Die Auswertung erfolgte auf Basis absoluter (nicht relativer) Reaktionen in Bezug auf die jeweilige Kategorie. War die Reaktion eines Probanden auf eine Kategorie in der PPG z.B. größer als 10 % einer vollständigen Erektion, so wurde diesem Probanden sexuelles Interesse an dieser Kategorie attestiert. Die Vorhersage der Gruppenzugehörigkeit gelang den beiden Instrumenten nur bei der Kategorie „männliche Kinder“ signifikant besser als auf Zufallsniveau. Hier lag die Spezifität des VT bei 94 % und die Sensitivität bei 75 %. Bei der PPG lag die Spezifität bei 96 % und die Sensitivität bei 63 %. Bei allen anderen Kategorien gelang keine signifikant bessere Klassifikationsgenauigkeit als durch Raten zu erwarten wäre (mittlere Genauigkeit der beiden Verfahren lag zwischen 29.5 % und 66.0 %). Als Einschränkungen sind hinsichtlich der PPG die ungenügenden Stimuli (keine exakte Altersangabe bei den auditiven Stimuli) und die Auswertung mittels absoluter Reaktionen zu werten. Zudem ist die Güte der Gruppeneinteilung rein auf Basis des letzten bekannten Opfers fragwürdig.

Eine spezielle Stichprobe wurde von Tong (2007) untersucht. Er verglich PPG, AASI (Abel et al., 1998; VT- und Fragebogenergebnisse) und den MSI bei 53 Probanden, die in einem Sorgerechtsstreit von den Müttern beschuldigt worden waren, die gemeinsamen Kinder sexuell missbraucht zu haben. Bei allen bis auf drei Probanden kam die PPG zu dem Ergebnis, dass es sich um teleiophile Probanden handelte. Bei diesen mittels PPG negativ getesteten Probanden lag die Übereinstimmung zwischen PPG und den anderen Verfahren zwischen 98 % und 100%. Daraus schlussfolgerte der Autor, dass eine Verwendung des AASI (statt der PPG) zulässig sei. Problematisch ist allerdings, dass keiner der Probanden die ihnen vorgeworfenen Taten zugab. Wie hoch der Prozentsatz der Pädophilen in der Stichprobe war, ist somit vollkommen unklar. Sofern in der Stichprobe Pädophile waren, waren diese zudem vermutlich hoch motiviert, die Messung zu beeinflussen, um nicht als pädophil erkannt zu werden. Somit lassen sich nur Aussagen darüber

treffen, wie groß der Zusammenhang zwischen den Instrumenten bei durch die PPG als nicht-pädophil identifizierten Personen ist. Ob es einigen der Probanden in der PPG (und somit evtl. auch in anderen Instrumenten) möglich war, die Ergebnisse zu verfälschen bleibt unklar. Der Autor gab zudem an, dass bei den in der PPG positiv getesteten Probanden ($N = 3$) die anderen Verfahren alle negativ ausfielen – in dieser Gruppe lag die Übereinstimmung also bei 0 %. Eine Schlussfolgerung, welches Verfahren besser ist, lässt sich somit aus dieser Studie nicht ableiten.

Bei allen Studien, welche verschiedene objektive Verfahren verglichen haben, liegen meiner Einschätzung nach ganz erhebliche methodische Mängel vor. So wurden unübliche Auswertungsroutinen genutzt (Auswertung anhand absoluter Reaktionen), oft fehlte eine Kontrollgruppe, in vielen Studien war die Anzahl der Probanden relativ klein und teils ließ die Beschreibung der Methodik sowie der Stichprobe zu wünschen übrig. Die interne Konsistenz scheint bei beiden Verfahren ausreichend bis gut zu sein. Darüberhinaus ergibt sich ein inkonsistentes Bild: Bei Gray und Plaud (2005) erwies sich die PPG als das schlechtere der beiden Instrumente, während sie bei Harris et al. (1996) überlegen zu sein schien. Abel et al. (1998) hingegen, kamen zu dem Schluss, dass beide Verfahren gleichwertig seien. In der Studie von Letourneau (2002) erreichten beide Verfahren schließlich schlechte Klassifikationsleistungen in drei von vier Kategorien und Mackaronis et al. (2014) konnten keine signifikante Korrelation zwischen den Indizes der beiden Verfahren berichten. Die Ergebnisse der vergleichenden Arbeiten sind zur Übersicht in Tabelle 2 zusammengefasst.

Die Steigerung der Validität durch die Verwendung mehrerer Verfahren wurde bisher selten untersucht. Allerdings wiesen z.B. Banse et al. (2010) darauf hin, dass eine Kombination von direkten und indirekten Verfahren eine höhere Genauigkeit aufweist, als auf Basis nur eines Verfahrens zu erreichen ist. Sie führen dies darauf zurück, dass mittels unterschiedlicher Verfahren verschiedene Aspekte des Konstruktes der sexuellen Orientierung erfasst werden können. Dieser Gedanke lag auch der Konstruktion des EISIP zugrunde (Schmidt et al., 2014). Auch De Houwer, Teige-Mocigemba, Spruyt und Moors (2009) wiesen darauf hin, dass – aufgrund des allgemeingültigen Konvergenzprinzips (Epstein, 1983) – mehrere objektive Verfahren kombiniert werden können, um die Genauigkeit der Messung zu erhöhen.

Tabelle 2: Übersicht über bisher veröffentlichte Studien, die PPG und VT miteinander verglichen und Angaben zu Reliabilität (Cronbachs α), Kriteriumsvalidität (mittlere Genauigkeit in Prozent) und/oder konvergenter Validität (Korrelation zwischen VT / PPG) gemacht haben.

	Verfahren	Cronbachs α (im Mittel und Range)	Mittlere Genauigkeit	Korrelation VT / PPG	Besonderheiten
Gray und Plaud (2005)	VT	NA	79 %	NA	- Non-reflexive Probanden wurden mittels VT signifikant besser klassifiziert als durch PPG
	PPG	NA	64 %		- Messung sexueller Orientierung anhand absoluter Reaktionen - Keine Kontrollgruppe
Mackaronis et al. (2014)	VT	NA	NA	.27	- Jugendliche Sexualstraftäter (Alter 15-20)
	PPG				- Keine Kontrollgruppe, weshalb keine Werte für Sensitivität und Spezifität berichtet wurden - Keine signifikanten Korrelationen bei Nutzung von z-Standardisierung
Harris et al. (1996)	VT	NA	NA (d = 1)	NA	- Anteil der pädophilen Täter an den 26 Personen, die sexuellen Kindesmissbrauch begangen hatten, ist unklar
	PPG	NA	90 % (d=2.1)		
Abel et al. (1998)	VT	.87 (.86-.90)	81 %	NA	- Teils vage Ausdrucksweise der Autoren (“Visual reaction time data were screened for disparate responses and these were removed prior to data analysis”)
	PPG	.87 (.66-.97)	82 %		- Keine Kontrollgruppe
Letourneau (2002)	VT	.82 (.60-.90)	61 %	.36 (.18-.61)	- Gruppeneinteilung anhand der aktuellen Opfer
	PPG	.87 (.82-.91)	60 %		- Keine Kontrollgruppe - Messung sexueller Orientierung anhand absoluter Reaktionen

1.3 Ableitung der Fragestellung

Da einige Autoren bereits die Verwendung der PPG aufgeben und neue Verfahren anbieten (Abel et al., 1998; Abel et al., 1994), scheint eine Aussage darüber, welches Verfahren sich besser zur Messung der sexuellen Orientierung eignet notwendig. Bisherige Vergleichsstudien wiesen teilweise erhebliche methodische Mängel auf und führten zu inkonsistenten Ergebnissen. Deshalb kann aus der Literatur keine eindeutige Vorhersage abgeleitet werden, welches Verfahren sich am besten zur Messung der sexuellen Orientierung eignet. PPG gilt weiterhin als „Gold Standard“ (Blanchard et al., 2001; Freund & Blanchard, 1989), aber ein direkter Vergleich mit VT und fMRT wurde bisher nicht zufriedenstellend umgesetzt. Deshalb soll in der vorliegenden Arbeit die externe Validität und die Reliabilität (interne Konsistenz) der drei Verfahren anhand derselben Stichprobe bestimmt und miteinander verglichen werden. Auf die Berechnung der internen Konsistenz muss beim fMRT verzichtet werden, da hier keine Reaktionen auf einzelne Items berechnet werden können. Die Reaktion jedes Probanden auf die einzelnen Bilder wird im ersten Auswertungsschritt zusammengefasst, da das Signal-Rausch-Verhältnis nicht stark genug ist, um die Reaktion auf einzelne Reize zu erfassen.

Darüber hinaus sollen die Korrelationen zwischen den einzelnen Verfahren bestimmt werden, um zu ermitteln, ob die verschiedenen Tests dasselbe Konstrukt messen (konvergente Validität). Die verschiedenen Verfahren sollen alle eine Bestimmung der sexuellen Orientierung ermöglichen, basieren allerdings – wie in den einzelnen Abschnitten von 1.2 dargelegt – auf unterschiedlichen kognitiven bzw. physiologischen Prozessen. Die PPG basiert auf der genitalen Reaktion (also eine periphere Reaktion; dritte Phase in Singers (1984) Modell), während das VT mit Reaktionszeitmessungen arbeitet und am ehesten der ersten Stufe von Singers (1984) Modell zuzuordnen ist. Neuere Forschungsergebnisse legen allerdings eine rein kognitive Grundlage des VT-Effekts nahe (Imhoff, 2012). Das fMRT hingegen erfasst die zentralnervöse, präferenzspezifische Aktivierung; zu welcher Phase aus Singers (1984) Modell diese Aktivierungsmuster am ehesten passen, ist noch nicht hinreichend zu beantworten, da die Aktivierung weite Teile des Gehirns betrifft (Ponseti et al., 2012). Somit basieren die Verfahren auf unterschiedlichen Aspekten, um eine Messung der sexuellen Orientierung zu ermöglichen. Deshalb erwarte ich, dass die zur Bestimmung der sexuellen Orientierung genutzten Indizes der einzelnen Verfahren nur in geringem bis mittlerem Maße miteinander korreliert sind ($.1 \leq r \leq$

.5; Cohen, 1988). Dies lassen vorherige Studien auch vermuten (Letourneau, 2002; Mackaronis et al., 2014).

Aufgrund des Aggregationsprinzips (De Houwer et al., 2009; Epstein, 1983) erwarte ich weiterhin, dass die Kombination der unterschiedlichen Instrumente die Vorhersagegenauigkeit erhöht. Somit sollte die Kombination der Verfahren bessere Klassifikationsleistungen erbringen als die einzelnen Verfahren für sich genommen. Da sich homosexuelle Teleiophile und Pädophile in vielen Studien (VT, PPG und fMRT) stärker voneinander unterschieden als heterosexuelle Teleio- und Pädophile, gehe ich davon aus, dass die Diskriminationsleistung bei heterosexuellen Probanden (im Trend) geringer ausfallen wird als bei homosexuellen Probanden. Darüber hinaus werde ich überprüfen, ob der statistische Zusammenhang zwischen VT und den anderen Verfahren tatsächlich größer ist, wenn die Rohdaten verwendet werden, wie die Arbeit von Mackaronis et al. (2014) nahelegt.

Abschließend werde ich die einzelnen Verfahren bezüglich ihrer Stärken und Schwächen darstellen und miteinander vergleichen. Hierbei werde ich nicht ausschließlich auf psychometrische Kennwerte, sondern auch auf andere Aspekte – wie z.B. ökonomische und ethische Gesichtspunkte – eingehen. Da es schwieriger ist mittels objektiver Verfahren gleichzeitig die Alters- und Geschlechtsorientierung vorherzusagen (Freund et al., 1979) soll in der vorliegenden Arbeit nur der prädiktive Wert der drei Verfahren in Hinblick auf die Altersorientierung untersucht werden. Diese ist im Sinne von Therapie und Beratung auch die Wichtigere. Dabei soll auch nicht weiter zwischen Pädo- und Hebephilie getrennt werden; die Auswertung lehnt sich damit an die traditionelle Auswertung phallometrischer Messdaten an (Blanchard et al., 2001).

2 Methoden

2.1 Probanden

Für die vorliegende Arbeit habe ich auf die im Rahmen des Forschungsverbundes NeMUP erhobene Stichprobe zurückgegriffen. Im NeMUP Verbund wurden sowohl pädo-/hebephile Personen untersucht, die sexuellen Kindesmissbrauch begangen haben als auch solche, die bisher keinen sexuellen Kindesmissbrauch begangen hatten. Ebenso wurden Teleiophile untersucht, die selbst Kindesmissbrauch begangen hatten und solche, die keinen Kindesmissbrauch begangen hatten (Kontrollgruppe). Ich konnte nur die in Kiel erhobenen

Daten nutzen, da nur an diesem Standort die PPG durchgeführt wurden. Kontrollprobanden wurden überwiegend über Aushänge und Internetanzeigen rekrutiert. Zudem wurden lokale Homosexuellen-Stammtische kontaktiert und die Probanden gebeten, weiteren Personen von der Studie zu erzählen. Pädo-/hebephile Probanden wurden zum einen ebenfalls über Aushänge und Internetanzeigen akquiriert. Zum anderen wurden sie über verschiedene ambulante Therapieeinrichtungen (Fachambulanz Gewalt im Zentrum für integrative Psychiatrie und bei Pro Familia, Therapieprojekt „Kein Täter werden“, Beier et al., 2009) und spezielle Internetforen (<http://www.jungsforum.net>, <http://www.girlloverforum.net>) rekrutiert. Teilnehmen konnten alle pädo-/hebephilen oder teleiophilen Männer zwischen 18 und 60 Jahren. Ausschlusskriterien im Rahmen von NeMUP waren eine akute psychotische Erkrankung, aktuelle Medikation, welche die sexuelle Reaktionsfähigkeit beeinträchtigt (wie z.B. Antiandrogene, GnRH-Analoga oder auch eine Reihe von Antidepressiva), Kontraindikationen für die fMRT-Messung (wie z.B. Herzschrittmacher, bestimmte Metallimplantate o.ä.), gravierende Erkrankungen, die einen Einfluss auf die Hirnanatomie oder -funktion haben (Tumore, HIV o.ä.), aktueller Substanzmissbrauch oder -abhängigkeit sowie eine akute, schwere Depression oder Angststörungen (HAMA bzw. HAMD > 25 bzw. 15; Maier, Buller, Philipp & Heuser, 1988; Williams, 1988) oder schwere Persönlichkeitsstörungen.

Probanden wurden der Kontrollgruppe (teleiophile Probanden) zugeordnet, wenn sie in den Selbstauskünften eine sexuelle Orientierung auf das erwachsene Köperschema angaben (v.a. in Bezug auf sexuelle bzw. Masturbations-Fantasien, d.h. eine *ausschließliche* Ansprechbarkeit auf das Tannerstadium 4 und/oder 5) *und* keine sexuellen Übergriffe auf Personen unter dem 18. Lebensjahr vorlagen.

Teleiophile Probanden, die sexuellen Kindesmissbrauch begangen hatten, wurden ebenfalls in die Studie aufgenommen, wenn sie in den Selbstauskünften eine sexuelle Orientierung auf das erwachsene Köperschema angaben, älter als 25 Jahre alt waren, zusätzlich zu den sexuellen Übergriffen auf Kinder sexuelle Kontakte zu Personen angaben, die älter als 17 Jahren waren, *und* einen SSPI (Seto & Lalumiere, 2001) Wert kleiner als vier hatten.

Probanden wurden als pädo- bzw. hebephil aufgenommen, wenn sie in den Selbstauskünften eine sexuelle Orientierung auf das kindliche bzw. frühe pubertierende Köperschema berichteten (v.a. in Bezug auf sexuelle Fantasien sowie Masturbations-Fantasien; *stärkste* Ansprechbarkeit auf Tanner 1 bzw. 2 oder 3) *oder* eine solche Präferenz bestritten, aber aus der Aktenlage (bzw. eingeräumten Straftaten) eine entsprechende Orientierung hinreichend sicher erschien. Dies wurde operationalisiert durch einen SSPI Wert größer als zwei, sexuelle

Kontakte zu mindestens drei Personen, die jünger als 15 Jahre alt waren, und ein eigenes Alter über 25 Jahre. Die Altersbegrenzung sollte sicherstellen, dass es sich nicht um jugendliche Täter handelt, bei denen eine Präferenz anhand der Taten schwer nachvollziehbar sein kann.

Von den insgesamt in Kiel gemessenen 88 Probanden konnten 11 nicht in die Analysen aufgenommen werden, da auf Basis des Interviews ihre sexuelle Orientierung nicht hinreichend eindeutig war. Bei weiteren fünf Probanden lagen keine vollständige Datensätze für alle drei untersuchten Verfahren vor: Bei einem Probanden wurde vergessen, das VT durchzuführen, ein Proband passte wegen zu breiter Schultern nicht in den MRT, ein Proband verweigerte die Teilnahme an der PPG, bei einem war aufgrund eines Mikropenis die PPG nicht möglich und bei einem weiteren kam es zu technischen Problemen⁴ bei der Datenerhebung der PPG, weswegen diese Messung nicht auswertbar war.

Insgesamt konnten somit 72 Probanden in die Analysen aufgenommen werden. 38 Probanden wurden als exklusiv teleiophil, 17 als primär hebephile und 17 als primär pädoophile klassifiziert. Im Folgenden werden die pädo- und hebephilen Probanden gemeinsam betrachtet und als pädo-hebephile Gruppe bezeichnet. Drei teleiophile (teleiophile Probanden mit Kindesmissbrauch) und 13 pädo- bzw. hebephile Probanden hatten sexuellen Kindesmissbrauch begangen – bei einem Probanden ereigneten sich die Taten im Dunkelfeld, bei einem weiteren kam es trotz Anzeige zu keiner Verurteilung. Alle anderen Taten wurden rechtskräftig verurteilt. Von den als pädoophile klassifizierten Probanden bestritten drei ihre sexuelle Orientierung auf Kinder.

Die pädo-hebephilen und die teleiophilen Probanden unterschieden sich nicht signifikant hinsichtlich Alter (Teleiophile: $M = 34.63$, $SD = 10.44$; Pädo-hebephile: $M = 38.76$, $SD = 11.58$; $t(70) = -1.59$, $p = .12$; Gesamtstichprobe: $M = 36.58$, $SD = 11.11$), Intelligenz (Teleiophile: $M = 109.18$, $SD = 17.19$; Pädo-hebephile: $M = 108.88$, $SD = 18.18$; $t(69) = .07$, $p = .94$; Gesamtstichprobe: $M = 109.04$, $SD = 17.53$) oder der sexuellen Geschlechtsorientierung (aufgeteilt nach der Hauptpräferenz in hetero- und homosexuell; $\chi^2 = .02$, $p = .54$). Die Pädo-hebephilen waren im Trend allerdings häufiger linkshändig als die Teleiophilen (gemessen mit dem 15-Item Edinburgh Inventar; Teleiophile: $M = 74.70$, $SD = 34.49$; Pädo-hebephile: $M = 54.61$, $SD = 57.39$; $t(52.85) = 1.78$, $p = .08$; Gesamt: $M = 65.21$, $SD = 47.46$). Von den pädo-hebephilen Probanden waren 16 hetero- und acht homosexuell sowie zehn bisexuell. Zudem

⁴ Wir nutzten eine selbstgefertigte Steckverbindung zwischen Messschleife und Vorverstärker. Diese Verbindung wies nach einer gewissen Zeit einen Wackelkontakt auf und wurde daraufhin ersetzt. Eine Messung, die mit dieser defekten Verbindung durchgeführt wurde, konnte nicht ausgewertet werden. Bei neun weiteren (davon drei pädo-hebephile) Probanden kam es ebenfalls zu diesem Problem; diese Probanden erklärten sich aber zu einer Wiederholung der Messung bereit, sodass sie nicht aus den Analysen ausgeschlossen werden mussten.

gaben 20 eine nicht exklusive Altersorientierung (sogenannte pädo- bzw. hebephile Nebenströmung) an. Bei den teleiophilen Probanden waren 24 hetero-, 13 homo- und ein Proband bisexuell. Für eine detaillierte Aufteilung der pädo-hebphilen Probanden siehe Tabelle 3.

Die Untersuchung wurde von der Ethikkommission des Universitätsklinikums Schleswig-Holstein (UKSH) genehmigt. Alle Probanden unterzeichneten eine schriftliche Einverständniserklärung, bevor Sie in die Studie aufgenommen wurden.

Tabelle 3: Stichprobenbeschreibung der einzelnen Gruppen bezüglich Alter, IQ, sexueller Geschlechtsorientierung, Exklusivität der sexuellen Altersorientierung und Händigkeit

	Alter		IQ		Geschlechts- und Altersorientierung					Händigkeit	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	hetero-	homo-	bisexuell	exklusiv	gemischt	rechts	links
Teleio	34.63	10.44	109.18	17.19	24	13	1	38	0	36	2
Hebe	40.00	10.10	113.24	14.24	12	2	3	5	12	14	3
Pädo	37.53	13.09	104.25	21.09	4	6	7	9	8	13	4

2.2 Untersuchungsablauf und Methodik

2.2.1 Untersuchungsablauf im Projekt NeMUP

Mit jedem Probanden wurde im ersten Schritt ein kurzes Telefoninterview durchgeführt, um mögliche Ausschlusskriterien zu prüfen, die Methoden und Ziele der Untersuchung zu erläutern und Fragen zu klären. Anschließend wurden dem Probanden die ausführliche Probandeninformation (Anhang 1) und die auszufüllenden Fragebögen zugesandt, welche meist bereits ausgefüllt zum ersten Messtermin mitgebracht wurden. Die Messungen dauerten insgesamt zwischen fünf und zehn Stunden und verteilten sich - in Absprache mit den Probanden - auf zwei bis fünf Messtermine. Die Dauer der Untersuchung hing maßgeblich von den längeren Interviewzeiten bei vorliegenden psychischen Erkrankungen bzw. sexuellen Übergriffen ab. Sie verlängerte sich allerdings auch bei kognitiv weniger leistungsfähigen Probanden. Bei Ankunft im Institut fand zunächst die mündliche Aufklärung statt und die Probanden unterzeichneten die schriftliche Einverständniserklärung (siehe Anhang 2). Neben den unten detaillierter erläuterten Methoden fanden zudem folgende Messungen im Rahmen des Forschungsprojektes NeMUP statt (von denen die meisten für die vorliegende Arbeit nicht von Bedeutung sind):

- Fragebögen zu traumatischen Kindheitserlebnissen (Childhood Trauma Questionnaire; Klinitzke, Romppel, Hauser, Brahler & Glaesmer, 2012), sexueller Inhibitions- und Exzitationsbereitschaft (SIS/SES; Janssen, Vorst, Finn & Bancroft, 2002), Empathiefähigkeit (Saarbrückener Persönlichkeits-Fragebogen; Paulus, 2009), Impulsivität (BIS/BAS; Strobel, Beauducel, Debener & Brocke, 2001) und zur Aufmerksamkeits- und Konzentrationsstörung (ADHS-SB und WURS-K; Rösler, Retz-Junginger, Retz & Stieglitz, 2008). Die Probanden brauchten ca. 30 Minuten um diese Fragebögen auszufüllen.
- Ein ausführliches Interview zur Händigkeit (basierend auf dem Edinburgh Inventar; Oldfield, 1971), demographischen Angaben, Delinquenzgeschichte, Sexualität (inklusive Paraphilien und sexuellen Funktionsstörungen) sowie zu möglichen Kontraindikation für die MRT-Untersuchung. Dieses Interview wurde eigens für die NeMUP Studie entwickelt (siehe Anhang 3). Anhand dieses Interviews wurden die Probanden für die vorliegende Arbeit in die entsprechenden Gruppen eingeteilt. Die Gruppenzuordnung wurde dabei gegebenenfalls durch die Aktenlage oder (nach erfolgter, schriftlicher Schweigepflichtsentbindung) Rücksprache mit dem aktuellen Therapeuten unterstützt. Das Interview dauerte etwa 30 bis 90 Minuten.
- Das SKID I und SKID II Interview zur Erfassung psychischer Störungen (Wittchen, Zaudig & Fydrich, 1997), die Fremdbeurteilungsverfahren zur Einschätzung der Schwere von Depression (Hamilton Depression Rating Scale; Williams, 1988), Angst- (Hamilton Anxiety Scale; Maier et al., 1988) und Zwangsstörungen (Yale-Brown Obsessive Compulsive Scale; Hand & Büttner-Westühal, 1991) HAMA, HAMD und sowie ein Interview zum Alkohol- und Substanzkonsum im Lebenslängsschnitt. Die Dauer dieser Interviewteile schwankte zwischen 30 Minuten und 3 Stunden.
- Vier Untertests des Wechsler Intelligenztets für Erwachsene (WAIS III; Molz, Schulze, Schroeders & Wilhelm, 2010) zur Beurteilung der allgemeinen kognitiven Leistungsfähigkeit: Wortschatztest, Gemeinsamkeiten finden, Matrizentest und Mosaiktest (Dauer ca. 30 Minuten).
- Folgende Aufgaben aus der computerbasierten, neuropsychologischen Testbatterie CANTAB (Cambridge Neuropsychological Testing Automated Battery; CANTABeclipse®): motorische Steuerung (MOT), Stop-Signal-Task (SST), Intra/Extradimensional Shift (IED), Information sampling task (IST), spatial working memory (SWM), stockings of cambridge (SOC). Die Durchführung dieser Untertests dauerte ca. 90 Minuten.

- Eine computerbasierte Erfassung der Empathiefähigkeit, welche etwa 30 Minuten dauerte und eigens für NeMUP entwickelt wurde.
- Eine Messung des Längenverhältnisses von Ring- und Zeigefinger (siehe z.B. Mularczyk, Zietek-Czeszak & Zietek, 2014) sowie eine Blutentnahme (Dauer ca. 10 Minuten).
- Das VT-Verfahren (siehe 2.2.2), welches ca. 5 Minuten dauerte.
- Die PPG (siehe 2.2.3), welche ca. 60 Minuten dauerte.
- Zwei Messtermine im MRT (siehe 2.2.4.), welche jeweils ca. 60 Minuten dauerten.

Alle computerbasierten Tests (MET, VT, CANTAB) wurden an demselben Tablet durchgeführt, das bei CANTAB mitgeliefert wurde. Die einzelnen Methoden wurden in keiner standardisierten Reihenfolge erhoben und v.a. an die aktuellen zeitlichen Gegebenheiten angepasst. Nach Abschluss aller Untersuchungsteile bekamen die Teilnehmer die Aufwandsentschädigung ausgezahlt.

Neurologen inspizierten die am MRT erhobenen Daten auf Zufallsbefunde. Hierzu wurde ein anatomisches, T2-gewichtetes Bild aufgenommen (siehe 2.2.4.2). Bei einer Rückmeldung seitens der Neurologen wurde dies den Probanden am zweiten Messtermin mitgeteilt und eine weitere Abklärung nahegelegt. Dies betraf insgesamt vier Probanden.

2.2.2 Untersuchungsablauf und Methodik VT

2.2.2.1 Stimuli und Methodik beim VT

Als Stimuli nutzten wir insgesamt 80 Bilder in 10 verschiedenen Kategorien: Männliche und weibliche Personen jeweils in den Tannerstadien 1-5 (acht Bilder pro Kategorie). Die Bilder stammten aus dem „Not Real People Set“ (NRP; Mokros, Butz, Dombert, Santtila, Bäuml, et al., 2011) und zeigten am Computer generierte Personen in Badebekleidung. Der Hintergrund war in einem Grauton gehalten. Das NRP wurde speziell für die indirekten Verfahren zur Messung der sexuellen Orientierung erstellt. Die Bildpräsentation erfolgte über ein Tablet mit einem 12 Zoll großen Display. Das Experiment wurde in Presentation © (<https://www.neurobs.com/>) programmiert.

2.2.2.2 Untersuchungsablauf VT

Die Probanden waren aufgefordert, zu jedem Bild anzugeben, wie sexuell erregend sie dies fanden (für die genauen Instruktionen siehe Anhang 4). Dazu stand ihnen eine 4-Punkt Likert-Skala zur Verfügung (von „gar nicht erregend“ bis „sehr erregend“). Die Antwort erfolgte mit der rechten Hand über die Tasten „4“ bis „7“ auf einer an dem Tablet

angeschlossenen Computertastatur. Die Bilder wurden für fünf Sekunden präsentiert bzw. bis eine Antwort erfolgte – je nachdem was zuerst eintrat. Erfolgte innerhalb der fünf Sekunden keine Reaktion, wurde das nächste Bild präsentiert und der Durchgang als fehlend gewertet. Die Zeit – von Präsentation des Bildes bis zur Eingabe der Antwort – wurde erfasst, ohne dass die Probanden im Vorhinein darüber informiert worden waren. Die Bildpräsentation erfolgte in randomisierter Reihenfolge mit einem variablen, mittleren Interstimulusintervall von 1992 ms (Spannweite: 1492 bis 2492 ms).

2.2.2.3 Auswertung VT

Die Auswertung der VT-Daten erfolgte in SPSS 20. Zunächst habe ich die Reaktionszeiten pro Kategorie gemittelt und anschließend zwei Altersindizes berechnet: Einen für die weiblichen und einen für die männlichen Stimuli. Dafür habe ich von der mittleren Reaktionszeit in Bezug auf die erwachsenen Stimuli (Tanner 5) die mittlere Reaktionszeit in Bezug auf die kindlichen Stimuli (Tanner 1-3) abgezogen.

Im NeMUP Verbund wurde eine andere Art der Datenauswertung genutzt. In Anlehnung an Bourke und Gormley (2012) wurde besonders lange und kurze Reaktionszeiten (Ausreißer) vor der weiteren Analyse ausgeschlossen. Dies betraf alle Werte die mehr als drei Interquartilsabstände von den Interquartilsgrenzen (nach oben und nach unten) entfernt waren. Anschließend fand eine z -Transformation der Daten statt. Um die Ergebnisse auf Grundlage beider Datenarten (Roh- und z -transformierte Daten) zu prüfen, führte ich ebenfalls diese Berechnungen durch und bildete schließlich auf Grundlage dieser Daten dieselben Indizes wie auf Grundlage der Rohdaten. Dabei musste ich einen Probanden von den weiteren Analysen ausschließen, da dieser Proband durch den Ausschluss von Ausreißern bei den männlichen, erwachsenen Reizen (Tanner 5) alle Reaktionszeiten wegfielen und somit kein Index für das männliche Geschlecht gebildet werden konnte. Anschließend habe ich Diskriminanzanalysen und ROC gerechnet.

2.2.3 Untersuchungsablauf und Methodik PPG

2.2.3.1 Stimuli und Methodik der PPG

Für die Stimulation haben wir eine audiovisuelle Präsentation gewählt. Bei den auditiven Stimuli handelte es sich um die Übersetzungen von etablierten Stimuli (Freund & Blanchard, 1989), die ich im Rahmen meiner Diplomarbeit angefertigt habe (Pohl, 2009). Die Geschichten erzählen eine sexuelle Interaktion zwischen einer Hauptperson – die in der 2.

Person Singular angesprochen wird – und einer männlichen oder weiblichen Person in verschiedenen Altersstufen. Wir haben keine Stimuli verwendet, die Gewalt oder Demütigung beinhalteten. Es gab insgesamt je fünf Kurzgeschichten, die eine Interaktion mit männlichen und weiblichen Erwachsenen (ohne Altersangabe in den Stimuli) und Kindern (Alter 5-9 Jahre) beschrieben. Für die intrapuberalen Jugendlichen (Alter 11-13 Jahre) gab es nur vier Kurzgeschichten. Zudem wurde noch eine neutrale Kurzgeschichte insgesamt zweimal dargeboten, sodass insgesamt 30 auditive Stimuli präsentiert wurden. Die heterosexuellen Stimuli wurden von einer weiblichen Sprecherin, die homosexuellen von einem männlichen Sprecher gesprochen. Die neutrale Geschichte wurde je einmal mit der männlichen und einmal mit der weiblichen Stimme präsentiert.

Die visuellen Stimuli zeigten frontale Ganzkörper- bzw. Genitalaufnahmen von vorpubertären Kindern (Tanner 1), intrapuberalen Jugendlichen (Tanner 2-3) und Erwachsenen (Tanner 5). Es handelte sich bei den Kindern um nicht-pornographische Abbildungen. Quellen waren z.B. im Internet frei zugängliche Nudistenwebseiten. Die Bilder mit erwachsenen Modellen waren pornographische Bilder, die bereits in vorherigen Studien genutzt wurden sind (z.B. Ponseti et al., 2012; Ponseti et al., 2009). Auch die meisten kindlichen Bilder (Tanner 1) konnten aus vorherigen Studien übernommen werden. Die Bilder der Tannerstadien 2-3 wurden eigens für NeMUP erstellt. Um ein valides Stimulusset zu gewährleisten, haben fünf Mitarbeiter der Sexualmedizin die auf den Bildern gezeigten Personen bezüglich ihres Tannerstadiums eingeschätzt. Es wurden die Bilder als Stimuli für die entsprechende Kategorie ausgewählt, bei denen alle Mitarbeiter übereinstimmend dasselbe Tannerstadium angegeben haben. Da hierdurch einige Bilder der Tannerkategorie 1 entfernt wurden, mussten diese durch neue Stimuli ersetzt werden. Bei allen Bildern wurde der Hintergrund durch einen Grauton ersetzt, um Kontextinformationen zu entfernen.

Jede Kurzgeschichte (Dauer: 46 bis 68 Sekunden; im Mittel 55 Sekunden) wurde gepaart mit insgesamt drei Bildern dargeboten. Zunächst wurde für 23 Sekunden eine Ganzkörperaufnahme einer hinsichtlich Alter und Geschlecht passenden Person präsentiert. Anschließend wurde für 23 Sekunden eine Genitalaufnahme gezeigt, woraufhin noch einmal das zuvor bereits präsentierte Ganzkörperbild für 23 Sekunden gezeigt wurde. Während der gesamten neutralen Geschichten wurde nur eine Landschaftsaufnahme präsentiert. Die Stimuluspräsentation erfolgte blockweise in einer fixen Abfolge.

Zur Datenaufzeichnung nutzten wir Messschlaufen vom Umfang 80 mm von DM Davis, Inc. (D. M. Davis HgPc™ Penis Gage; www.dmdavis.com). Die Schlaufen waren mittels eines von Brainproducts angefertigten Vorverstärkers mit einem V-Amp 16 Kanal EEG

System über den Aux1-Kanal verbunden (<http://www.brainproducts.com/>). Der Output erfolgte in μV (Abtastrate 1000-mal pro Sekunde). Die Aufzeichnung des Signals erfolgte über den Vision Recorder von Brainproducts (Version 1.20.0401) auf einem Laptop (Windows 7). Die Stimuluspräsentation wurde über Presentation (<https://www.neurobs.com/>) an einem zweiten Rechner (Windows XP) gesteuert. Die beiden Rechner waren über den COM- bzw. USB-Port miteinander verbunden. Die Präsentation der auditiven Stimuli erfolgte über schnurlose Kopfhörer von Philips (Philips Wireless FM SHC8535); die Präsentation der Bilder erfolgte über einen 19 Zoll großen Monitor, der sich ca. $\frac{1}{2}$ Meter von dem Sessel entfernt befand, auf dem der Probanden während der Messung saß. Für die Aufgabe während der Messung stand dem Probanden eine Computermaus zur Verfügung, die er nach Belieben in der rechten oder linken Hand halten konnte.

2.2.3.2 Untersuchungsablauf PPG

Vor jeder Messung haben wir die Messschleife kalibriert, indem wir sie über einen Zylinder mit verschiedenen Umfängen gelegt und den μV Output aufgezeichnet haben. Dabei haben wir die Umfänge 90 - 130 mm in fünf mm Schritten gemessen. Anschließend wurde mittels SPSS eine lineare Regression von Umfang auf μV Output berechnet. Einerseits wurden so die korrekte Funktionsweise der Schleifen sowie die Linearität der Messwerte sichergestellt. Andererseits konnten wir die Regressionsgleichung im Anschluss an die Messung zur Bestimmung der Umfangswerte während der Messung nutzen.

Während der Messung saß der Proband alleine in einem halb-abgedunkelten Raum, während der Versuchsleiter sich in einem angrenzenden Vorraum befand. Die Kommunikation war durch die geschlossene Tür hindurch möglich. Im Hauptteil der PPG wurde eine Abwandlung der STT (Harris et al., 1999; Quinsey & Chaplin, 1988) genutzt, um die Aufmerksamkeit der Probanden sicher zu stellen und einer möglichen Verfälschungstendenz vorzubeugen. Die Probanden wurden gebeten, wann immer sie das Wort „Penis“ oder „Küssen“ hörten, die linke Maustaste zu drücken. Sie sollten ebenfalls die linke Maustaste drücken, sobald ein grauer Punkt auf dem Bildschirm zu sehen war. Ein entsprechender Punkt erschien bei jedem sexuellen Bild einmal. So sollte die Aufmerksamkeit für die auditiven und visuellen Reize sichergestellt werden. Um die relativ komplexe Aufgabe möglichst einfach zu halten, wurde auf die Verwendung verschiedener Tasten für die unterschiedlichen Reize verzichtet. In einem Probedurchgang (Instruktionen siehe Anhang 5) lernte der Proband zunächst die STT kennen. Hierzu wurde eine neutrale Geschichte präsentiert, die im Hauptdurchgang nicht vorkam. Anschließend hat der Versuchsleiter den Proband instruiert, wie er die Messschleife

anzulegen hat und es konnten letzte Fragen geklärt werden. Zudem sollte der Proband während der Messung die Unterhose bis zu den Knien herunterziehen, konnte sich aber mit einem Bettlaken zudecken, sofern er dies wollte. Sodann verließ der Versuchsleiter den Raum. Während der Proband die Schlaufe anlegte, kontrollierte der Versuchsleiter das Signal im Vision Recorder. Falls es zu Störungen im Signal kam, wurde der Proband gebeten, die Schlaufe noch einmal abzunehmen und erneut aufzusetzen. Sobald die einwandfreie Funktionsweise der Geräte sichergestellt war, startete der Versuchsleiter den Hauptdurchgang. Nachdem der Proband die Instruktionen gelesen hatte (siehe Anhang 6), initiierte dieser mit einem Klick die eigentliche Messung. Bei dem ersten und letzten Stimulus der Messung handelte es sich um eine neutrale Geschichte (gepaart mit einer Landschaftsaufnahme). Nach der ersten neutralen Geschichte erfolgte zunächst die Präsentation von je drei Geschichten in folgender Reihenfolge: heteropädophil⁵, homoteleiophil, homopädophil, heteroteleiophil. Anschließend erfolgte die Präsentation der beiden hebephilen Kategorien; zunächst die der Jungen, dann die der Mädchen. Abschließend wurden die letzten beiden Geschichten der ersten vier Kategorien in umgekehrter Reihenfolge präsentiert (heteroteleiophil, homopädophil, homoteleiophil, heteropädophil), gefolgt von der zweiten neutralen Geschichte. Zwischen den einzelnen Kategorien befand sich ein 30-sekündiges Interstimulusintervall, währenddessen der Proband einen schwarzen Bildschirm mit der Information „Pause – es geht gleich automatisch weiter“ sah. Die gesamte Stimuluspräsentation dauerte insgesamt (ohne Instruktion und Probedurchgang) 39 Minuten.

Nach Abschluss der Untersuchung legte der Proband die Schlaufe selbstständig ab und verließ den Raum. Die Messschlaufen wurden nach jeder Messung mittels Cidex[®] Opa (Johnson & Johnson Medical GmbH) desinfiziert und für mehrere Messungen genutzt.

2.2.3.3 Auswertung PPG

Die mittels Vision Recorder gespeicherten Daten habe ich zunächst im Vision Analyzer (Version 2.0.2.5859; <http://www.brainproducts.com/>) geöffnet und über ein Script (Solutions, Peak_Export) ausgelesen. Dabei wurde der μV Output pro präsentiertem Bild gemittelt und anschließend in eine Textdatei ausgeschrieben. Diese Daten öffnete ich dann in SPSS 20, um sie dort weiter auszuwerten. Zuerst löschte ich die Reaktionen auf die ersten drei Bilder pro Kategorie (bzw. die ersten 46 Sekunden bei den neutralen Stimuli), um Überhangseffekte zu reduzieren. Anschließend transformierte ich die Roh-Werte mittels der in der Kalibrierung

⁵ Eine „heteropädophile“ Geschichte soll eine Geschichte bezeichnen, in der eine sexuelle Interaktion mit einem jungen Mädchen geschildert wird.

errechneten Regressionsgleichung in Umfangswerte, um anhand dieser Daten, Low-Responder auszuschließen (Howes, 2003). Hierzu bildete ich einen Index, der ein Maß für die Reaktionsstärke während der gesamten Messung darstellt (*neutral_deviation*), indem ich von der mittleren Reaktion auf die neutralen Stimuli den höchsten Wert in Reaktion auf eines der sexuellen Bilder abzog. Die weitere Auswertung der PPG Daten habe ich anhand der Roh-daten vorgenommen, um Fehlervarianz, die durch die Umrechnung in Umfangswerte entstanden sein könnte, auszuschließen.

Da eine Transformation der Rohdaten erforderlich (Blanchard et al., 2001), die Verwendung von *z*-Transformationen aber nicht unumstritten ist, subtrahierte ich zunächst von den μ V Werten in Reaktion auf die jeweiligen sexuellen Kategorien (Kinder, Jugendliche und Erwachsene beider Geschlechter) die mittleren μ V Werte in Reaktion auf die neutralen Reize. Somit ergaben sich sechs Indizes: Zwei erwachsene, zwei kindliche und zwei für die jugendlichen Reize. Anschließend habe ich wie beim VT zwei Altersindizes erstellt: Einen für das weibliche und einen für das männliche Geschlecht. Hierzu subtrahierte ich von dem jeweiligen Erwachsenenindex den päd- oder hebephilen-Index – je nachdem, welcher größer war. Da die Mittelung des weiblichen und männlichen Altersindex zu einem Gesamtindex zu besseren Klassifikationsergebnissen führte, habe ich für alle weiteren Analysen den aus beiden Altersindizes gemittelten Index verwendet. Anschließend habe ich Diskriminanzanalysen und die AUC berechnet.

2.2.4 Untersuchungsablauf und Methodik fMRT

2.2.4.1 Stimuli und Methodik beim fMRT

Als Stimuli verwendeten wir Gesichts- und Ganzkörperabbildungen. Wir nutzten insgesamt sechs Kategorien: Männliche und weibliche Personen in den Tannerstadien 1 (pädophile Stimuli), 2 - 3 (hebephile Stimuli) und 5 (teleiophile Stimuli). Pro Kategorie wählten wir 60 Bilder aus (jeweils 30 Ganzkörper- und 30 Gesichtsaufnahmen), was zu einer Summe von insgesamt 360 Stimuli führte. Die Auswahl der Ganzkörperstimuli erfolgte dabei nach dem gleichen Procedere wie in der PPG (die Ganzkörperstimuli bei der PPG waren eine Teilmenge der Bilder aus dem fMRT-Paradigma). Ich habe die Gesichtsbilder aus frei zugänglichen Bildern aus dem Internet ausgewählt. Die Erwachsenenbilder wurden so gewählt, dass klar ersichtlich war, dass es sich um Personen jenseits der Pubertät handelt. Sie wurden bereits in vorherigen Studien verwendet (Ponseti et al., 2012). Da eine Einschätzung des Tannerstadiums, wie es bei den Ganzkörperbildern erfolgt war, bei den Gesichtsbildern nicht möglich war, haben

wir das chronologische Alter von Kindern und Jugendlichen auf 216 vorausgewählten Bildern durch 17 externe Beurteiler (zehn Männer, sieben Frauen, mittleres Alter: 36.5 Jahre) einschätzen lassen. Bei dem Rating wurden die Bilder in einer pseudorandomisierten Reihenfolge mittels Presentation© auf einem Laptop präsentiert. Die Probanden sollten das Alter (zwischen 1 und 20 Jahren) möglichst genau einschätzen. Dazu standen Ihnen 20 entsprechend beschriftete Tasten zur Verfügung. Wir wählten die Bilder aus, die von den Beurteilern im Mittel am jüngsten (präpubertäre Stimuli; maximales, eingeschätztes Alter lag dabei unter 10.5 Jahren) bzw. zwischen 11 und 14 Jahren (intrapubertäre Stimuli) eingeschätzt wurden.

Wir stellten in der Analyse im Rahmen des NeMUP Projektes fest, dass sich – im Gegensatz zur vorherigen Studie (Ponseti et al., 2015) – die Gesichtsbilder nur schlecht zur Klassifikation eigneten⁶. Auch die Bilder von intrapuberalen Jugendlichen wirkten sich nicht positiv auf die Klassifikationsleistung aus. Deshalb wurden zur Klassifikation der Probanden in der vorliegenden Studie nur die Ganzkörperbilder der Kinder und der Erwachsenen (beider Geschlechter) genutzt, sodass ich im Folgenden nur auf diese Reizklassen Bezug nehmen werde.

Die Stimulus-Präsentation wurde im fMRT über Presentation© gesteuert. Die Bilder wurden dem Probanden auf eine Leinwand, die sich hinter dem Scanner befand, projiziert. Diese konnte der Proband über einen Spiegel sehen, der sich auf der Kopfspule befand. Die Synchronisation zwischen Scanner und Presentation© erfolgte nach der Aufzeichnung der zwei Dummy-Scans. Die MRT-Messungen erfolgten in Kiel an einem 3T Philips Achieva Scanner. Das hochauflösende, strukturelle T1 gewichtete Bild wurde mit einer Echozeit (TE) von 3.75 ms und einer Wiederholungszeit (TR; repetition time) von 8.14 ms aufgenommen. Es wurden 192 Schichten ohne Distanzfaktor aufgenommen. Der Kippwinkel betrug 7°, das Messfeld (FOV; field of view) 256x256 mm und deckte den gesamten Kopf ab. Die Voxelgröße betrug 1x1x1mm. Die Messung dauerte ca. 5.5 Min.

Das funktionelle Klassifikationsparadigma erfolgte in zwei Sessions à ca. 10 Min. und wurde mit einer echo planar imaging (EPI) Sequenz mit folgenden Einstellungen aufgenommen: TR 2400 ms, TE 30 ms, FOV 240 mm, 38 Schichten (Schichtdicke: 3 mm; Distanzfaktor 10%), Kippwinkel 80°. Die Messreihenfolge war in beiden Durchgängen aufsteigend (ascending).

⁶ Diese Abweichung liegt vermutlich u.a. an dem im Vergleich zur Vorstudie verkürzten ISI. Auf die genauen Ursachen der Abweichungen soll hier aber nicht weiter eingegangen werden, da dies Schwerpunkt weiterer wissenschaftlicher Arbeiten sein soll.

2.2.4.2 Untersuchungsablauf fMRT

Die erste fMRT-Messung zur Erfassung der sexuellen Orientierung fand während des ersten MRT-Termins statt. Der Versuchsleiter fixierte den Kopf des Probanden mit Hilfe von Schaumstoffpolstern in der Kopfspule, um Bewegungsartefakte zu minimieren. Um eine bequemere Liegeposition zu ermöglichen, platzierte der Versuchsleiter anschließend eines von zwei Beinkeilen unter den Waden des Probanden. Sodann befestigte der Versuchsleiter an der linken Hosentasche des Probanden eine, während dieser in seiner rechten Hand eine Computermaus hielt, mit der er die verschiedenen Aufgaben bearbeiten konnte. Anschließend wurden dem Probanden noch Kopfhörer aufgesetzt, worüber die Kommunikation mit dem Versuchsleiter möglich war. Danach wurde der obere Teil der Kopfspule inklusive Spiegel aufgesetzt und der Proband in den Scanner eingefahren.

Anschließend wurden folgende MRT-Sequenzen durchgeführt: ein Localizer (Scout-Messung zur Planung der folgenden Sequenzen; ca. 1 Min.), eine Resting-State-Messung (ca. 12 Minuten; siehe z.B. Cordes, Haughton, Carew, Arfanakis & Maravilla, 2002), ein Go-/No-go-Paradigma (ca. 9,5 Minuten; siehe z.B. Brown et al., 2015), der erste fMRT-Durchgang zur Messung der sexuellen Orientierung (siehe 2.2.4.1), das hoch-auflösende strukturelle T1-gewichtete Bild (siehe 2.2.4.1), eine B0-Feldkarte (ca. 1 Minute), ein strukturelles T2-gewichtetes Bild (welches zur Befundung genutzt wurde; ca. 1 Minute), der zweite fMRT-Durchgang zur Messung der sexuellen Orientierung (siehe 2.2.4.1) sowie eine diffusionsgewichtete Sequenz (DTI, ca. 16,5 Minuten; siehe z.B. Qiu et al., 2012) zur Darstellung der weißen Substanz. Bei 25 Probanden konnte die DTI Sequenz aus zeitlichen Gründen nicht am ersten Messtermin erfolgen und wurde am zweiten MRT-Termin nachgeholt. Die gesamte Messsequenz dauerte ca. 75 Minuten (60 Minuten ohne DTI).

Zwischen den einzelnen Sequenzen wurde dem Probanden kurz der weitere Ablauf geschildert. Zudem konnte so in dringenden Fällen die Messung unterbrochen werden. Sofern der Proband zwischendurch aufstehen musste, was bei zwei Probanden vorkam, wurde anschließend ein neuer Localizer gemessen, um die nachfolgenden Messungen ggfs. neu zu planen. Unmittelbar vor der ersten fMRT-Messung zur Erfassung der sexuellen Orientierung wurde dem Probanden zudem die Aufgabe anhand eines ca. 1-minütigen Probedurchgangs verdeutlicht. Während der zwei 10-minütigen fMRT-Sessions wurde jeweils die Hälfte der Stimuli (180 Bilder) in einer pseudorandomisierten Reihenfolge dargeboten. Die Bilder wurden für eine Sekunde präsentiert. Das variable Interstimulusintervall (ISI) betrug im Mittel zwei Sekunden (Spannweite: 1.5 bis 2.5 Sekunden). Zusätzlich zu den Stimuli wurden pro Session

10 Nullevents (Zeiten, in denen die Probanden dasselbe sahen wie beim ISI) präsentiert. Zudem wurden den Probanden pro Durchgang zehn Bilder von Schaufensterpuppen gezeigt. Bei diesen Bildern sollte der Proband möglichst schnell die linke Maustaste drücken. Diese Aufgabe diente als Aufmerksamkeitskontrolle.

Am zweiten MRT-Termin (der mindestens eine Woche nach dem ersten MRT stattfand), wurden die 2 Sessions zur Erfassung der sexuellen Orientierung wiederholt. Zudem fand eine ca. 5-minütige Präsentation von erwachsenen und kindlichen Tieren statt. Die wiederholte Messung der sexuellen Orientierung diente dem Ziel der Erfassung der Retest-Reliabilität und der Verfälschbarkeit der Messergebnisse durch die Probanden, was allerdings nicht Gegenstand der vorliegenden Arbeit ist.

2.2.4.3 Auswertung fMRT

Die Datenvorverarbeitung (preprocessing) haben wir mit SPM8 (<http://www.fil.ion.ucl.ac.uk/spm/>) unter Matlab (R2012a, Version 7.14.0.739; MathWorks, Natick, Massachusetts, <http://de.mathworks.com/>) für jede Session in der folgenden Reihenfolge durchgeführt (in Klammern sind die Befehle und Parameter aus SMP aufgeführt): Zur Bewegungskorrektur haben wir die EPI-Bilder zunächst am mittleren aufgenommenen EPI-Bild der jeweiligen Session orientiert (realign: estimate&reslice; estimate: 2nd degree b-spline, reslice: 4th degree b-spline). Anschließend haben wir das T1-Bild in graue und weiße Substanz sowie Cerebrospinalflüssigkeit segmentiert (segment mit den tissue probability maps von SPM) und die neuorientierten EPI-Bilder mit dem segmentierten T1-Bild koregistriert (coregister: estimate). Anschließend haben wir die Normalisierungsparameter der Segmentierung genutzt, um die koregistrierten EPI-Bilder im MNI Raum (Chau & McIntosh, 2005) zu normalisieren (normalise: write; Voxel size 2x2x2mm, trilinear interpolation, no wrap). Diese Bilder wurden dann räumlich geglättet (smooth; FWHM 8x8x8mm, same data type).

Die statistische Datenanalyse und die Klassifikation der Probanden erfolgte in Anlehnung an die Arbeit von Ponseti et al. (2012). Im *ersten* Schritt (first-level analysis) wurde pro Proband ein lineares Modell spezifiziert. Dabei stellte jede der vier Stimulusgruppen (männliche und weibliche Körper der Tannerstadien 1 und 5) eine einzelne Bedingung dar. Zudem wurden die 6 Bewegungsparameter (aus dem Schritt realign) als Regressoren in das Modell aufgenommen, um mögliche Störeinflüsse der Bewegung abzufangen. Die stimulusspezifischen Reaktionen wurden mit der hämodynamischen Antwortfunktion gefaltet. Wir nutzten einen Hochpassfilter von 128 Sekunden, um langsame Änderungen im BOLD

Signal zu eliminieren. Anschließend bildeten wir für jedes Geschlecht einen t -Kontrast: (1) einen Kontrast, der die Reaktion auf Männer- mit der auf Jungenbilder vergleicht (Männer > Jungen) und (2) einen Kontrast, der die Reaktion auf Frauen- mit der auf Mädchenbilder vergleicht (Frauen > Mädchen). Für jeden Probanden gab SPM für beide t -Kontraste jeweils ein Kontrastbild (con imagae) aus. Diese Bilder enthalten Informationen über die unterschiedliche Aktivierung zwischen den beiden Bedingungen (erwachsene und kindliche Reize).

Im *zweiten* Schritt (second-level analysis) wurden die Gruppenkarten erstellt. Hierzu habe ich ein flexibles Design (flexibel factorial design) mit drei Faktoren genutzt. Der erste Faktor (Bedingung) bestand aus den Reaktionen der Probanden auf die einzelnen Stimulusarten und hatte 4 Abstufungen (Mädchen, Frauen, Jungen und Männer). Der zweite Faktor (Gruppe) hatte die Ausprägung teleiophil oder pädo-/hebephil. Als dritten Faktor (Störvariable) definierte ich einen Probandenfaktor (mit 72 Abstufungen), um weitere Störvarianz (z.B. aufgrund von Alter oder IQ) abzufangen. Zudem habe ich die Händigkeit der Probanden (in Form der Summe über die 15 Items des Händigkeitsfragebogens) als Kovariate ins Modell mit aufgenommen, da sich diese marginal signifikant zwischen den Gruppen unterschied. Ich berechnete dann über einen t -Kontrast zwei Interaktionsterme (einen für die männlichen und einen für die weiblichen Stimuli). Diese erstellten zwei statistische Gruppenkarten zur präferenzspezifischen Aktivierung (Frauen > Mädchen und Männer > Jungs). In diesen Gruppenkarten sind die Regionen dargestellt, bei denen die teleiophilen Probanden – im Vergleich zu den pädophilen Probanden – eine stärkere Aktivierung bei erwachsenen (im Vergleich zu kindlichen) Reizen aufwiesen. Es handelt sich also um eine zweifache Interaktion. Positive t -Werte liegen in den Regionen vor, in denen die Teleiophilen eine stärkere Aktivierung bei dem entsprechenden Kontrast (Frauen > Mädchen bzw. Männer > Jungs) zeigten als die Pädophilen. Negative t -Werte hingegen bedeuten, dass die Pädophilen in diesen Regionen bei dem entsprechenden Kontrast eine stärkere Aktivierung aufwiesen als die Teleiophilen.

Im *dritten* (t -Karten Projektion) Schritt wurde die Passung der einzelnen Probanden in Bezug auf die im Schritt zwei erstellten Gruppenkarten errechnet. Hierzu wurde eine Toolbox für SPM8 genutzt, die Oliver Granert im Rahmen vorheriger Arbeiten programmierte (Ponseti et al., 2012; Ponseti et al., 2009). Sie ermöglicht, die probandenweise Multiplikation der Kontrastbilder aus dem first-level (Schritt eins; Männer > Jungs bzw. Frauen > Mädchen) mit den t -Werten der Gruppenkarte (aus Schritt zwei). Diese Berechnung wird für jeden Voxel einzeln durchgeführt und die Produkte anschließend aufsummiert. Die hieraus resultierende Summe ist der sogenannte Expression Value. Dabei wurden nur die Voxel in die Berechnung

mit einbezogen, die in der Gruppenkarte einen t -Wert von mindestens vier aufwiesen, da dies in vorherigen Studien die besten Klassifikationsergebnisse erbrachte (Ponseti et al., 2012). Dieser Schritt wurde für beide Geschlechter einzeln durchgeführt, sodass jeder Proband einen Expression Value für die weiblichen und einen für die männlichen Kategorien hatte. Negative Werte sprechen für eine pädophile, positive für eine teleiophile Präferenz. Umso näher die Werte bei null liegen, umso untypischer ist das Aktivierungsmuster des Probanden im Vergleich zur Gruppenkarte.

Im *vierten* Schritt (Klassifikation mit Kreuzvalidierung) habe ich die Probanden mittels Diskriminanzanalyse (nach Fischer) anhand ihrer zwei Expression Values klassifizieren lassen. Um eine Überschätzung der Modellgüte und somit der Klassifikationsgenauigkeit zu vermeiden, habe ich eine sogenannte Leave-One-Out-Kreuzvalidierung genutzt. Hierbei wurde jeweils ein Proband aus der Stichprobe entfernt und die Berechnung der Gruppenkarte (Schritt zwei) ohne diesen Probanden durchgeführt. Anschließend wurde die t -Karten Projektion (Schritt drei) mit den first-level Kontrasten des ausgelassenen Probanden und der Gruppenkarte berechnet. Der Proband wurde dann aus den daraus resultierenden Expression Values klassifiziert. Dies wurde für jeden Probanden einzeln wiederholt (also insgesamt 72-mal). Als Ergebnis der Klassifikation erhielt jeder Proband einen Klassifikationswert, aus dem sich die Gruppenzugehörigkeit ableiten lässt. Hierbei sprechen positive Werte allerdings für eine Pädophilie, während negative für eine Teleiophilie sprechen.

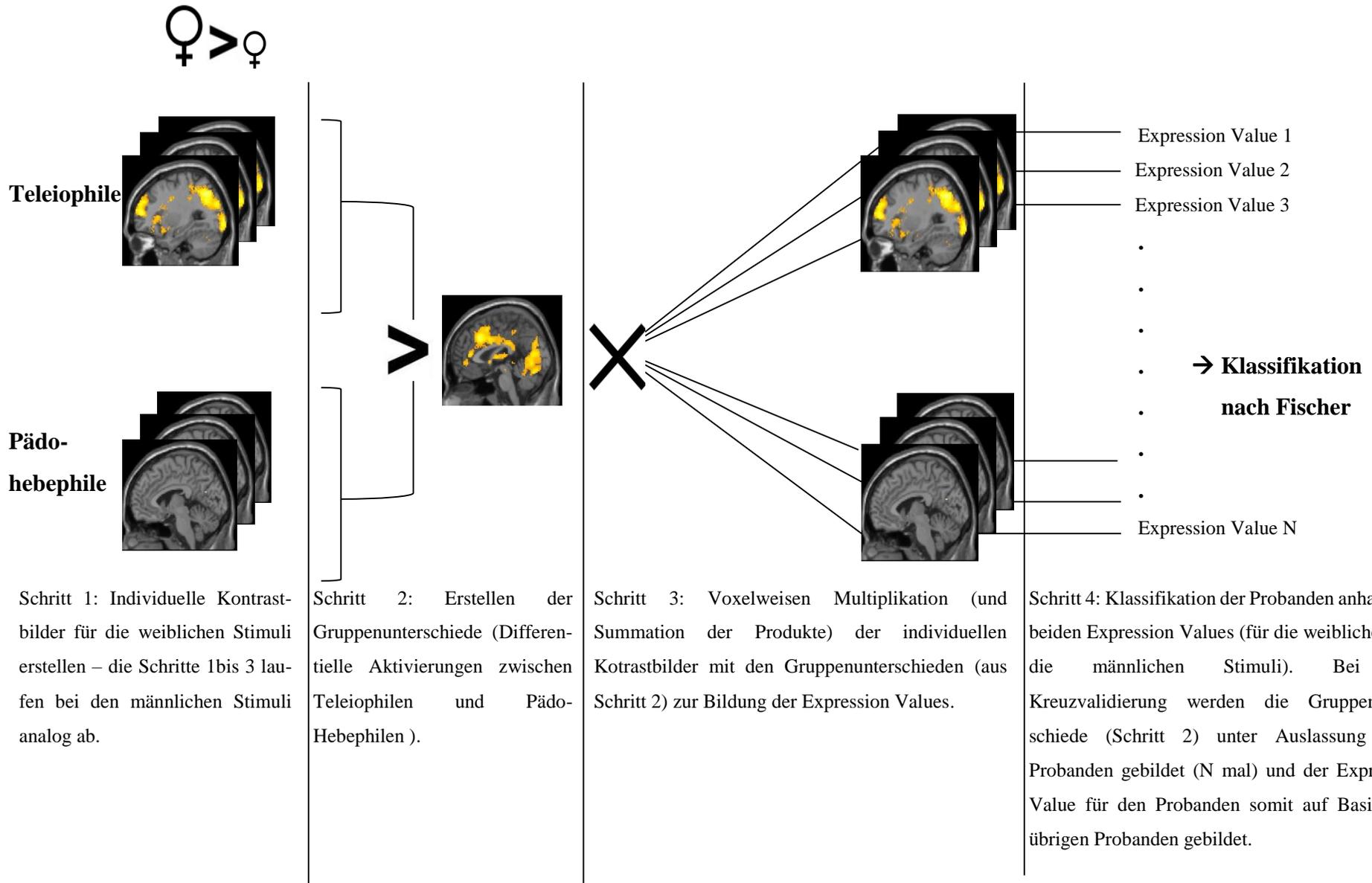


Abbildung 3: Schematische Darstellung der fMRT-Auswertung.

2.2.5 Berechnung der internen Konsistenz und Vergleich der verschiedenen Verfahren

Alle weiteren Analysen habe ich ebenfalls in SPSS 20 gerechnet. Zur Ermittlung der internen Konsistenz habe ich Cronbachs Alpha für die PPG und das VT berechnet. Zudem berechnete ich die Split Half Reliabilität (zwischen der Reaktion auf die letzten drei Bilder des ersten Blocks und den letzten drei Bilder des zweiten Blocks) für PPG, da die Voraussetzung (Unkorreliertheit der Fehler) für die Berechnung von Cronbachs Alpha bei der PPG wegen des fehlenden Interstimulusintervalls zwischen den einzelnen Bildern nicht erfüllt war, was zu einer fehlerhaften Schätzung führen kann. Da die genitale Erregung bei einem Bild erheblich durch die Erregung in Reaktion auf das vorhergegangene Bild beeinflusst wird, kann Cronbachs Alpha so systematisch überschätzt werden. Zur Berechnung der Split Half Reliabilität wurden von den Kategorien „Frauen“, „Männer“, „Mädchen“ und „Jungs“ jeweils die letzten drei Bilder eines jeden Blocks miteinander korreliert.

Die konvergente Validität errechnete ich als Korrelation zwischen den Indizes der verschiedenen Verfahren (Altersindex aus VT und PPG sowie Expression Values und Klassifikationswerte der fMRT). Hierzu wurden die Klassifikationswerte des fMRT mit -1 multipliziert, damit alle Skalen gleich gepolt waren. Die Expression Values und die Klassifikationswerte stammten aus dem Modell nach Kreuzvalidierung. Zur Bestimmung der inkrementellen Validität habe ich die Indizes der verschiedenen Verfahren kombiniert. Dabei habe ich die Indizes der PPG von allen Probanden (ohne Ausschluss von Low-Respondern) genommen. Dies habe ich zum einen gemacht, um die Anzahl der Probanden möglichst groß zu halten. Zum anderen wären die Analysen der Daten nach Ausschluss von Low-Respondern aufgrund der perfekten Diskriminationsleistung der PPG unter diesen Voraussetzungen (siehe 3.2) zwecklos gewesen. Bezüglich des fMRT habe ich die Klassifikationswerte aus den reduzierten Modellen (nach Kreuzvalidierung) verwendet. Beim VT habe ich jeweils die Indizes auf Basis der Rohwerte und der z -standardisierten Daten genutzt.

Die Klassifikationsgüte der einzelnen Verfahren habe ich mittels McNemars Test für verbundene Stichproben verglichen. Dies ist ein robuster Test zum Vergleich von Klassifikationsraten (Dietterich, 1998). Zudem eignet sich dieser Test bei abhängigen Stichproben. Da ich verschiedene Verfahren vergleiche, die an demselben Probandenpool erhoben worden sind, kann nicht von einer Unabhängigkeit der Daten ausgegangen werden. Ich habe diesen Test auch bei dem Vergleich der Klassifikationsleistung eines Verfahrens unter verschiedenen Voraussetzungen - z.B. beim Vergleich von Indizes, die auf Roh-Werten oder z -

standardisierten Werten beruhen - genutzt. Bei dem Vergleich verschiedener Subgruppen (z.B. Vergleich der Klassifikationsgenauigkeit bei hetero- und homosexuellen Probanden) griff ich stattdessen auf den Fischer-Exakt-Test zurück.

3 Ergebnisse

3.1 Vorbemerkungen und Reihenfolge der Testverfahren

Im Folgenden werde ich die Ergebnisse der einzelnen Verfahren vorstellen. Sofern nicht explizit anders erwähnt, beziehen sich alle p -Werte auf zweiseitiges Testen. Als Signifikanzschwelle habe ich einen p -Wert von 5 % gewählt, wobei ich p -Werte unter 10 % als marginal-signifikante Ergebnisse berichte.

Zunächst habe ich geprüft, in welcher Reihenfolge die drei objektiven Verfahren angewandt wurden. Hierzu habe ich bei jedem Probanden ausgezählt, welches Verfahren als erstes, zweites oder drittes durchgeführt wurde. Mittels Wilcoxon-Vorzeichen-Rang-Test habe ich geprüft, ob es zwischen den Verfahren Unterschiede in der Reihenfolge gab. Das VT-Verfahren wurde im Mittel vor den anderen beiden Verfahren durchgeführt (VT vs. PPG: $p = .002$; VT vs. fMRT: $p = .004$). PPG und fMRT wurden hingegen im Mittel an derselben Position durchgeführt ($p = .594$).

3.2 Reliabilität und Validität des VT

Die Betrachtungszeiten pro Gruppe verteilten sich erwartungskonform über die verschiedenen Stimuluskategorien. Wie in vorherigen Studien auch (Imhoff et al., 2010) stieg die Betrachtungszeit bei den teleiophilen Probanden mit ansteigendem Reifegrad der abgebildeten Personen an (siehe Abbildung 4). Dieser Anstieg war bei dem präferierten Geschlecht ausgeprägter als bei dem nicht-präferierten Geschlecht – die mittlere Differenz zwischen der Betrachtungszeit der weiblichen Tanner 5 vs. Tanner 1 Bilder war bei den heterosexuellen Teleiophilen größer als bei den homosexuellen ($t(36) = 3.99, p < .001$), bei den männlichen Bildern war dies genau anders herum ($t(13,69) = -4.79, p < .001$). Die Reaktionszeitprofile der pädo-hebephilen Gruppe sind weniger einfach zu interpretieren. Zum einen sind diese in der Literatur noch nicht so detailliert beschrieben worden. Zum anderen handelt es sich bei diesen Probanden (v.a. in der vorliegenden Arbeit) um eine heterogene

Gruppe, in der die Probanden selten eine exklusive Alters- bzw. Geschlechtsorientierung aufwiesen (siehe 2.1). Während die homopädo-hebephilen die höchste Reaktionszeit bei den männlichen Stimuli der Tanner-Kategorie 3 aufwiesen, lag das Maximum bei den heteropädo-hebephilen Probanden bei Tanner 4.

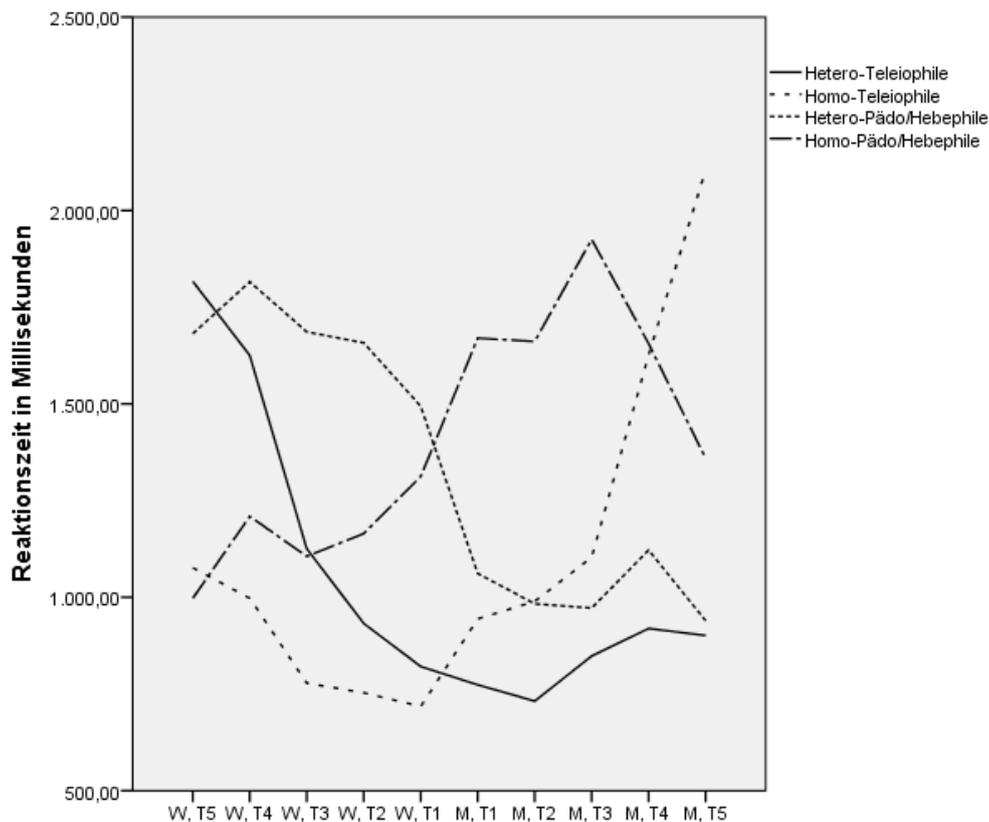


Abbildung 4: Mittlere Betrachtungszeit der zehn Stimulusgruppen: weibliche (w) und männliche (m) Reize der Tannerkategorien 1-5. Abgebildet ist die Betrachtungszeit (vor z-Standardisierung) für jede der vier Gruppen.

Die Berechnung von Cronbachs Alpha des VT ergab für die Rohwerte sehr gute Kennwerte Die z-transformierten Werte lagen jeweils unter den Kennwerten der Rohdaten (siehe Tabelle 4), waren aber immer noch akzeptabel.

Tabelle 4: Werte für die interne Konsistenz (Cronbachs Alpha) für die einzelnen Stimuluskategorien des VT – vor (roh) und nach z-Standardisierung (z). Abkürzungen: W: weibliche Stimuli, M: männliche Stimuli, T1-T5: Tannerkategorien 1-5.

	W, T1	W, T2	W, T3	W, T4	W, T5	M, T1	M, T2	M, T3	M, T4	M, T5
α (roh)	.911	.872	.866	.844	.901	.896	.904	.897	.880	.852
α (z)	.701	.750	.746	.737	.874	.625	.717	.673	.749	.877

Anschließend habe ich die Präferenzindizes gebildet (siehe 2.2.2.3), um die Gruppenzugehörigkeit der Probanden vorherzusagen. Mithilfe dieser Indizes gelang es mir, die

Probanden mittels Diskriminanzanalyse (nach Kreuzvalidierung) mit einer mittleren Genauigkeit von 88.7 % zu klassifizieren. Die Spezifität lag bei 92.1 %, die Sensitivität bei 85.3 %. Nach der probandenweisen z-Standardisierung der Reaktionszeiten lag die mittlere Klassifikationsgenauigkeit bei 87.3 % - bei einer Spezifität von 86.5 % und einer Sensitivität von 88.2 % (siehe Tabelle 5). Die beiden Datenvorverarbeitungsvarianten (mit Rohwerten und z-transformierten Daten) führten zu keiner signifikant unterschiedlichen, mittleren Klassifikationsgenauigkeit (McNemar's Test, ns). Zur Berechnung der AUC habe ich beide Altersindizes gemittelt und als Prädiktor aufgenommen. Dabei ergab sich für beide Datenaufbereitungsmethoden eine AUC von .93. Bei Rohwerten lag der optimale Cut-Off (im Sinne der gleichzeitigen Optimierung von Sensitivität und Spezifität) bei 237 ms, bei den z-Werten bei .22. Für die Verteilung der Probanden entlang der beiden Altersindizes siehe Anhang 7.

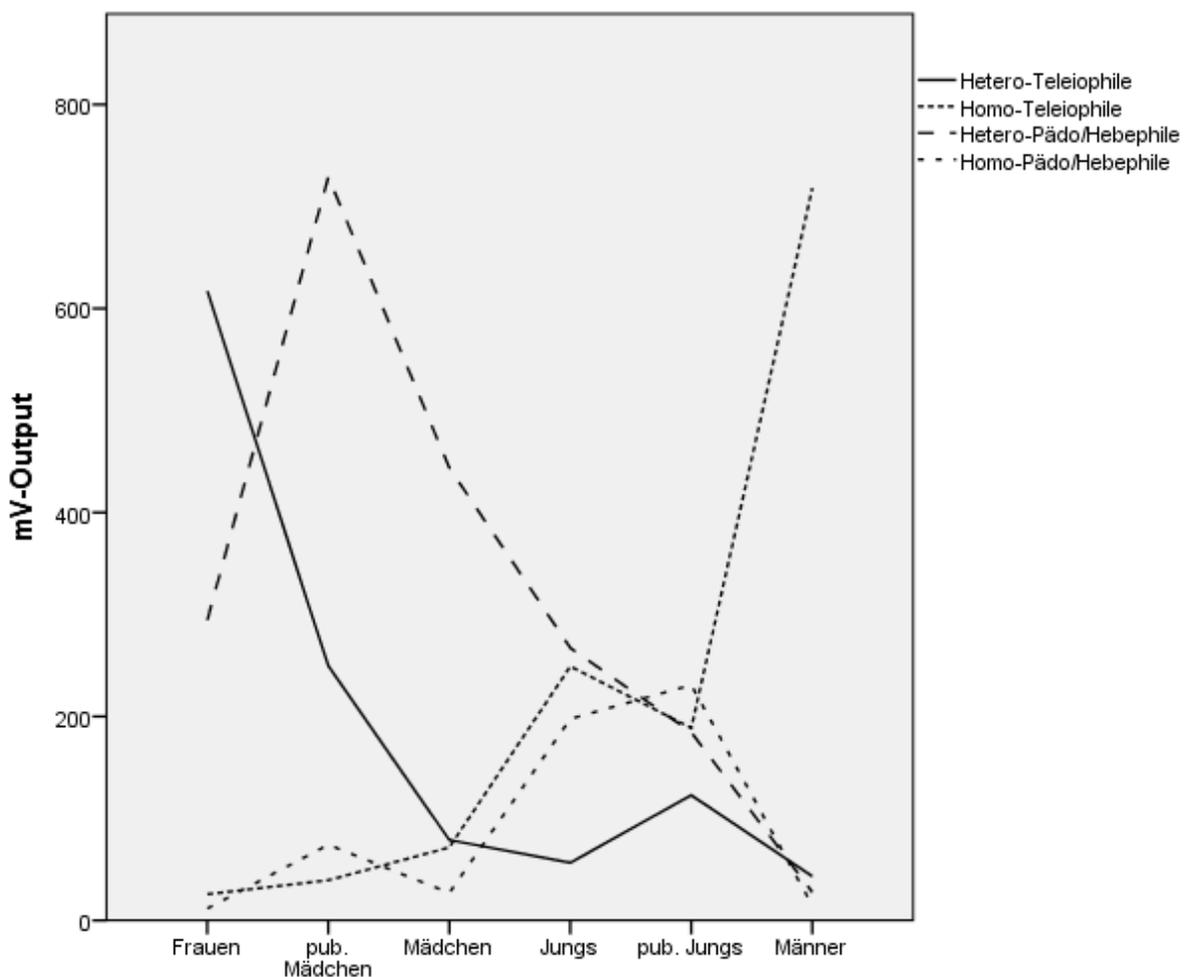
Tabelle 5: Klassifikationsleistung des VT mit den nicht-transformierten (Rohwerte) und den z-transformierten Daten anhand der zwei Indizes (Diskriminanzanalyse nach Kreuzvalidierung).

				Vorhergesagte Gruppenzugehörigkeit	
				Teleiophile	Pädo- Hebephile
Indexberechnung mit Rohwerten	Tatsächlich e	Absolut	Teleiophile	35	3
			Pädo- Hebephile	5	29
	Gruppenzu- gehörigkeit	In	Teleiophile	92.1	7.9
		Prozent	Pädo- Hebephile	14.7	85.3
Indexberechnung mit z- transformierten Werten	Tatsächlich e	Absolut	Teleiophile	32	5
			Pädo- Hebephile	4	30
	Gruppenzu- gehörigkeit	In	Teleiophile	86.5	13.5
		Prozent	Pädo- Hebephile	11.8	88.2

3.3 Reliabilität und Validität der PPG

Die Auswertung über die vier Gruppen (hetero- und homosexuelle Teleiophile bzw. Pädo-Hebephile) ergab erwartungskonforme Reaktionsprofile, welche denen aus einer vorherigen Studie ähnelten (Blanchard et al., 2010). Die heteroteleiphilen Probanden zeigten die stärkste Reaktion bei den erwachsenen Frauen, die homoteleiphilen die stärkste bei den

erwachsenen Männern (siehe Abbildung 5). Die gemischte Gruppe der homosexuellen Pädophil-Hebephilen zeigte eine ähnlich hohe Reaktion bei Jungen der Tanner-Kategorien 1 bzw. 2-3, während die Gruppe der heterosexuellen Pädophil-Hebephilen ein deutliches Maximum bei den weiblichen Stimuli der Tanner-Kategorie 2-3 aufwies. Die Reaktion der Probanden in der homopädophil-hebephilen Gruppe war geringer als in den drei anderen Gruppen ($t(33.93) = -3.14$, $p < .01$). Ausgehend vom maximalen Reaktionswert der Gruppen ist typischerweise eine Abnahme der Reaktionsstärke zu beobachten. So weisen z.B. die heteropädophil-hebephilen Probanden ihre stärkste Reaktion bei pubertierenden Mädchen auf. Die benachbarten Kategorien (erwachsene Frauen und präpubertäre Mädchen) rufen bereits verminderte Reaktionen hervor, während in Bezug auf die männlichen Reize eine sehr geringe Reaktion



auftrat.

Abbildung 5: μ -Volt-Schwankungen (im Vergleich zu den neutralen Stimuli) in Reaktion auf die sechs Stimulus-Gruppen in der PPG aufgeteilt nach der Gruppenzugehörigkeit (vor Low-Responder Ausschluss). Dargestellt sind die Rohwerte (μ -V-Output) auf die einzelnen Kategorien, nachdem hiervon die mittlere Reaktion auf die beiden neutralen Reize (Baseline) abgezogen wurde.

Damit die Berechnung der Reliabilitätswerte nicht durch Extremwerte verzerrt wurde, habe ich zunächst alle Probanden ausgeschlossen, die als Ausreißer bei mindestens einem Stimulus identifiziert wurden. Dies betraf insgesamt 11 Probanden. Die Berechnung der internen Konsistenz der einzelnen Stimuluskategorien ergab sehr hohe Werte (alle Alphas > .993). Die Kennwerte der Split-Half-Reliabilität nach Guttman waren ebenfalls sehr gut (siehe Tabelle 6). Aufgrund der überraschend hohen Indizes habe ich im Anschluss von dem μV -Wert bei Präsentation eines sexuellen Stimulus den μV -Wert bei Präsentation des ersten neutralen Stimulus abgezogen, um einen um den Ausgangswert bereinigte Reaktion auf die einzelnen Stimuli angeben zu können. Dies habe ich getan, da ich vermutete, dass Personen, die während der Messung sehr wenig bzw. gar nicht reagieren, die Reliabilitätswerte nach oben hin verfälschen könnten (bei einer fehlenden Reaktion bei allen Probanden würde eine Reliabilität von 1 erreicht werden). Auch die Korrelation dieser neuen Variablen erbrachten Split-Half-Reliabilitäten nach Guttman zwischen .81 und .92. Selbst nach dem Ausschluss von Low-Respondern (Mindestreaktion von 7) lagen alle Reliabilitätswerte über .72 (.72 bis .92).

Tabelle 6: Split-Half-Reliabilitätskoeffizienten der PPG für die einzelnen Stimulus-Gruppen nach Guttman (berechnet anhand der Rohwerte).

	Mädchen	Männer	Jungs	Frauen	neutral
Split-Half-Reliabilität	.972	.965	.969	.973	.955

Vor einem Ausschluss von Low-Respondern ergab sich bei der PPG nach Kreuzvalidierung eine mittlere Genauigkeit von 90.3 %. Die Spezifität betrug dabei 97.4 %, die Sensitivität 82.4 % (siehe Tabelle 7). Die AUC betrug .96. Der optimale Cut-Off Wert (im Sinne einer Maximierung der mittleren Klassifikationsgenauigkeit) betrug -73.31 μV . Pädophile Probanden wurden nicht signifikant schlechter klassifiziert als hebephile Probanden, in der Tendenz war die Klassifikationsleistung bei pädophilen Probanden allerdings schlechter als bei Hebephilen (Fischer exakter Test: $p = .174$).

Tabelle 7: Klassifikationsleistung der PPG vor dem Ausschluss von Low-Respondern mittels Diskriminanzanalyse nach Kreuzvalidierung.

			Vorhergesagte Gruppenzugehörigkeit	
			Teleiophile	Pädo-Hebephile
Tatsächliche	Absolut	Teleiophile	37	1
		Pädo-Hebephile	6	28
Gruppenzugehörigkeit	In	Teleiophile	97.4	2.6
	Prozent	Pädo-Hebephile	17.6	82.4

Die Rate an korrekt klassifizierten Probanden war positiv mit dem Index für die Reaktionsstärke korreliert ($r = .32, p < .01$). Zudem war der Reaktionsindex bei den teleiophilen Probanden positiv ($r = .87, p < .001$) und bei den pädophilen Probanden negativ mit dem Pädophilie-Index korreliert ($r = -.32, p = .06$). Da dies die Annahme aus der Literatur bestätigt, dass die Klassifikation bei Probanden mit einer stärkeren Reaktion besser funktioniert, als bei Probanden mit geringer Reaktion, wurde die Diskriminanzanalyse noch einmal nach einem Ausschluss von Low-Respondern durchgeführt. Da das von Howes (2003) vorgeschlagene Kriterium (9.4 mm) sehr konservativ ist (was zum Ausschluss von 38.9 % der Stichprobe geführt hätte), habe ich zunächst das von (Kuban et al., 1999) vorgeschlagene Kriterium (2.5 mm) verwendet. Dies führte bereits zu einer Steigerung der Klassifikationsgenauigkeit auf 93.8 %. Nach einer manuellen Inspektion der einzelnen Reaktionsprofile habe ich versucht, einen Cut-off zu finden, ab dem die Schwankungen in den aufgezeichneten Daten nicht mehr zufällig erschienen. Dies schien mir bei den Probanden der Fall zu sein, die mindestens eine Reaktionsstärke von mehr als sieben mm aufwiesen. Dieses Kriterium führte zu einem Ausschluss von insgesamt 18 Probanden (25 % der Stichprobe). Die Stichprobe bestand danach aus 27 teleiophilen und 27 pädo-hebephilen Probanden. Nach dem Ausschluss von Low-Respondern anhand dieses Kriteriums wurde eine perfekte Diskriminationsleistung erzielt: Sowohl die Sensitivität als auch die Spezifität betragen 100 %, die AUC betrug demzufolge 1, der optimale Cut-Off-Wert lag weiterhin bei $-73.31 \mu\text{V}$. Auch nach dem Ausschluss aller Probanden mit einer Reaktion von weniger als 9.5 mm betrug die AUC weiterhin 1. Die mittlere Klassifikationsgenauigkeit nach Kreuzvalidierung sank allerdings auf 95.5 %. Für die Verteilung der Probanden in Bezug auf die Altersindizes für männliche und weibliche Reize siehe Anhang 7.

3.4 Validität der fMRT

Die Gruppenkarten (aus der second level analysis) zeigen die Bereiche, in denen die Teleiophilen im Vergleich zu den Pädophilen eine stärkere Aktivierung bei den entsprechenden Kontrasten (Männer > Jungen bzw. Frauen > Mädchen) aufwiesen. Wie in der Studie von Ponseti et al. (2012), weist dieser Kontrast bei den männlichen Stimuli höhere t-Werte auf als bei den weiblichen Stimuli (Abbildung 6). Im Gegensatz zu der vorherigen Arbeit (in der es bei dem männlichen Kontrast deutlich mehr signifikante Regionen gab) ist die Aktivierung bei beiden Kontrasten in ähnlichen Regionen zu finden und der weibliche Kontrast weist sogar mehr signifikante Unterschiede – vor allem in frontalen und präfrontalen Regionen – auf.

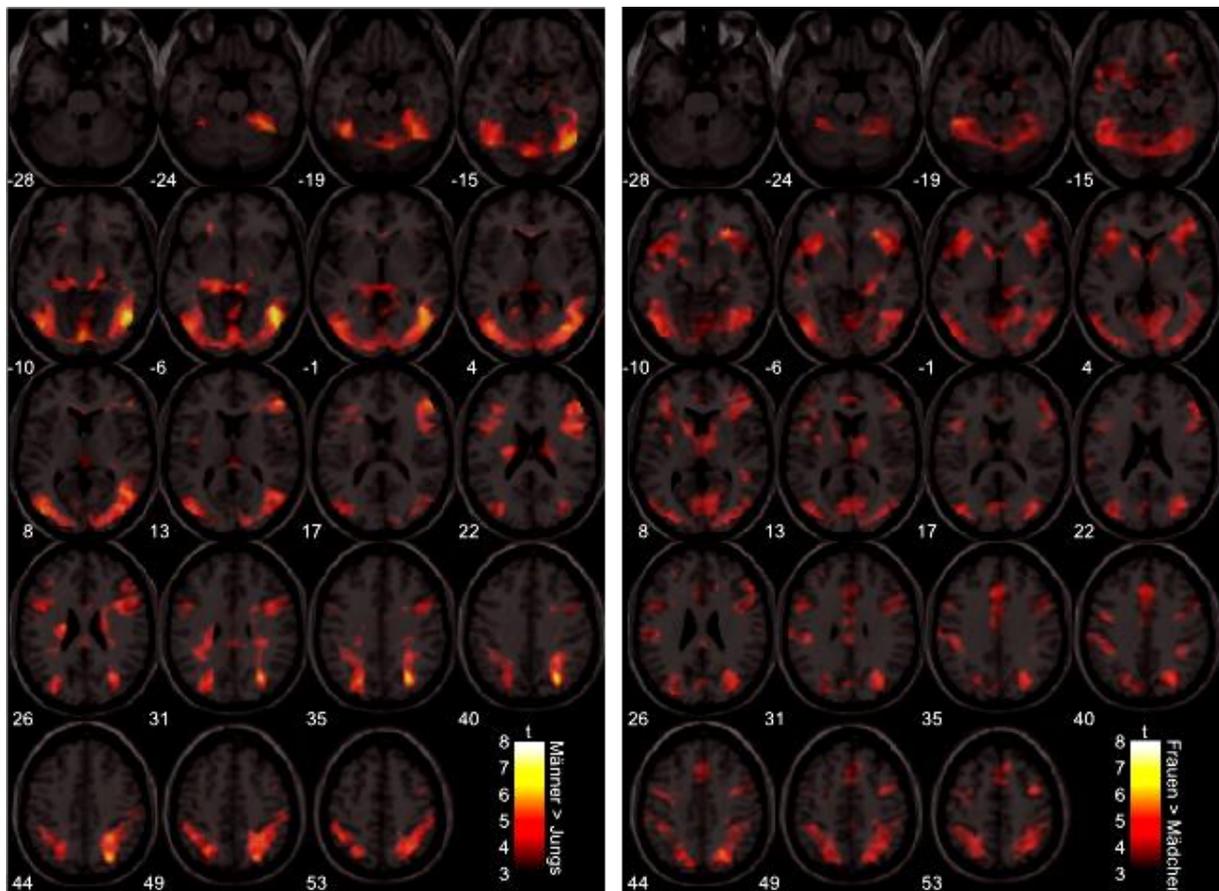


Abbildung 6: T-Karten der Regionen mit erhöhter Aktivierung bei den Kontrasten Männer > Jungen (links) bzw. Frauen > Mädchen (rechts). Hohe t-Werte bedeuten, dass Teleiophile in diesen Regionen eine höhere Aktivierung als Pädophile aufwiesen (in Bezug auf den jeweiligen Kontrast).

Tabelle 8: Absolute Anzahl und Prozentsätze der korrekt klassifizierten Probanden bei den Teleiophilen (Spezifität), Pädophilie (Sensitivität) sowie über die gesamte Stichprobe (mittlere Genauigkeit) beim fMRT. Zudem die *p*-Werte (nach Fischer's exaktem Test) zu den Unterschieden in der Klassifikationsgenauigkeit zwischen den Substichproben.

Gruppenzugehörigkeit	korrekt klassifizierte Probanden		Signifikanzniveau
	absolut	in Prozent	
Teleiophile	34 (von 38)	89.4	
Täter	1 (von 3)	33.3	<i>p</i> = .06
Nicht-Täter	33 (von 35)	94.3	
Androphile	12 (von 14)	85.7	<i>p</i> = .94
Gynäphile	20 (von 22)	90.9	
Pädo-Hebophile	28 (von 34)	82.4	
Pädophile	16 (von 17)	94.1	<i>p</i> = .18
Hebophile	12 (von 17)	70.6	
Androphile	11 (von 11)	100.0	<i>p</i> = .16
Gynäphile	17 (von 23)	73.9	
Gesamte Stichprobe	62 (von 72)	86.1	
Androphile	24 (von 26)	92.3	<i>p</i> = .44
Gynäphile	38 (von 46)	82.6	

Die automatische, kreuzvalidierte Klassifikation nach Fischer führte zu einer mittleren Genauigkeit von 86.1 %. Die Spezifität lag dabei mit 89.5 % knapp oberhalb der Sensitivität mit 82.4 % (siehe Tabelle 8). Von den vier falsch klassifizierten Teleiophilen, hatten zwei sexuellen Kindesmissbrauch begangen. Die teleiophilen Täter ($n = 3$) wurden im Trend signifikant schlechter klassifiziert, als die teleiophilen nicht-Täter ($n = 35$; einseitiger exakter Test nach Fischer: $p = .06$). Zudem waren zwei der falsch klassifizierten Probanden im Vergleich zur restlichen Stichprobe relativ alt (61 und 63 Jahre, jenseits von zwei Standardabweichungen des Mittelwertes). Ein dritter älterer Proband (62 Jahre) wurde korrekt klassifiziert. Die Inspektion der strukturellen Bilder ergab bei diesen beiden falsch klassifizierten Probanden relativ starke altersbedingte Abbauprozesse (siehe Abbildung 7) – welche nach Einschätzung der hiesigen Neurologen aber noch als normale Alterungsprozesse aufzufassen sind.

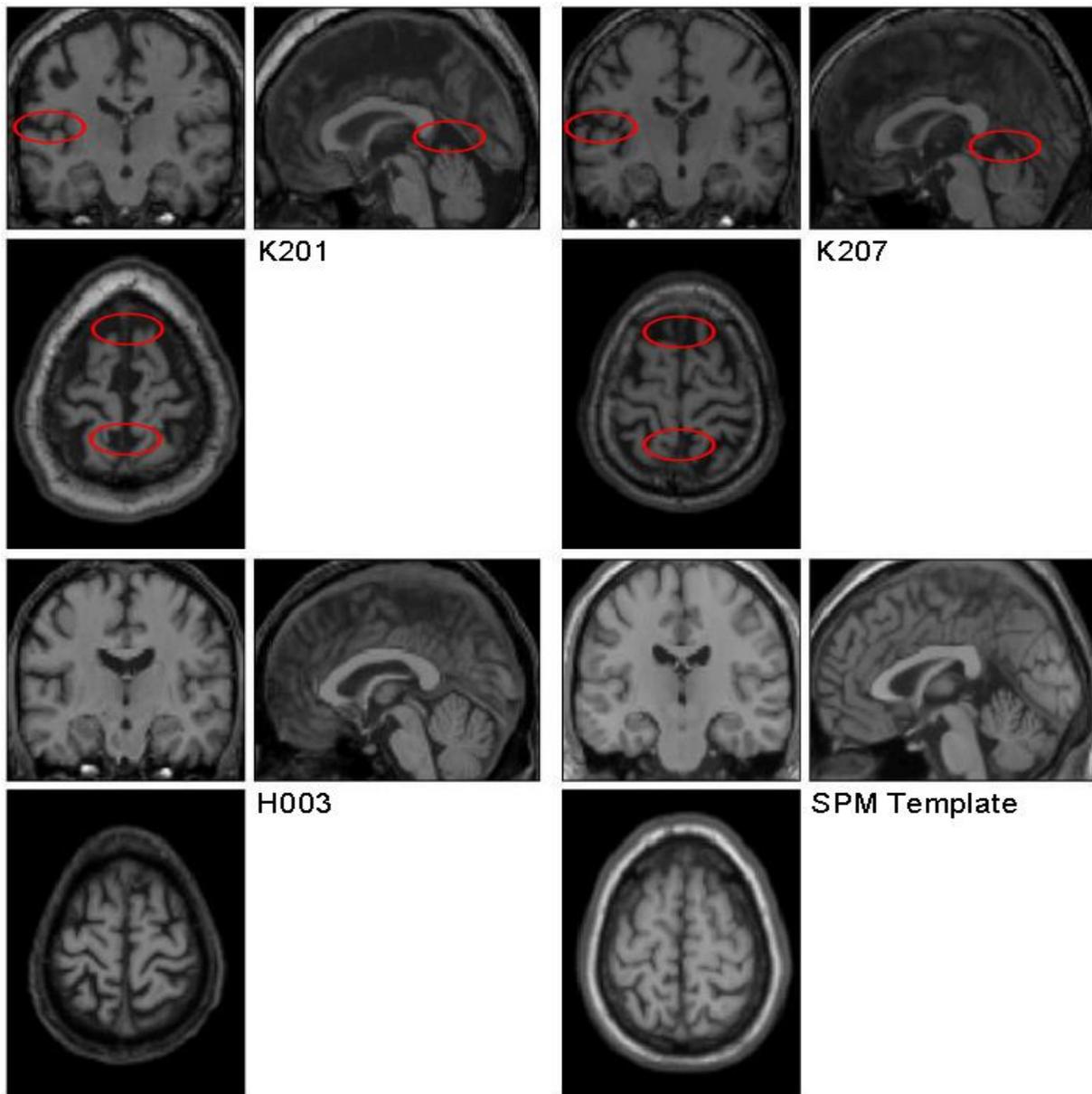


Abbildung 7: Altersbedingte Atrophien zweier älterer Probanden (K201 und K207) im Vergleich zu einem weiteren Probanden im selben Alter (H003) sowie der Vorlage aus SMP, welche u.a. zur Segmentierung der strukturellen Bilder genutzt worden ist. Es sind beispielhaft Regionen markiert, in denen die beiden im fMRT falsch klassifizierten Probanden stärkere Abbauprozesse erkennen lassen als dies z.B. bei H003 der Fall ist.

3.5 Vergleich der Validität der einzelnen Verfahren

In Tabelle 9 ist die Sensitivität, Spezifität und mittlere Klassifikationsgenauigkeit der einzelnen Verfahren aufgeführt. Nach dem Ausschluss von Low-Respondern erzielte die PPG eine signifikant bessere, mittlere Klassifikationsgenauigkeit als alle anderen Verfahren (McNemar Test, alle $p_s < .05$). Zudem lag die Sensitivität der PPG marginal-signifikant über der des fMRT (McNemar Test, $p = .06$) und die Spezifität der PPG marginal signifikant über

der des VT (mit z -transformierten Werten; McNemar Test, $p = .06$). Vor dem Ausschluss von Low-Respondern unterschied sich die PPG nicht signifikant in Hinblick auf die mittlere Genauigkeit von den anderen Verfahren. Alle weiteren Paarvergleiche waren ebenfalls nicht signifikant (McNemar Test, alle $ps > .13$).

Tabelle 9: Klassifikationsleistungen der drei Verfahren in der Übersicht. Prozentzahlen, die mit gleichen Buchstaben markiert sind, unterscheiden sich signifikant, solche mit identischen Ziffern unterschieden sich marginal-signifikant.

	VT		PPG		fMRT
	Rohwerte	z -Werte	mit Low-Respondern	ohne Low-Responder	
Sensitivität	85.3 %	88.2 %	82.4 %	100 % ¹	82.4 % ¹
Spezifität	92.1 %	86.5 % ²	97.4 %	100 % ²	89.5 %
Im Mittel	88.9 % ^A	87.3 % ^B	90.3 %	100 % ^{A,B,C}	86.1 % ^C

3.6 Konvergente und inkrementelle Validität

Die Berechnung der konvergenten Validität ergab mittlere bis hohe Korrelationen zwischen den Indizes der drei Verfahren (siehe Tabelle 10). Im Gegensatz zu der Arbeit von Mackaronis et al. (2014) ergaben sich keine erheblichen Differenzen bzgl. der Korrelationen in Abhängigkeit von der Datenaufbereitung des VT.

Tabelle 10: Korrelationen zwischen den Alters-Indizes der drei Verfahren. Mit ** markierte Korrelationen sind hoch signifikant ($p < .01$).

	VT-Altersindex (roh)	VT-Altersindex (z)	PPG-Altersindex	Klassifikationswert (umgepolt)
VT-Altersindex (roh)	1	.787**	.483**	.524**
VT-Altersindex (z)		1	.494**	.532**
PPG-Altersindex			1	.487**
fMRT-Klassifikationswert (umgepolt)				1

Anschließend habe ich die Korrelationen in Bezug auf die nach Geschlecht aufgeteilten Altersindizes der einzelnen Verfahren berechnet. Da sich die Korrelationen zwischen unstandardisierten und z -standardisierten VT-Werten nicht erheblich unterschieden, sind in Tabelle 11 nur die Korrelationen in Bezug auf die z -standardisierten VT-Werte abgebildet. Hierbei ergeben sich jeweils zwischen den weiblichen und männlichen Indizes der verschiedenen Verfahren mittlere bis hohe Korrelationen. Die Korrelationen zwischen den männlichen und weiblichen Indizes (sowohl innerhalb eines Verfahrens als auch zwischen den

Verfahren) waren meist nicht signifikant miteinander korreliert. Eine Ausnahme war z.B. die signifikante Korrelation zwischen den beiden Indizes des VT.

Tabelle 11: Korrelationen zwischen den Alters-Indizes der drei Verfahren (aufgeteilt nach Geschlecht). Mit ** markierte Korrelationen sind hoch signifikant ($p < .01$), mit * markierte Korrelationen sind signifikant ($p < .05$). Bzgl. des VT sind nur Korrelationen auf Basis der z-standardisierten Werte dargestellt. Abkürzung: EV: Expression Value aus der fMRT.

	VT- Frauenindex	VT-Män- nerindex	PPG-Frau- enindex	PPG-Män- nerindex	EV (weiblich)	EV (männlich)
VT- Frauenindex	1	.235*	.572**	.015	.433**	.213
VT- Männerindex		1	.081	.408**	.102	.539**
PPG- Frauenindex			1	.028	.389**	.292*
PPG- Männerindex				1	-.122	.398**
EV (weiblich)					1	.245*
EV (männlich)						1

Da eine explorative Datenanalyse bzgl. der nach Geschlecht aufgeteilten Indizes eine Reihe von Ausreißern offenbarte, hätte dies die Korrelationen erheblich beeinflussen können. Wie zu erwarten war, wiesen die homosexuellen Probanden z.B. im Mittel höhere Werte in den männlichen Indizes auf, als die heterosexuellen Probanden. Dies führte dazu, dass die homosexuellen Probanden in Bezug auf die männlichen Indizes gehäuft als Ausreißer identifiziert wurden. Deshalb habe ich im zweiten Schritt eine Partialkorrelation zwischen den Indizes – bereinigt um den Einfluss der sexuellen Altersorientierung – gerechnet. Dies hat die Ergebnisse nicht erheblich beeinflusst (siehe Tabelle 12). Weitere Korrelationen zwischen den Stimulus-Kategorien der einzelnen Verfahren sind in Anhang 8 aufgeführt.

Tabelle 12: Partialkorrelation (bereinigt um den Einfluss der sexuellen Geschlechtsorientierung) zwischen den Alters-Indizes der drei Verfahren (aufgeteilt nach Geschlecht). Mit ** markierte Korrelationen sind hoch signifikant ($p < .01$), mit * markierte Korrelationen sind signifikant ($p < .05$). Bzgl. des VT sind nur Korrelationen auf Basis der z-standardisierten Werte dargestellt. Abkürzung: EV: Expression Value aus der fMRT.

	VT Fraue- nindex	VT Män- nerindex	PPG-Frau- enindex	PPG-Män- nerindex	EV (weiblich)	EV (männlich)
VT-Frauenindex	1	.311**	.589**	.137	.367**	.243*
VT-Männerindex		1	.090	.379**	.171	.543**
PPG-Frauenindex			1	.048	.400**	.301*
PPG- Männerindex				1	-.023	.401**
EV (weiblich)					1	.268*
EV (männlich)						1

In den folgenden Analysen habe ich die Prädiktoren der verschiedenen Variablen kombiniert, um die inkrementelle Validität zu bestimmen. Die Diskriminanzanalyse mit mehreren Prädiktoren erzielte dabei jeweils eine bessere Klassifikationsleistung als die Analysen auf Basis nur eines Verfahrens (siehe Tabelle 13). Dabei erwies sich das Modell, welches auf allen drei Verfahren basierte als signifikant besser als das Modell, welches nur auf den fMRT-Daten ($p = .039$) oder nur auf den z -transformierten VT-Daten basierte ($p = .031$). Im Vergleich zu den anderen Modellen auf Basis nur eines Verfahrens, war dieses kombinierte Modell nicht signifikant besser (alle anderen $p > .180$). Die verschiedenen Kombinationen mehrerer Verfahren unterschieden sich ebenfalls nicht signifikant bzgl. ihrer Klassifikationsgenauigkeit voneinander (alle $p > .125$).

Tabelle 13: Spezifität, Sensitivität sowie mittlere Klassifikationsgenauigkeit bei der Kombination von mehreren Prädiktoren.

	VT + PPG		VT + fMRT		PPG + fMRT	VT + PPG + fMRT	
	Roh-Werte	z -Werte	Roh-Werte	z -Werte		Roh-Werte	z -Werte
Sensitivität	88.2 %	91.2 %	88.2 %	88.2 %	88.2 %	91.2 %	94.1 %
Spezifität	97.4 %	94.6 %	89.5 %	94.6 %	94.7 %	97.4 %	97.3 %
Im Mittel	93.1 %	93.0 %	88.9 %	91.5 %	91.7 %	94.4 %	95.8 %

4 Diskussion

4.1 Vorbemerkungen

Zunächst werde ich die Ergebnisse der einzelnen Verfahren zu den Ergebnissen aus vorherigen Studien in Bezug setzen und ggfs. kurz auf Abweichungen bzw. ungewöhnliche Ergebnisse eingehen. Hier werde ich auch ggfs. Vorschläge für die Modifikation der im NeMUP genutzten Methoden anbringen. Erst danach werde ich die drei Verfahren direkt miteinander vergleichen und auf die Aspekte der inkrementellen und konvergenten Validität eingehen. Zum Abschluss diskutiere ich Grenzen bei der Verwendung objektiver Verfahren sowie ethische Aspekte.

Die Reihenfolge in der die Verfahren durchgeführt worden sind, könnte zu einer Beeinflussung und somit zu Reihenfolgeeffekten führen. Das VT-Verfahren wurde signifikant häufiger vor den anderen beiden Verfahren durchgeführt. Da sich sowohl die Stimuli als auch

die Mechanismen auf denen die Messung beruht zwischen den Verfahren erheblich unterschieden, gehe ich davon aus, dass dies keinen ausschlaggebenden Effekt auf die Ergebnisse gehabt haben wird. FMRT und PPG wurden hingegen im Mittel an derselben Position durchgeführt weshalb ich eine systematische Beeinflussung dieser Verfahren untereinander ausschließen kann. Deshalb gehe ich auf die Reihenfolge der einzelnen Messverfahren nicht weiter ein.

4.2 Diskussion VT

Bezüglich der internen Konsistenz erreichte die Verfahren ähnliche Kennwerte wie andere Autoren. In vorherigen Studien wurden Kennwerte zwischen .72 und .98 berichtet (Banse et al., 2010; Letourneau, 2002; Mokros et al., 2013; Silva, 2011; Worling, 2006). In der vorliegenden Arbeit wurden Werte zwischen .84 bis .91 für die Rohwerte und .70 bis .88 für die *z*-transformierten Werte erreicht. Auch die Mittelwertsverläufe der Betrachtungszeit im VT-Verfahren (siehe Abbildung 4) ähneln den Verläufen, wie sie in vorherigen Arbeiten berichtet wurden (Imhoff et al., 2012). Dies weist darauf hin, dass es im NeMUP Projekt gelungen ist, ausgeprägte und erwartungskonforme Reaktionszeitverläufe – auch ohne explizit sexuelle Stimuli – zu generieren. Dies spiegelt sich auch in der relativ guten Klassifikationsgenauigkeit wider. Während die Studien von Abel und Kollegen (Abel et al., 1998; Abel et al., 2001) eine mittlere Klassifikationsgenauigkeit von ca. 85 % berichteten, erreichten andere Studien nur Werte um 70 % (siehe z.B. Gress, 2005; Mokros, Dombert, Osterheider, Zappala & Santtila, 2010). Somit konnten in der vorliegenden Studie gute Validitätswerte (mittlere Klassifikationsgenauigkeit zwischen 87.5 % und 88.7 %) erreicht werden. Zudem wurden in den Studien von Abel und Kollegen teils nicht ausschließlich VT-Indizes zur Klassifikation genutzt und die Auswertung war nicht ausreichend transparent beschrieben. Im Vergleich zu den anderen Studien (Gress, 2005; Mokros et al., 2010) ergeben sich somit sogar bessere Klassifikationswerte. Auch der hier berichteten AUC Wert von .93 lag deutlich über denen aus den Studien von Banse et al. (2010) und Worling (2006).

Da im NeMUP Projekt ein klassisches VT-Design genutzt wurde und ich mich bzgl. der Auswertung an Vorgaben aus der Literatur orientiert habe, können die im Vergleich zu Vorstudien besseren Klassifikationsergebnisse nicht darin begründet sein, dass ein überlegenes Design genutzt wurde. Wir nutzten sogar ein fast identischen VT wie Banse et al. (2010), die für die AUC Werte von .86 erreichten. Ein Grund für die höhere Zuverlässigkeit des VT in der vorliegenden Arbeit könnte an der Gruppenzusammensetzung liegen. Ich nutzte eine päd-

hebephile Stichprobe, in der sich zum Zeitpunkt der Testung keiner der Probanden in Haft befand. Um als pädo- bzw. hebephiler Proband aufgenommen zu werden, musste zudem eine Reihe von Kriterien erfüllt sein (siehe 2.1). Weiterhin gaben die meisten Probanden dieser Gruppen ihre sexuelle Orientierung offen zu. Die Kontrollgruppe bestand aus teleiophilen Probanden aus der Allgemeinbevölkerung (sowie drei Probanden, die sexuellen Kindesmissbrauch begangen hatten). Auch in dieser Gruppe habe ich zwei Probanden von den Analysen ausgeschlossen, da diese bzgl. ihrer sexuellen Altersorientierung keine hinreichend eindeutigen Angaben machten. Somit ergab sich ein geringeres Risiko, Personen aus Versehen der falschen Gruppe zuzuordnen. Banse et al. (2010) hingegen schlossen inhaftierte Personen in die Experimentalgruppe ein, die sexuellen Kindesmissbrauch begangen hatten. Wie hoch der Anteil an pädophilen Probanden und wie hoch der Anteil an Non-Admittern in der Stichprobe war, haben die Autoren nicht angegeben. Auch andere Studien (Gress, 2005; Letourneau, 2002; Mokros et al., 2010; Worling, 2006) stützten sich auf forensische Stichproben und berichteten nicht immer den Anteil von Pädophilen bzw. Non-Admittern in ihren Stichproben. Dieser Unterschied in der Gruppenzusammensetzung mag sich auf zweierlei Arten ausgewirkt haben: (1) Zum einen kann es sein, dass nicht alle Probanden aus der Experimentalgruppe aus vorherigen Studien pädo- bzw. hebephil waren; (2) zum anderen kann bei Probanden, die zum Zeitpunkt der Testung inhaftiert waren, eine erhöhte Motivation vorgelegen haben, die Testergebnisse zu beeinflussen, vor allem, wenn sie davon ausgingen, dass die Testergebnisse der Haftanstalt übermittelt werden könnten (Letourneau, 2002). Diesen Personen könnte eine willentliche Beeinflussung der Testergebnisse durchaus gelungen sein (Imhoff et al., 2012). Während die erste Möglichkeit zu einer systematischen Unterschätzung der wahren Klassifikationsgenauigkeit führen würde (Blanchard et al., 2001), stellt die zweite Fehlerquelle eine tatsächliche Einschränkung der Validität dar. Ob es sich in der vorliegenden Arbeit daher um eine genauere oder eine zu liberale Schätzung der Validität des VT handelt, ist schwer einzuschätzen.

Auf Grundlage der Daten und der Erfahrungen mit dem VT lassen sich keine zwingenden Empfehlungen für eine Weiterentwicklung ableiten. Bei weiteren vergleichenden Arbeiten wäre ggfs. die Verwendung von möglichst ähnlichen Stimuli bei allen Verfahren zu empfehlen. Beim VT könnten z.B. dieselben Stimuli wie beim fMRT genutzt werden. Die Verwendung von Roh- bzw. z -transformierten Werten hatte keinen erheblichen Einfluss auf die Validität der Messung. Allerdings konnten die Daten eines Probanden wegen zu langer Reaktionszeiten nach z -Transformation und Ausschluss von Ausreißern nicht mehr ausgewertet werden. Zudem erzielten die z -transformierten Daten geringere Werte in Bezug auf die interne Konsistenz

(Cronbachs Alpha), weswegen ich eine Auswertung des VT anhand von Indizes auf Basis von Rohdaten empfehle.

4.3 Diskussion PPG

Insgesamt konnte ich erwartungskonforme Gruppenverläufe berichten. Von diesen idealtypischen Verläufen gab es drei Abweichungen: (1) Die erhöhte Reaktion der Homoteleiohilien bei den männlichen Stimuli der Tanner-Kategorie I, (2) die erhöhte Reaktion der Heteroteleiohilien bei den männlichen Stimuli der Tanner-Kategorie II sowie (3) die erhöhte Reaktion der Homopädohebehilien bei der weiblichen Tanner-Kategorie II. Die Abweichungen von idealtypischen Gruppen-Reaktionsprofilen gehen vermutlich auf verschiedene Quellen zurück: Während (1) und (2) vermutlich Hinweise auf Reihenfolgeneffekte liefern, ist (3) wahrscheinlich durch die Inhomogenität der pädohebehilien Gruppen in Bezug auf ihre Geschlechtsorientierung verursacht (in dieser Gruppe befinden sich sowohl bisexuelle als auch homosexuelle Probanden). Die Vermutung, dass Reihenfolgeneffekte die Ergebnisverläufe der PPG beeinflusst haben, bestätigte sich durch die Inspektion einiger Reaktionsprofile einzelner Probanden. So reichte bei einigen Probanden das ISI von 30 s zwischen den Kategorien nicht aus und es kam (sogar trotz des Entfernens der ersten drei Stimuli jeder Kategorie) bei manchen Probanden zu Überhangseffekten, da die Erektion in diesem Zeitraum nicht vollständig nachließ. Dies entspricht auch dem Vorgehen anderer Forschergruppen (siehe z.B. Blanchard et al., 2001).

Wie auch beim VT, erzielte ich im Vergleich zu vorherigen Studien bei der PPG überdurchschnittlich gute Ergebnisse. Sowohl die Kennwerte der Reliabilität (Cronbachs Alpha und Split-Half-Reliabilität) als auch der Validität lagen über den Werten, die vorher in der Literatur berichtet wurden – dies galt sowohl vor (90.3 %) als auch nach dem Ausschluss von Low-Respondern (100 %). Bisher gelang nur einer Autorengruppe eine perfekte Vorhersageleistung (Chaplin et al., 1995), allerdings bei einer geringeren Anzahl von Probanden.

Die überdurchschnittlich guten Ergebnisse führe ich – wie beim VT – auf die Gruppenkomposition zurück. Im Vergleich zu anderen Studien (siehe z.B. Blanchard et al., 2001) haben wir fast keine Non-Admitter in der Gruppe der pädohebehilien Probanden gehabt und die Probanden befanden sich zum Zeitpunkt der Testung nicht in Kontakt mit Justizbehörden, was die Absicht, die Messungen zu verfälschen erheblich beeinflusst haben kann (Blanchard et al., 2007). Es ist durchaus möglich, dass die Klassifikationsergebnisse bei

einem höheren Anteil von Probanden, die motiviert gewesen wären, die Messung zu verfälschen, schlechter ausgefallen wären (Freund, 1961). Inwieweit die Verwendung der STT möglichen Verfälschungstendenzen einiger Probanden entgegen wirken konnte, kann leider nicht bestimmt werden. Gleichzeitig ist es auch möglich, dass die positiven Ergebnisse der vorliegenden Arbeit zumindest partiell dadurch zu erklären sind, dass es durch den Prozess der Gruppeneinteilung zu wenigen oder keinen falschen Diagnosen gekommen ist. Dies würde unmittelbar eine exaktere Einschätzung der Validität im Vergleich zu vorherigen Studien erlauben (Blanchard et al., 2001).

Der Ausschluss von Low-Respondern hatte einen erheblichen Einfluss auf die Klassifikationsgenauigkeit und scheint daher wichtig zu sein, um die Messergebnisse korrekt interpretieren zu können. Der mittels ROC-Analyse ermittelte optimale Cut-Off Wert von etwa $-73\mu\text{V}$ weist darauf hin, dass einige teleiophile Probanden im Mittel auf Stimuli von vor- und/oder intrapuberalen Kindern stärker reagiert haben als bei der Präsentation von erwachsenen Reizen. Dies ist allerdings nicht ungewöhnlich. In der Literatur wird immer wieder berichtet, dass ca. 20 % aller Probanden aus der Kontrollgruppe (vermeintlich teleiophile Männer) stärker oder gleich stark auf vor- und/oder intrapuberalen Kinder (im Vergleich zu Erwachsenen) reagieren (Fedora et al., 1992; Firestone, Bradford, Greenberg & Nunes, 2000; Freund & Watson, 1991; Hall, Hirschman & Oliver, 1995; Kolla et al., 2010; Marshall, Barbaree & Christophe, 1986; Seto & Lalumiere, 2001; Seto et al., 2000)⁷. Somit sind die Ergebnisse der PPG-Messungen vergleichbar mit den Ergebnissen aus vorherigen Studien – abgesehen von den besseren Kennwerten bzgl. Reliabilität und Validität, die aber auch beim VT zu verzeichnen sind.

Um das Problem von Überhangseffekten zu umgehen (und so auch die Anzahl der effektiv genutzten Stimuli zu erhöhen), empfehle ich bei der zukünftigen Nutzung der PPG eine Pause zwischen den einzelnen Stimulusblöcken, die solange anhält, bis die Probanden zu ihrem Ausgangsniveau zurückgekehrt sind – also z.B. dem Penisumfang, der bei Präsentation der neutralen Stimuli erreicht wurde. Dies kann gegebenenfalls mit einer Aufgabe erreicht werden, die mit der sexuellen Erregung interferiert. Lalumière und Harris (1998) empfehlen ein solches Vorgehen in PPG-Messungen. Dies konnte allerdings wegen technischer Beschränkungen in

⁷ Die Gründe für diesen hohen Anteil an teleiophilen Probanden, der stärker auf kindliche als auf erwachsene Reize reagiert, wurden bisher nicht ausführlich in der wissenschaftlichen Literatur diskutiert. Ein aktueller Aufsatz, der allerdings nicht in einem wissenschaftlichen Journal publiziert worden ist, widmet sich diesem Thema (Schuster, F. (2014). *Warum werden rund 20 Prozent der Männer in phallometrischen Studien stärker durch Vorpubertierende oder Pubertierende als durch Erwachsene sexuell erregt?*; abgerufen von https://www.ipce.info/sites/ipce.info/files/biblio_attachments/warum_20_prozent.pdf am 11.05.2015)

der vorliegenden Arbeit nicht umgesetzt werden. Die verwendete Hardware machte ein solches Vorgehen unmöglich. Auch wäre bei einer größeren Anzahl unabhängiger Stimuli – also Stimuli, bei denen sich die Reaktionen nicht wechselseitig bedingen – eine bessere Schätzung der Reliabilität möglich. Das im Vergleich zur vorherigen Studie modifizierte Stimulusset sollte dabei beibehalten werden (Pohl, 2009), da die Verwendung der hebephilen Stimuli die Klassifikationsgenauigkeit erhöhte. Als ich zur Klassifikation der Probanden nur die teleio- und pädo-philien Stimuli verwendet habe, sank die mittlere Klassifikationsgenauigkeit auf 75 % vor bzw. 83.3 % nach dem Ausschluss von Low-Respondern.

4.4 Diskussion fMRT

Die Genauigkeit der Klassifikation mittels fMRT lag unterhalb der auf Basis der Studie von Ponseti et al. (2012) erwarteten Werten. Dies spiegelte sich auch in den Gruppenkarten (second level) wider. Die Aktivierungsmuster wiesen in der vorliegenden Arbeit zwar z.T. höhere *t*-Werte auf (was vermutlich an der größeren Stichprobe in dieser Studie liegt), es waren aber weniger Regionen signifikant. Dies zeigte sich auch in geringeren Aktivierungen auf dem 1st level. Die Gründe hierfür sind schwer auszumachen. Die Messeinstellungen stimmten weitestgehend mit denen aus der Studie von Ponseti et al. (2012) überein. Es gab nur kleine Abweichungen (die TE betrug statt 36.8 nun 30 ms, die TR 2400 statt 2500 ms der Flipwinkel 80 statt 90 °), die keinen erheblichen Einfluss auf die Stärke des gemessenen Effekts haben sollten.

Die Zusammensetzung der Experimentalgruppen unterschied sich allerdings zwischen den Studien. Während bei Ponseti et al. (2012) nur pädo-philie Probanden untersucht wurden, waren im NeMUP Projekt auch hebephile Probanden mit aufgenommen worden. Da die Klassifikationstoolbox, die ich zur automatische Klassifikation genutzt habe, nur eine Arbeit mit zwei Gruppen erlaubt, war eine Aufteilung der pädo-hebephilen Probanden in zwei separate Subgruppen nicht möglich. Eine Auswertung mit einer derart gemischten Gruppe ist allerdings auch bei der PPG üblich (siehe z.B. Freund & Watson, 1991); nur auf Gruppenebene ist bisher zwischen pädo- und hebephilen Probanden differenziert worden (Blanchard et al., 2010). Da ich daher bzgl. der anderen Verfahren die Gruppe der pädo- und hebephilen Probanden als eine Gruppe behandelt habe, stellt dies kein Mangel des fMRT-Paradigmas dar. Nach der Klassifikation über die zwei Gruppen habe ich allerdings getrennt nach Pädo- und Hebephilen die Anzahl an korrekt und falsch klassifizierten Probanden ausgewertet (siehe Tabelle 8). Dies ergab eine in der Tendenz geringere Trefferquote bei den hebephilen im Vergleich zu den

pädophilen Probanden, was wiederum mehrere Ursachen haben könnte. Zum einen ist die korrekte Gruppenzuordnung der hebephilen Probanden schwieriger und weniger eindeutig gewesen. Während im Interview bei den pädophilen Probanden exploriert werden musste, ob sich die stärkste sexuelle Ansprechbarkeit auf vorpubertäre Kinder (Tanner 1) oder Erwachsene (Tanner 4 oder 5) richtet, musste bei den hebephilen zwischen einer primären Präferenz für Jugendliche (Tanner 2 oder 3) oder Erwachsene (Tanner 4 oder 5) differenziert werden. Dies könnte zu einer höheren Rate an Fehlzuweisungen geführt haben, da einige als hebephil klassifizierte Probanden evtl. eher eine teleiophile bzw. ephebophile Orientierung aufwiesen. Sollte es zu falschen Gruppenzuordnungen gekommen sein, hätte dies allerdings auch die anderen Verfahren beeinflussen müssen.

Wie oben dargestellt, wurde in Bezug auf VT und PPG aber eher eine höhere als zu erwartende Genauigkeit erzielt. Es könnte ebenfalls sein, dass die hebephilen Probanden mittels der pädophilen Stimuli (im fMRT) schlecht zu klassifizieren sind. In Analysen, die ich mit Frau Gewinn gemeinsam durchgeführt habe (welche allerdings nicht Gegenstand der vorliegenden Arbeit sind), stellte sich heraus, dass auch die Hinzunahme unserer als hebephil konzipierten Stimuli die Klassifikationsgenauigkeit in dieser Gruppe nicht erheblich verbesserte. Es könnte aber sein, dass die Stimuli den von hebephilen Männern präferierten Altersbereich nicht gut genug widerspiegelten. Vielleicht waren die abgebildeten Personen zu alt, was eine Differenzierung zwischen hebe- und teleiophilen Personen ggfs. schwer möglich machte. Auf der anderen Seite handelt sich hier um nicht signifikanten Befund; es kann also auch sein, dass sich dieser Befund in späteren Studien nicht replizieren lässt. Des Weiteren nahmen an der Studie drei (vermutlich) nicht-pädophile Personen teil, die sexuellen Kindesmissbrauch begangen hatten. Diese wurden marginal-signifikant schlechter klassifiziert als die restlichen teleiophilen Probanden, was auf eine Schwäche des fMRT bei Ersatztätern hinweisen könnte. Zu beachten ist aber die geringe Anzahl von Probanden in dieser Substichprobe. Um diese Sachverhalte abschließend zu klären, sind Studien mit mehr Probanden notwendig.

Eine weitere potentielle Erklärung für die Abweichungen zwischen den beiden Studien stellt die Veränderung bzgl. der genutzten Stimuli dar. Ponseti et al. (2012) präsentierten ihren Probanden Bilder von Erwachsenen und vorpubertären Kindern (jeweils Bilder von Gesichtern und Genitalien sowie Ganzkörperaufnahmen) sowie nicht-sexuelle Bilder aus dem IAPS (Lang, Bradley & Cuthbert, 1999). Im NeMUP Projekt wurden Bilder von Erwachsenen sowie vor- und intrapubertären Kindern (jeweils Gesichts- und Ganzkörperaufnahmen) gezeigt. Zur Auswertung nutzten Ponseti et al. (2012) nur die Ganzkörper- und Genitalaufnahmen, wobei auch mittels Gesichtsbilder eine gute Klassifikationsgenauigkeit erreicht wurde (Ponseti et al.,

2015). Zudem wurden in der vorherigen Studie 35 Bilder pro Kategorie genutzt, wir verwendeten nur 30 Bilder pro Kategorie. Es ist nicht auszuschließen, dass sich die andere Art der Stimuluspräsentation auf die Klassifikationsgenauigkeit ausgewirkt hat, wobei eine Reduktion in der Stärke der first level Kontraste hierdurch nur schwer zu erklären ist. Dies ließe sich evtl. besser durch die Verringerung des ISI von (im Mittel) vier auf zwei Sekunden erklären. Durch das kurze ISI ist eine Überlappung der Reaktionen auf verschiedene Reizklassen denkbar, was zu einer Verwischung des Effekts beigetragen haben mag.

Das hohe Alter sowie damit einhergehende altersbedingte Atrophien (siehe Abbildung 7) könnten zudem die Fehlklassifikation einzelner Probanden erklären. Es ist durchaus denkbar, dass die Aktivierungen aufgrund der atrophischen Prozesse bzgl. der räumlichen Lage nicht mit den Aktivierungen jüngerer Probanden in Übereinstimmung gebracht werden können. Zusammenfassend lässt sich also sagen, dass die fMRT-Klassifikation (im Gegensatz zum VT und zur PPG) hinter den Erwartungen zurück blieb, allerdings immer noch deutlich bessere Ergebnisse lieferte als auf Grundlage der Ratewahrscheinlichkeit zu erwarten wäre.

Zur Optimierung der Power des fMRT-Paradigmas sollte das ISI wieder auf den ursprünglichen Wert von im Mittel 4 Sekunden erhöht werden. Alternativ könnte eine Aufstockung der Nullevents geschehen. Zudem könnte es notwendig sein, die Testung auf verschiedene Altersbereiche einzuschränken bzw. Personen mit erhöhten Atrophien auszuschließen. Ob sich unter diesen Bedingungen Gesichtsstimuli (wieder) zur Klassifikation eignen, sollte geklärt werden. Zudem sollte getestet werden, ob mit dem veränderten ISI hebephile Stimuli die Klassifikationsleistung erhöhen, oder ob die bisher verwendeten hebephilen Stimuli nicht zur fMRT-Klassifikation dienlich sind.

4.5 Vergleich der Verfahren

Die PPG erreichte in den Auswertungen zwar die höchste mittlere Klassifikationsgenauigkeit (90.3 %), unterschied sich allerdings nicht signifikant von den anderen beiden Verfahren (mittlere Genauigkeit zwischen 86.1 und 88.9 %). Einzig nach dem Ausschluss von Low-Respondern konnte die PPG aufgrund einer perfekten Diskriminationsleistung signifikant bessere Werte erzielen als die anderen Verfahren. In diesem Fall konnte die PPG aber in rund 25 % der Fälle aufgrund zu geringer Reaktionen gar keine Aussage treffen. Deshalb konnte sich meiner Einschätzung zufolge auf Basis der vorliegenden Daten – gemessen an der Validität – keines der drei Verfahren als klar überlegen herausstellen. Auch bezüglich der Reliabilität ergaben sich keine erheblichen Unterschiede zwischen PPG

und VT. Zwar erreichte die PPG bessere Werte für die interne Konsistenz (alle $\alpha > .95$), aber auch beim VT lag das Cronbachs Alpha bei den Rohwerten durchweg in einem guten bis exzellenten Bereich (0,87 - 0,91). Die Korrelationen bei den z-transformierten Daten war geringer (.70 - .88). Somit scheint die PPG dem VT in Bezug auf Reliabilität und Validität – zumindest im Trend – überlegen zu sein.

Bei der Verwendung von psychometrischen Messverfahren sind neben der Zuverlässigkeit allerdings noch weitere Aspekte zu beachten. Hierzu gehören z.B. ökonomische Aspekte wie Personal-, Zeit und Kostenaufwand. In Hinblick auf diese ökonomischen Aspekte weist das VT klare Vorteile gegenüber den anderen Verfahren auf. Die Durchführung dauert nur etwa fünf Minuten und es ist sowohl für die Durchführung als auch für die Auswertung keine aufwändige Schulung der Mitarbeiter nötig. Zudem ist die Durchführung weitgehend kostenneutral und es sind keine sexuell expliziten Stimuli notwendig. Ein erheblicher Schwachpunkt des Verfahrens ist allerdings die vermutlich erhebliche Beeinflussbarkeit seitens des Probanden (Imhoff et al., 2012), was die Verwendung des VT bei Non-Admittern erschweren könnte.

Die PPG hingegen benötigt die Anschaffung von geeigneter Hardware (Messschlaufen und Vorverstärker) und Software (zur Stimuluspräsentation, Datenaufzeichnung und Auswertung). Zudem benötigt die Durchführung einer PPG-Messung ein erheblich größeres Maß an Vorbereitung und Schulung. Die Messung an sich dauert etwa eine Stunde. Des Weiteren ist die Verwendung sexueller Stimuli notwendig. Zwar ist auch die PPG willentlich durch den Probanden verfälschbar (Freund et al., 1988), es gibt aber auch Möglichkeiten, dieser Beeinflussung entgegenzuwirken – auch wenn nicht alle Studien zu dem Thema die Effektivität entsprechender Kontrollmechanismen belegen konnten (Harris et al., 1999; McAnulty & Adams, 1991; Quinsey & Chaplin, 1988).

Bezüglich der fMRT-Klassifikation ist noch unklar, inwieweit diese der willentlichen Beeinflussbarkeit durch den Probanden unterliegt. Auch das fMRT ist vermutlich nicht auf explizit sexuelle Stimuli angewiesen (Ponseti et al., 2015), benötigt allerdings mit Vor- und Nachbereitung ebenfalls mindestens zwanzig Minuten pro Testung. Zudem können einige Probanden z.B. aufgrund des starken Magnetfeldes nicht an der Messung teilnehmen. Darüber hinaus ist die Auswertung komplizierter als bei den anderen Methoden. Unter rein ökonomischen Aspekten ist das VT den anderen Verfahren damit klar überlegen.

4.6 Konvergente und inkrementelle Validität

Wie bei der PPG zu erwarten ist, waren die Altersindizes für das männliche und weibliche Geschlecht nicht miteinander korreliert. Beim VT ergab sich zwischen den beiden Indizes eine geringe, signifikante Korrelation. Dies ist vor dem Hintergrund der sequentiellen Prüfstrategie, welche nach Imhoff et al. (2012) Grundlage des VT-Effektes ist, zu erwarten. Diese führt dazu, dass die Reaktionszeiten bei Teleiophilen mit zunehmendem Reifegrad ansteigen – wenngleich dieser Anstieg in den Antwortlatenzen bei dem präferierten Geschlecht stärker ausfällt. Die Ergebnisse des fMRT können aufgrund fehlender Vorarbeiten nicht in Bezug zu vorherigen Studien gesetzt werden. Bezogen auf die Interkorrelationen zwischen den einzelnen Verfahren ergaben sich ebenfalls erwartungskonforme Ergebnisse. Die Werte des weiblichen Altersindex eines Verfahrens waren mit den weiblichen Altersindizes der anderen Verfahren signifikant korreliert. Dasselbe galt für die männlichen Indizes. Sofern der Altersindex eines Geschlechts mit beiden Altersindizes eines anderen Verfahrens korreliert war, war die Korrelation mit dem Index desselben Geschlechts höher als die Korrelation mit dem Index des Gegengeschlechts. Diese Befunde können als Indikator dafür gewertet werden, dass die verschiedenen Verfahren dasselbe Konstrukt (sexuelle Orientierung) messen. Dass die Korrelationen sich überwiegend in einem mittleren Bereich bewegen, deutet darauf hin, dass die drei Verfahren unterschiedliche Aspekte der sexuellen Orientierung nutzen. Während das VT über kognitive Prozesse arbeitet, misst die PPG periphere physiologische Reaktionen. Das fMRT erfasst vermutlich eine Vielzahl von Prozessen: Möglicherweise von der sexuellen Reizverarbeitung bis hin zur Vorbereitung motorischer Reaktionen. Diese Prozesse unterliegen unterschiedlichen Störeinflüssen und weisen – wie oben diskutiert – unterschiedliche idealtypische Verläufe auf.

Die korrelativen Befunde lassen sich nur schwer mit denen aus vorherigen Arbeiten in Einklang bringen. Während Letourneau (2002) in der Kategorie „female adolescent“ keine signifikanten Korrelationen zwischen VT und PPG nachweisen konnte, habe ich in dieser Kategorie robusteste Korrelationen gefunden (siehe Anhang 8). Letourneau (2002) fand dafür hohe Korrelationen bei vorpubertären Mädchen und Jungen, was ich nur partiell replizieren konnte. Die Stichprobe in der Studie von Letourneau bestand fast ausschließlich aus Personen, die sexuellen Kindesmissbrauch begangen hatten. Der Anteil der pädo-, hebe- oder teleiophilen Probanden war nicht bekannt. Dies könnte – z.B. bei einer sehr geringen Anzahl von teleiophilen Probanden – zu einer sehr geringen Varianz innerhalb einiger Stimulus-Kategorien geführt haben, was die Ergebnisse verfälschen kann. Zudem könnten die Probanden hoch motiviert gewesen sein, die Messergebnisse zu verfälschen, was ebenfalls die Unterschiede in

den Ergebnissen erklären könnte – vor allem wenn dies bei einem Verfahren besser möglich war als bei dem anderen. Im Gegensatz zu der Studie von Mackaronis et al. (2014) sind die un- sowie z -standardisierten VT Indizes etwa im gleichen Maße mit den anderen Indizes korreliert. Mackaronis et al. (2014) fanden zudem überhaupt keine signifikanten Zusammenhänge zwischen den beiden Verfahren innerhalb der einzelnen Stimulus-Kategorien oder zwischen den Altersindizes der beiden Verfahren. Allerdings erzielten sie auch in anderen Bereichen ungewöhnliche Ergebnisse. So errechneten die Autoren z.B. teils negative Korrelationen zwischen dem SSPI und der PPG. Zum Teil mag dies an der Stichprobe dieser Studie (jugendliche Sexualstraftäter) und der geringen Stichprobengröße ($N = 16$) liegen.

Die Tatsache, dass die Indizes der drei in dieser Arbeit verwendeten Verfahren miteinander korreliert sind, lässt vermuten, dass sie auf demselben Konstrukt laden. Zusammen mit den Ausführungen zur theoretischen Einbettung der Konstrukte anhand des Modells der sexuellen Reaktion nach Singer (Singer, 1984) legt dies nahe, dass sie verschiedene Aspekte der sexuellen Orientierung erfassen. Eine Kombination der Verfahren sollte somit die Klassifikationsgenauigkeit erhöhen (Epstein, 1983), was die Berechnungen bezüglich der inkrementellen Validität bestätigen. Eine Kombination mehrerer Indizes steigerte die Vorhersagegenauigkeit. Dieser Anstieg erwies sich allerdings nur als signifikant, wenn man alle drei Verfahren miteinander kombinierte und mit dem VT (auf Grundlage der z -transformierten Werte) oder dem fMRT verglich. Diese Ergebnisse stehen im Einklang mit der Studie von Banse et al. (2010).

Zur Berechnung der inkrementellen Validität habe ich die Indizes der drei Verfahren gleichzeitig in die Diskriminanzanalyse aufgenommen. Ich hätte ebenfalls die Verfahren getrennt auswerten und nur in den Fällen, in denen zwei oder drei Verfahren zu demselben Ergebnis kommen, eine Aussage treffen können. In 70.8 % der Fälle kamen alle drei Verfahren zu demselben Ergebnis. Bezogen auf die daraus resultierende neue Grundgesamtheit von 51 Probanden, lagen alle drei Instrumente nur bei 1.9 % Probanden ($n = 1$) übereinstimmend falsch. In über 98 % wäre also eine korrekte Gruppenzuordnung erfolgt; bezogen auf die ursprüngliche Grundgesamtheit aller Probanden, wäre dann aber bei 29.2 % der Probanden aufgrund fehlender Übereinstimmung zwischen den Verfahren überhaupt keine Aussage möglich, weshalb ich die Kombination mehrerer Indizes in einer Diskriminanzanalyse für die elegantere Methode halte.

4.7 Problematische Fälle

Bevor ich ein abschließendes Fazit ziehe, möchte ich auf einige problematische Fälle zu sprechen kommen, die – selbst bei den guten Ergebnissen der vorliegenden Studie – zu einer Unterschätzung der Validität der Verfahren geführt haben könnten. Bei dem besten Modell (auf Grundlage der z -transformierten VT-Daten sowie der PPG- und fMRT-Indizes) wurden nur 2 pädophile und ein teleiophiler Proband falsch klassifiziert. Bei einem pädophilen Probanden handelte es sich um den Probanden K201. Dieser könnte – wie oben bereits beschrieben – eventuell aufgrund atropher Prozesse im Gehirn im fMRT falsch klassifiziert worden sein. Dies würde allerdings nicht erklären, warum er auch in den beiden anderen Verfahren jeweils als teleiophil identifiziert worden ist. Es handelte sich bei dieser Person um einen Probanden, der mehrfachen sexuellen Kindesmissbrauch begangen hatte und überwiegend aufgrund der Anzahl der Taten in die Gruppe der pädophilen bzw. hebephilen Probanden aufgenommen wurde. Allerdings fanden alle Missbrauchshandlungen in einem häuslichen (nicht inzestuösen) Kontext statt und der Proband bestritt, Kinder sexuell attraktiv zu finden. Es könnte also auch sein, dass es sich hierbei um einen Probanden handelt, den wir fälschlicherweise der Gruppe der pädophilen Probanden zugewiesen haben. Andererseits könnte er die Messungen auch bewusst beeinflusst haben, um nicht als pädophil erkannt zu werden. Abschließend lässt sich dies nicht klären, könnte aber unter Umständen zu einer Unterschätzung der Validitätswerte geführt haben. Der zweite falsch-negative Proband wies in der PPG eine starke Reaktion bei den intrapuberalen Mädchen auf, was auf Basis der PPG zu der Diagnose einer Pädophobie bzw. Hebephilie geführt hat. In den anderen Verfahren zeigte er aber Reaktionen, die ganz klar zu denen der teleiophilen Probanden passten. Im Rahmen seiner Selbstauskünfte gab dieser Proband eine nicht-exklusive Heterohebephilie an. Bei diesem Probanden stimmten also nicht alle Verfahren überein, was Messfehler einzelner Verfahren nahe legt (vermutlich trotz kooperativer Teilnahme).

Zudem erwies sich die Klassifikation der nicht-pädophilen Missbrauchstäter und der hebephilen Probanden mittels fMRT als problematisch. Die anderen Verfahren (VT und PPG) erwiesen sich bei diesen Subgruppen allerdings als robust, sodass dies eher auf eine Schwäche der fMRT-Diagnostik bei diesen Gruppen hindeutet. Mögliche Gründe hierfür wurden bereits unter 4.3 diskutiert. Die Interpretation kann allerdings aufgrund der geringen Anzahl von Probanden innerhalb dieser Subgruppen ($n = 3$ bzw. $n = 17$) nur unter Vorbehalt geschehen. Die PPG hingegen wies bei den Hebephilen eine gute Klassifikationsleistung auf, erreichte dafür aber eine geringere Trefferquote bei den pädophilen Probanden. Dieser Interaktionseffekt

kann ebenfalls für die inkrementelle Validität bei der Kombination mehrerer Verfahren verantwortlich sein.

4.8 Verwendung objektiver Verfahren in der Therapie und Prognostik

Die guten Ergebnisse der vorliegenden Arbeit, die teils sogar höhere Klassifikationsleistungen nahe legen als bisherige Studien, sprechen prinzipiell für die Verwendung objektiver Verfahren zur Messung der sexuellen Orientierung. Diese Verfahren eignen sich – wie bereits in der Einleitung verdeutlicht – unter anderem zur Unterstützung der Prognosestellung bei Sexualstraftätern. So können in einem etablierten Instrument zur Risikoeinschätzung bei Sexualstraftätern neben der Diagnose einer Pädophilie nach DSM auch der SSPI Wert oder Testergebnisse aus der PPG zur Einschätzung genutzt werden (Quinsey, Harris, Rice & Cormier, 2006). Auch in der therapeutischen Arbeit unterstützen diese Instrumente die Therapieplanung. Die sexuelle Orientierung kann so verdeutlicht, dem Patienten rückgemeldet und mit ihm diskutiert werden.

Die mäßigen Korrelationen zwischen den Indizes der einzelnen Verfahren deuten dabei darauf hin, dass diese verschiedene Aspekte eines umfassenden Konstruktes erfassen. Durch die Verwendung mehrerer Verfahren lässt sich die diagnostische Sicherheit erhöhen. Sofern möglich, spricht dies daher für die Kombination verschiedener Verfahren. Bei einer vorhandenen Täuschungsabsicht des Patienten sollte es zudem leichter fallen, diese mithilfe verschiedener Techniken aufzudecken.

4.9 Verwendung objektiver Verfahren als Screening-Instrumente

An anderer Stelle wird darüber hinaus bereits die Verwendung objektiver Verfahren als Screening-Instrument empfohlen (Abel et al., 1994). Auf der Homepage des von Gene Abel entworfenen Instrumentes heißt es z.B.: „by identifying men and women who are most likely to cross sexual boundaries with children and teens, it helps to reduce adult on child sexual misconduct [...]” (http://diana.public.abelscreen.com/?page_id=371; abgerufen am 05.03.2015). Beim Einsatz der oben genannten Verfahren als Screening-Instrument ist nicht nur die Sensitivität und Spezifität der Verfahren entscheidend (Abel et al., 1994). Wenn ermittelt werden soll, wie wahrscheinlich eine pädophile oder hebephile Orientierung bei einem Probanden ist, nachdem er in einem Test als pädo- bzw. hebephil identifiziert worden ist, muss dies mathematisch anhand einer bedingten Wahrscheinlichkeit geschehen. Hierfür wird

wiederum eine Schätzung der Basiswahrscheinlichkeit dieser Orientierung in der Population benötigt, wozu es noch keine verlässlichen Daten gibt. Aktuelle Studien deuten auf eine Prävalenz der Pädophilie in der männlichen Bevölkerung von ca. 1 % bis 5 % hin (Ahlers et al., 2011; Dombert et al., 2015; Santtila et al., 2015; Santtila et al., 2010; Seto, 2007; Wurtele, Simons & Moreno, 2014). Bei Frauen liegt die Prävalenz vermutlich deutlich darunter. Der positive Vorhersagewert gibt den Anteil der richtig positiven (rp) Ergebnisse an der Gesamtheit der positiven (p) Ergebnisse an:

$$P(rp/p) = \frac{P(rp)}{P(rp)+P(fp)}$$

Bei einer Spezifität von 97.4 Prozent, einer Sensitivität von 91.2 % (Werte für die Kombination aller drei objektiver Verfahren der vorliegenden Arbeit) und einer angenommenen Basiswahrscheinlichkeit von 1 % ergibt sich somit:

$$P(rp/p) = \frac{.912 * .01}{.912 * .01 + .026 * .99} = \frac{.00912}{.00912 + .02574} = .2616$$

Das heißt, dass selbst bei einem Test mit hoher Spezifität und Sensitivität ein Populationsscreening sehr fehlerbehaftet ist, wenn die Basiswahrscheinlichkeit des aufzudeckenden Merkmals gering ist. Tatsächlich würden von 100 Personen, die durch solch einen Screening-Test als pädophil diagnostiziert wurden, nur etwa 26 Personen tatsächlich pädophil sein; in 74 % der Fälle würden wir uns bei einem positiven Ergebnis somit von dem Test in die Irre führen lassen, was für die Betroffenen mit schwerwiegenden Konsequenzen (Stigmatisierung, Berufsverbot, Inhaftierung o.ä.) einhergehen würde. Der positive Vorhersagewert steigt in Populationen an, in denen die Basiswahrscheinlichkeit einer pädophilen Orientierung höher ist (wie dies z.B. bei Sexualstraftätern der Fall ist). Die Schätzungen der Kennwerte für Spezifität und Sensitivität beruhen allerdings auf vorherigen Untersuchungen. Die theoretischen Kennwerte sind niemals bekannt. Es können somit immer erhebliche Fehler begangen werden, selbst wenn Studien eine Spezifität von 100 % nahelegen, da der tatsächliche Wert von diesem geschätzten Wert abweichen kann. Die Schätzung der Spezifität und Sensitivität der vorliegenden Arbeit basieren zudem auf einer Stichprobe, die überwiegend aus Personen bestand, die offen ihre sexuelle Orientierung einräumten. Die Ergebnisse lassen sich nicht ohne weiteres auf eine Population generalisieren, die zumindest partiell aus Non-Admittern besteht (Blanchard et al., 2001; Blanchard et al., 2007). Da zudem die Basiswahrscheinlichkeit pädophiler Orientierungen auch in bestimmten Subpopulationen (z.B. Lehrern oder Erziehern) noch vollkommen unbekannt ist, halte ich ein Screening entsprechender Berufsgruppen für äußerst problematisch.

Durch wiederholte Messungen ließen sich diese messtheoretischen Beschränkungen ggfs. minimieren. Der positiv prädiktive Wert beträgt im obigen Beispiel bei zweimaliger Testung 92.6 %, was allerdings auch zu einer Verringerung der korrekt positiven Testergebnisse führen würde, da nur bei zweimaliger positiver Testung eine zuverlässige Aussage möglich wäre. Zudem verdoppelt dieses Vorgehen – im Falle eines erstmalig positiven Tests – die Kosten der Testung.

Hinzu kommt, dass die vorgestellten Verfahren für eine Diagnose pädo-hebephiler Orientierungen konzipiert sind und nicht für die Vorhersage von kriminellem Verhalten. Wie viele pädo- bzw. hebephile Männer sexuellen Kindesmissbrauch begehen, muss anhand repräsentativer Befragungen geklärt werden. Entsprechende Daten liegen nicht vor, sodass unklar ist, wie viele pädo-hebephile Männer sexuellen Missbrauch begehen. Seto et al. (2015) konnte zeigen, dass die Nutzung kinderpornographischen Materials in einer großen Studie mit sexuellem Missbrauch assoziiert ist. Allerdings ist der Anteil pädophiler Probanden an der Stichprobe unklar und die Nutzung kinderpornographischen Materials war ebenfalls mit Indikatoren für antisoziales Verhalten assoziiert. Vogt (2006) lieferte Daten zur Frage, wie hoch der Anteil der Probanden, die sexuellen Kindemissbrauch begehen, in der Population der pädo-hebephilen Männer ist. Die Daten legen die Schlussfolgerung nahe, dass über die Lebensspanne jeder pädo- bzw. hebephile Mann pädosexuelle Kontakte zu Kindern eingeht. Der Autor befragte allerdings fast ausschließlich homo- bzw. bisexuelle Pädophile, welche in Bezug auf sexuellen Missbrauch eine höhere Rückfallwahrscheinlichkeit aufweisen als heterosexuelle Pädophile (Seto et al., 2004). Es erscheint daher auch naheliegend zu sein, dass homopädophile Männer mit einer höheren Wahrscheinlichkeit sexuellen Kindesmissbrauch begehen als heteropädophile Männer. Zudem führte Vogt nicht auf, was unter pädosexuellen Kontakten subsumiert worden ist, sodass unklar bleibt, ob es sich dabei wirklich um strafrechtlich relevante und/oder für das Kind schädliche Erfahrungen gehandelt hat. Denkbar wäre auch, dass die Teilnehmer diese Frage aufgrund freundschaftlicher („platonischer“) Kontakten zu Kindern bejahten. Die Ergebnisse einer groß angelegten Onlineumfrage von über 8700 deutschen Männern legen nahe, dass nicht jeder pädophile Mann sexuellen Kindesmissbrauch begeht (Dombert et al., 2015). Gut die Hälfte der etwa 500 Probanden, die sexuelle Fantasien in Bezug auf Kinder angaben, sagte, dass sie in der Vergangenheit weder Kinderpornographie genutzt noch sexuellen Kindesmissbrauch begangen hat. Williams, Cooper, Howell, Yuille und Paulhus (2008) gelang es nicht, signifikanten Korrelationen zwischen pädophilen Fantasien und Taten in zwei studentischen Populationen zu berechnen. Zudem war in ihrer Studie der Zusammenhang zwischen (anderen) devianten Fantasien und Verhalten durch das Vorliegen

von selbst-berichteten psychopathischen Persönlichkeitsmerkmalen mediiert. Da mittels objektiver Verfahren die sexuelle Orientierung – nicht aber potentiell straffälliges Verhalten – diagnostiziert werden kann, halte ich ein Screening auch unabhängig von der Genauigkeit der verwendeten Verfahren für problematisch.

4.10 Stigmatisierung

Dass die beiden Konzepte der sexuellen Orientierung und des sexuellen Verhaltens sowohl in der Fachliteratur (siehe z.B. Abel et al., 1994) als auch in der Allgemeinbevölkerung nicht ausreichend differenziert betrachtet werden, verschärft die Problematik weiterhin. Pädophile leiden unter systematischer Stigmatisierung (Jahnke & Hoyer, 2013; Jahnke, Schmidt, Geradt & Hoyer, under review; Kramer, 2011), selbst wenn sie keine sexuellen Übergriffe begehen bzw. begangen haben (für ein erschreckendes Beispiel siehe Anhang 9). So wurde erst jüngst ein pädophiler Erzieher, der sich im Rahmen einer Therapie auf seiner Arbeit outete, entlassen. Ein auf diesem Gebiet renommierter Therapeut verurteilte die Konsequenz, die der Arbeitgeber des Mannes gezogen hat (<http://www.sueddeutsche.de/muenchen/professor-paedophile-offenbaren-sich-nur-sehr-selten-1.2487471-2>; abgerufen am 27.05.2015). Im Gegensatz zu anderen Bereich, in denen Stigmatisierung untersucht wurde, ist das Ausmaß der Stigmatisierung mit Maßen der sozialen Erwünschtheit korreliert (Imhoff, 2015). Menschen gehen also wahrscheinlich davon aus, dass es angemessen und erwünscht ist, Pädophile zu stigmatisieren. Die Ablehnung von Personen mit pädophilen Interessen steigt weiterhin an, wenn man die Orientierung als krankhaft bezeichnet und ein entsprechendes Label verwendet (Imhoff, 2015).

Diese Stigmatisierung könnte den Umgang mit der eigenen Orientierung weiter erschweren. Im Kontext von psychischen Erkrankungen wurde bereits gezeigt, dass Stigmatisierung den Heilungsprozess behindert und einen erheblichen Vulnerabilitätsfaktor darstellt (Link, Struening, Neese-Todd, Asmussen & Phelan, 2001). Jahnke (2015) vermutet zudem, dass die Stigmatisierung pädophiler Männer die Wahrscheinlichkeit, dass diese Männer benötigte Therapieangebote aufsuchen, reduziert. Dies konnte durch die Daten zwar nicht gestützt werden. Allerdings gab ein Großteil der Pädophilen in ihrer Studie an, Therapeuten nicht vertrauen zu können. Hierdurch könnte die Wahrscheinlichkeit für Übergriffe zum Nachteil von Kindern erhöht werden. Sicherlich stellt Stigmatisierung aber zumindest eine ernsthafte Bedrohung für die Gesundheit und das Leben pädo-hebephiler Männer dar (Hatzenbuehler et al., 2014). Durch den Einsatz objektiver Verfahren im Sinne eines Screenings

könnte die massive Stigmatisierung weiter zunehmen, was eine Gefährdung für pädophile Männer und auch für potentielle Opfer darstellt. Allein der Verdacht, eine Person könne pädophil sein, kann zu erheblichen Ausgrenzungen bis zu Gewalttaten führen (Seto, 2008). Alles in allem halte ich den Einsatz von objektiven Verfahren im Sinne eines Screening ganzer Bevölkerungsgruppen daher für übereilt, unethisch und sogar potentiell schädlich. Ethische Überlegungen sind im Bereich der Pädophilie leider selten und sollten in Zukunft einen höheren Stellenwert bekommen. Eine ausgewogene Übersicht lieferte jüngst Moen (2015).

5 Zusammenfassung

Zusammenfassend deuten die Ergebnisse der vorliegenden Arbeit auf eine etwa gleich gute Klassifikationsleistung von VT, PPG und fMRT hin. Alle drei Verfahren erreichten gute bis sehr gute Validitätswerte und VT sowie PPG erwiesen sich darüber hinaus als reliabel. Sie erreichten dabei sogar Indizes, die über den aus der Literatur bekannten Werten lagen, während die fMRT leicht hinter den Ergebnissen der ersten Studie zurückblieb. Allerdings erreichte keines der Instrumente (abgesehen von der PPG nach dem Ausschluss von Low-Respondern) eine perfekte Vorhersageleistung. Zudem offenbaren die Instrumente weitere Schwächen wie potentielle Beeinflussung seitens des Probanden. Nichtsdestotrotz weisen auch traditionelle Messungen – im Rahmen der Diagnostik sexueller Orientierung also Selbst- und Fremdauskünfte sowie das Aktenstudium – Mängel auf. Hier ist z.B. die Gefahr gegeben, dass das gutachterliche Urteil auf relativ unsystematischen Daten und persönlichen Eindrücken beruht. Mit anderen Worten kann eine geringe Objektivität der Messung vorliegen. Hier können objektive Verfahren neue Aspekte und Informationen liefern. Sofern sich objektive Verfahren nicht als unzuverlässig erweisen, sehe ich deshalb keinen Grund, wertvolle Informationen bei gewichtigen Fragestellungen zu ignorieren (De Houwer et al., 2009). Deshalb halte ich die untersuchten Verfahren für wertvoll und sehe eine Reihe von Einsatzmöglichkeiten. Voraussetzung ist für mich dabei, dass der Einsatz durch geschultes und in diesem Bereich erfahrenes Personal erfolgt, welches sich nicht nur mit der Datenerhebung und -auswertung sondern auch der Interpretation und Integration der Befunde auskennt.

Bei Begutachtungsverfahren in Strafverfahren zu sexuellem Kindesmissbrauch kann man Studien zufolge etwa mit einer 50%igen Wahrscheinlichkeit von einer pädophilen Motivation (bei einer Erststraftat, Seto, 2007) ausgehen. Und auch im therapeutischen Setting stellen sich immer wieder Personen mit einer fraglich pädophilen Orientierung vor, die nicht ohne weiteres

allein anhand von Selbstauskünften abklärbar ist. Zur Unterstützung im diagnostischen Prozess halte ich die Verfahren für geeignet. Gerade die Kombination mehrerer Methoden kann die diagnostische Sicherheit erhöhen (Selbst- und Fremdanamnese, sowie indirekter Verfahren; Banse et al., 2010), indem verschiedene Aspekte der sexuellen Orientierung abgeklärt werden – nämlich Selbstkonzept, genitale Erregung, zentralnervöse Verarbeitung und Aufmerksamkeitsprozesse (Schmidt et al., 2014). Geschultes Personal kann mithilfe entsprechender Verfahren zusätzliche Informationen zur sexuellen Orientierung des Probanden erhalten. Diese dürfen aber nicht als Beweise für die Schuld bzw. Schuldfähigkeit oder eine fehlende berufliche Eignung (z.B. für die Arbeit mit Kindern) missinterpretiert werden. Sehr wohl können entsprechende Erkenntnisse aber für die Prognosebegutachtung genutzt werden.

Wie bereits erwähnt, lassen sich die berichteten Ergebnisse nicht ohne weiteres generalisieren. Dies liegt unter anderem daran, dass die Stichprobe der pädo-hebephilen Männer fast ausschließlich aus Admittern bestand. Es ist durchaus denkbar, dass sich manche Verfahren als robuster gegenüber willentlicher Beeinflussung durch den Probanden erweisen als andere. Sollte sich die fMRT-Diagnostik z.B. als robust gegenüber Verfälschungstendenzen erweisen, wäre dies ein erheblicher Vorteil gegenüber der PPG (Freund et al., 1979). Die Ergebnisse der vorliegenden Arbeit sind deshalb als vorläufig zu betrachten und müssen durch weitere Studien gestützt bzw. ergänzt werden. Dabei sollte explizit die Verfälschbarkeit der Messung durch experimentelle Manipulation oder durch die Verwendung von Non-Admittern getestet werden.

Zudem wurden bei pädo-hebephilen Probanden nur die Personen aufgenommen, die die stärkste Erregung in Bezug auf prä- bzw. intrapubere Kinder berichteten. Dies könnte die Generalisierbarkeit der Befunde weiter einschränken. Dombert et al. (2015) berichteten, dass in ihrer Stichprobe von 8700 Männern 482 (5.5 %) sexuelle Fantasien mit Kindern (jünger als 12 Jahre) angaben. Allerdings berichteten nur 12 Männer (.1 % der Gesamtstichprobe) mehr sexuelle Fantasien in Bezug auf Kinder im Vergleich zu Erwachsenen zu erleben. Von diesen 12 gaben 6 (.05 % der Gesamtstichprobe) einen Leidensdruck an. Dies entspricht etwa 1 % der Männer aus der Studie, die sexuelle Fantasien von Kindern berichten. Da ein Großteil der Probanden der vorliegenden Arbeit aus dem Präventionsprojekt Dunkelfeld stammt, setzt sich die Stichprobe vermutlich hauptsächlich aus diesem einem Prozentsatz der pädophilen Männer zusammen. Dies führt zu einer erheblichen Einschränkung der Repräsentativität der Stichprobe (in Bezug auf die Gesamtheit der Männer mit sexuellen Fantasien von Kindern).

Um das Verständnis von Pädophilie, sexuellen Fantasien mit Kindern und sexuellen Missbrauch zu vertiefen, wären weitere repräsentative Umfragen, wie sie von Ahlers et al.

(2011), Dombert et al. (2015) und Santtila et al. (2015) durchgeführt worden sind, notwendig. Forscher sollten in diesen Studien auf eine Trennung zwischen pädo- und hebephilen Fantasien achten und zusätzlich zu Selbstauskünften objektive Verfahren (VT, PPG, fMRT etc.) verwenden. Anhand dieser größeren und repräsentativen Stichprobe ließe sich herausfinden, wie gut objektive Verfahren in der Lage sind, exklusive und nicht-exklusive Altersorientierungen vorherzusagen und ob die Klassifikationsgenauigkeit von weiteren Faktoren abhängt (Leidensdruck, Geschlechtsorientierung etc.). Idealerweise könnte durch die Erfassung der strafrechtlichen Vorgeschichte der Probanden in Bezug auf sexuellen Kindesmissbrauch auch der Nutzen der objektiven Verfahren in Bezug auf eine Risikoprognose in einer repräsentativen Bevölkerungsstichprobe untersucht werden. So könnte gezeigt werden, ob forensische Untersuchungen auch auf die Allgemeinbevölkerung anwendbar sind (Hanson & Morton-Bourgon, 2005).

6 Literaturverzeichnis

- Abel, G. G., Blanchard, E. B., & Barlow, D. H. (1981). Measurement of sexual arousal in several paraphilias: the effects of stimulus modality, instructional set and stimulus content on the objective. *Behaviour Research and Therapy*, *19*, 25-33.
- Abel, G. G., Huffman, J., Warberg, B., & Holland, C. L. (1998). Visual reaction time and plethysmography as measures of sexual interest in child molesters. *Sexual Abuse: A Journal of Research and Treatment*, *10*, 81-95.
- Abel, G. G., Jordan, A., Hand, C. G., Holland, L. A., & Phipps, A. (2001). Classification models of child molesters utilizing the abel assessment for sexual interest. *Child Abuse & Neglect*, *25*, 703-718.
- Abel, G. G., Lawry, S. S., Karstrom, E., Osborn, C. A., & Gillespie, C. F. (1994). Screening tests for pedophilia. *Criminal Justice and Behavior*, *21*, 115-131.
- Ahlers, C. J., & Schaefer, G. A. (2010). Pädophilie, Pädosexualität und sexueller Kindesmissbrauch: Über die Notwendigkeit einer differenzierten Betrachtung. *BZgA-Forum*, *3*, 46-49.
- Ahlers, C. J., Schaefer, G. A., & Beier, K. M. (2005). Das Spektrum der Sexualstörungen und ihre Klassifizierbarkeit in DSM-IV und ICD-10. *Sexuologie*, *12*, 120-152.
- Ahlers, C. J., Schaefer, G. A., Mundt, I. A., Roll, S., Englert, H., Willich, S. N., & Beier, K. M. (2011). How unusual are the contents of paraphilias? Paraphilia-associated sexual arousal patterns in a community-based sample of men. *The Journal of Sexual Medicine*, *8*, 1362-1370.
- Ahlmeyer, S., Heil, P., McKee, B., & English, K. (2000). The impact of polygraphy on admissions of victims and offenses in adult sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, *12*, 123-138.
- Ashby, F. G. (2015). An Introduction to fMRI. In Forstmann, B. U. & Wagenmakers, E.-J. (Eds.), *An introduction to model-based cognitive neuroscience*. New York: Springer.
- American Psychological Association (2009). Report of the american psychological association's task force on appropriate therapeutic responses to sexual orientation. Washington, DC: American Psychological Association.
- American Psychiatric Association (2013). Diagnostic and statistical manual of mental disorders, fifth edition - DSM 5 (Vol. 5 ed.). Arlington, VA: American Psychiatric Publishing.
- Bailey, J. M. (2009). What is sexual orientation and do women have one? In D. A. Hope (Ed.), *Contemporary perspectives on lesbian, gay, and bisexual identities*. (Vol. 54, pp. 43-63). New York: Springer.
- Bailey, J. M. (2014). A failure to demonstrate changes in sexual interest in pedophilic men: Comment on Muller et al. (2014). *Archives of Sexual Behavior*, *44*, 249-252.
- Bancroft, J. J. H. C., & Pullan, B. P. (1966). A simple transducer for measuring penile erections with comments on its use in the treatment of sexual disorders. *Behaviour Research and Therapy*, *4*, 239-241.
- Banse, R., Schmidt, A. F., & Clabour, J. (2010). Indirect measures of sexual interest in child sex offenders: A multimethod approach. *Criminal Justice and Behavior*, *37*, 319-335.
- Barbaree, H. E., Baxter, D. J., & Marshall, W. L. (1989). Brief research report: the reliability of the rape index in a sample of rapists and nonrapists. *Violence and Victims*, *4*, 299-306.
- Barbaree, H. E., Marshall, W. L., Yates, E., & Lightfoot, L. O. (1983). Alcohol intoxication and deviant sexual arousal in male social drinkers. *Behaviour Research and Therapy*, *21*, 365-373.

- Barlow, D. H., Becker, R., Leitenberg, H., & Agras, W. S. (1970). A mechanical strain gauge for recording penile circumference change. *Journal of Applied Behavior Analysis*, *3*, 73-76.
- Beccuti, G., & Ghizzoni, L. (2000). Normal and Abnormal Puberty. In De Groot, L. J. et al. (Eds.), *Endotext*. South Dartmouth (MA).
- Becker, J. V., Hunter, J. A., Goodwin, D., Kaplan, M. S., & Martinez, D. (1992). Test-retest reliability of audio-taped phallometric stimuli with adolescent sexual offenders *Sexual Abuse: A Journal of Research and Therapy*, *5*, 45-51.
- Beckstead, A. L. (2012). Can we change sexual orientation? *Archives of Sexual Behavior*, *41*, 121-134.
- Beier, K. M. (1998). Differential typology and prognosis for dissexual behavior--a follow-up study of previously expert-appraised child molesters. *International Journal of Legal Medicine*, *111*, 133-141.
- Beier, K. M., Ahlers, C. J., Goecker, D., Neutze, J., Mundt, I. A., Hupp, E., & Schaefer, G. A. (2009). Can pedophiles be reached for primary prevention of child sexual abuse? First results of the Berlin Prevention Project Dunkelfeld (PPD). *The Journal of Forensic Psychiatry & Psychology*, *20*, 851-867.
- Beier, K. M., Amelung, T., Kuhle, L., Grundmann, D., Scherner, G., & Neutze, J. (2013). Hebephilia as a sexual disorder. *Fortschritte der Neurologie und Psychiatrie*, *81*, 128-137.
- Beier, K. M., Bosinski, H., & Loewit, K. (2005). *Sexualmedizin* (2. Auflage). München: Elsevier.
- Blanchard, R., & Barbaree, H. E. (2005). The strength of sexual arousal as a function of the age of the sex offender: comparisons among pedophiles, hebephiles, and teleiophiles. *Sexual Abuse: A Journal of Research and Treatment*, *17*, 441-456.
- Blanchard, R., Klassen, P., Dickey, R., Kuban, M. E., & Blak, T. (2001). Sensitivity and specificity of the phallometric test for pedophilia in nonadmitting sex offenders. *Psychological Assessment*, *13*, 118-126.
- Blanchard, R., Kolla, N. J., Cantor, J. M., Klassen, P. E., Dickey, R., Kuban, M. E., & Blak, T. (2007). IQ, handedness, and pedophilia in adult male patients stratified by referral source. *Sexual Abuse: A Journal of Research and Therapy*, *19*, 285-309.
- Blanchard, R., Kuban, M. E., Blak, T., Cantor, J. M., Klassen, P. E., & Dickey, R. (2009). Absolute versus relative ascertainment of pedophilia in men. *Sexual Abuse: A Journal of Research and Therapy*, *21*, 431-441.
- Blanchard, R., Kuban, M. E., Blak, T., Klassen, P. E., Dickey, R., & Cantor, J. M. (2010). Sexual attraction to others: A comparison of two models of alloerotic responding in men. *Archives of Sexual Behavior*, *41*, 13-29.
- Blanchard, R., Lykins, A. D., Wherrett, D., Kuban, M. E., Cantor, J. M., Blak, T., Dickey, R., & Klassen, P. E. (2009). Pedophilia, hebephilia, and the DSM-V. *Arch Sex Behav*, *38*, 335-350.
- Boetticher, A., Nedopil, N., Bosinski, H. A. G., & Saß, H. (2007). Mindestanforderungen für Schuldfähigkeitsgutachten. *Forensische Psychiatrie, Psychologie, Kriminologie*, *1*, 3-9.
- Bourke, A. B., & Gormley, M. J. (2012). Comparing a pictorial stroop task to viewing time measures of sexual interest. *Sexual Abuse: A Journal of Research and Treatment*, *24*, 479-500.
- Brown, M. R. et al. (2015). Neural correlates of high-risk behavior tendencies and impulsivity in an emotional Go/NoGo fMRI task. *Frontiers in Systems Neuroscience*, *9*, 24.
- Bundeskriminalamt (2012). Polizeiliche Kriminalstatistik (PKS). Abgerufen am 05.08.2014 unter:

http://www.bka.de/DE/Publikationen/PolizeilicheKriminalstatistik/pks__node.html?__nnn=true.

- Buxton, R. B. (2009). Introduction to functional magnetic resonance imaging: Principles and techniques (Vol. 2.). Cambridge: Cambridge University Press.
- Cantor, J. M. (2015). Purported changes in pedophilia as statistical artefacts: Comment on Muller et Al. (2014). *Archives of Sexual Behavior*, 44, 253-254.
- Chaplin, T. C., Rice, M. E., & Harris, G. T. (1995). Salient victim suffering and the sexual responses of child molesters. *Journal of Consulting and Clinical Psychology*, 63, 249-255.
- Chau, W., & McIntosh, A. R. (2005). The Talairach coordinate of a point in the MNI space: how to interpret it. *Neuroimage*, 25, 408-416.
- Chivers, M. L., Rieger, G., Latty, E., & Bailey, J. M. (2004). A sex difference in the specificity of sexual arousal. *Psychological Science*, 15, 736-744.
- Chow, E. W. C., & Choy, A. L. (2002). Clinical characteristics and treatment response to SSRI in a female pedophile. *Archives of Sexual Behaviour*, 31, 211-215.
- Cohen, J. (1988). *Power Analysis for the behavioural sciences*. Hillsdale, New York: Erlbaum.
- Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science*, 1, 98-101.
- Cordes, D., Haughton, V., Carew, J. D., Arfanakis, K., & Maravilla, K. (2002). Hierarchical clustering to measure connectivity in fMRI resting-state data. *Magnetic Resonance Imaging*, 20, 305-317.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Damsa, C., Kosel, M., & Moussally, J. (2009). Current status of brain imaging in anxiety disorders. *Current Opinion in Psychiatry*, 22, 96-110.
- Day, D., Miner, M., Sturgeon, V., & Murphy, J. (1989). Assessment of sexual arousal by means of physiological and self-report measure. In Laws, D. R. (Ed.), *Relapse prevention with sex offenders* (pp. 115-123). New York: Guilford.
- De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin*, 135, 347-368.
- Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computing and Applications*, 10, 1895-1923.
- Dombert, B. et al. (2013). The virtual people set: Developing computer-generated stimuli for the assessment of pedophilic sexual interest. *Sexual Abuse: A Journal of Research and Treatment*, 25, 557-82.
- Dombert, B., Schmidt, A. F., Banse, R., Briken, P., Hoyer, J., Neutze, J., & Osterheider, M. (2015). How common is males' self-reported sexual interest in prepubescent children? *The Journal of Sex Research*, 4, 1-10.
- Duchesne, S., Caroli, A., Geroldi, C., Barillot, C., Frisoni, G. B., & Collins, D. L. (2008). MRI-based automated computer classification of probable AD versus normal controls. *IEEE Transactions on Medical Imaging*, 27, 509-520.
- Earls, C., & Marshall, W. (1983). The current state of technology in the laboratory assessment of sexual arousal patterns. In Greer, J. & Stuart, I. (Eds.), *The sexual aggressor: Current perspectives on treatment*. Van Nostrand: Reinhold.
- Earls, C. M., Morales, A., & Marshall, W. L. (1988). Penile sufficiency: An operational definition. *The Journal of Urology*, 139, 536-538.
- Earls, C. M., Quinsey, V. L., & Castonguay, L. G. (1976). A Comparison of two methods of scoring the penile circumference response: Magnitude and area. *Behavior Therapy*, 7, 702-703.

- Earls, C. M., Quinsey, V. L., & Castonguay, L. G. (1987). A comparison of three methods of scoring penile circumference changes. *Archives of Sexual Behaviour, 16*, 493-500.
- Epstein, S. (1983). Aggregation and beyond: Some basic issues on the prediction of behaviour. *Journal of Personality, 51*, 360-392.
- Farkas, G. M., Evans, I. M., Sine, L. F., Eifert, G., Wittlieb, E., & Vogelmann-Sine, S. (1979). Reliability and validity of the mercury-in-rubber strain gauge measure of penile circumference. *Behavior Therapy, 10*, 551-561.
- Fedora, O., Reddon, J. R., Morrison, J. W., Fedora, S. K., Pascoe, H., & Yeudall, L. T. (1992). Sadism and other paraphilias in normal controls and aggressive and nonaggressive sex offenders. *Archives of Sexual Behavior, 21*, 1-15.
- Fedoroff, J. P., Curry, S., Muller, K., Ranger, R., Briken, P., & Bradford, J. (2015). Evidence that arousal to pedophilic stimuli can change: Response to bailey, cantor, and lalumiere. *Archives of Sexual Behavior, 44*, 259-263.
- Fedoroff, J. P., Fishell, A., & Fedoroff, B. (1999). A case series of women evaluated for paraphilic sexual disorders. *Canadian Journal of Human Sexuality, 8*, 127-141.
- Fernandez, Y. (2000). Phallometric testing with sexual offenders against female victims: Limits to its value. Unpublished doctoral thesis, Queen's University Kingston, Ontario (Canada). Abstract retrieved May 2013 from: <https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=195448>.
- Firestone, P., Bradford, J. M., Greenberg, D. M., & Nunes, K. L. (2000). Differentiation of homicidal child molesters, nonhomicidal child molesters, and nonoffenders by phallometry. *American Journal of Psychiatry, 157*, 1847-1850.
- Flak, V. E., Beech, A. R., & Fisher, D. (2007). Forensic assessment of deviant sexual interests: The current position. *Issues in Forensic Psychology, 6*, 70-83.
- Frenzel, R. R., & Lang, R. A. (1989). Identifying sexual preferences in intrafamilial and extrafamilial child sexual abusers. *Annals of Sex Research, 2*, 255-275.
- Freund, K. (1957). Diagnostika homosexuality u muza [Diagnosis of male homosexuality.]. *Cesk Psychiatry, 53*, 382-394.
- Freund, K. (1961). Laboratory differential diagnosis of homo- and heterosexuality--an experiment with faking. *Review of Czechoslovak medicine, 7*, 20-31.
- Freund, K. (1967). Erotic preference in pedophilia. *Behavioural Research and Therapy, 5*, 339-348.
- Freund, K., & Blanchard, R. (1989). Phallometric diagnosis of pedophilia. *Journal of Consulting and Clinical Psychology, 57*, 100-105.
- Freund, K., Chan, S., & Coulthard, R. (1979). Phallometric diagnosis with nonadmitters. *Behaviour Research and Therapy, 17*, 451-457.
- Freund, K., Langevin, R., & Barlow, D. (1974). Comparison of two penile measures of erotic arousal. *Behaviour Research and Therapy, 12*, 355-359.
- Freund, K., & Watson, R. (1991). Assessment of sensitivity and specificity of a phallometric test: An update of phallometric diagnosis of pedophilia. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3*, 254-260.
- Freund, K., Watson, R., & Rienzo, D. (1988). Signs of feigning in the phallometric test. *Behaviour Research and Therapy, 26*, 105-112.
- Freund, K., & Watson, R. J. (1992). The proportions of heterosexual and homosexual pedophiles among sex offenders against children: An exploratory study. *Journal of Sexual and Marital Therapy, 18*, 34-43.
- Fu, C. H., Mourao-Miranda, J., Costafreda, S. G., Khanna, A., Marquand, A. F., Williams, S. C., & Brammer, M. J. (2008). Pattern classification of sad facial processing: Toward the development of neurobiological markers in depression. *Biological Psychiatry, 63*, 656-662.

- Geer, J. H., & Bellard, H. S. (1996). Sexual content induced delays in unprimed lexical decisions: Gender and context effects. *Archives of Sexual Behavior*, 25, 379-395.
- Geer, J. H., & Fuhr, R. (1976). Cognitive factors in sexual arousal: The role of distraction. *Journal of Consulting and Clinical Psychology*, 44, 238-243.
- Gerwinn, H. et al. (2015). The (in)consistency of changes in brain macrostructure in male paedophiles: A combined T1-weighted and diffusion tensor imaging study. *Journal of Psychiatric Research*, 68, 246-253.
- Glueck, B. C., Jr. (1955). Final report: Research project for the study and treatment of persons convicted of crimes involving sexual aberrations. June 1952 to June 1955. New York: New York State Department of Mental Hygiene.
- Golde, J. A., Strassberg, D. S., & Turner, C. M. (2000). Psychophysiological assessment of erectile response and its suppression as a function of stimulus media and previous experience with plethysmography. *Journal of Sex Research*, 37, 53-59.
- Goodkind, M. et al. (2015). Identification of a common neurobiological substrate for mental illness. *JAMA Psychiatry*, 72, 305-15.
- Gray, S. R., & Plaud, J. J. (2005). A comparison of the abel assessment for sexual interest and penile plethysmography in an outpatient sample of sexual offenders. *Journal of Sexual Offender Civil Commitment: Science and the Law*, 1, 1-10.
- Gress, C. L. Z. (2005). Viewing time measures and sexual interest: another piece of the puzzle. *Journal of Sexual Aggression*, 11, 117-125.
- Groenewold, N. A., Opmeer, E. M., de Jonge, P., Aleman, A., & Costafreda, S. G. (2013). Emotional valence modulates brain functional abnormalities in depression: Evidence from a meta-analysis of fMRI studies. *Neuroscience & Biobehavioral Reviews*, 37, 152-163.
- Haldeman, D. C. (1994). The practice and ethics of sexual orientation conversion therapy. *Journal of Consulting and Clinical Psychology*, 62, 221-227.
- Hale, V. E., & Strassberg, D. S. (1990). The role of anxiety on sexual arousal. *Archives of Sexual Behavior*, 19, 569-581.
- Hall, G. C. (1991). Sexual arousal as a function of physiological and cognitive variables in a sexual offender population. *Archives of Sexual Behavior*, 20, 359-369.
- Hall, G. C., Proctor, W. C., & Nelson, G. M. (1988). Validity of physiological measures of pedophilic sexual arousal in a sexual offender population. *Journal of Consulting and Clinical Psychology*, 56, 118-122.
- Hall, G. C. N., Hirschman, R., & Oliver, L. L. (1995). Sexual arousal and arousability to pedophilic stimuli in a community sample of normal men. *Behavior Therapy*, 26, 681-694.
- Hall, R. C. W., & Hall, R. C. W. (2007). A profile of pedophilia: Definition, characteristics of offenders, recidivism, treatment outcomes, and forensic issues. *FOCUS Fundamentals of Life: Disorders of Sleep, Eating and Sex*, 7, 522-537.
- Hand, I., & Büttner-Westühal, H. (1991). Die Yale-Brown Obsessive Compulsive Scale (Y-BOCS). Ein halbstrukturiertes Interview zur Beurteilung des Schweregrades von Denk- und Handlungszwängen. *Verhaltenstherapie*, 1, 223-225.
- Hanson, R. K., & Bussiere, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, 66, 348-362.
- Hanson, R. K., & Morton-Bourgon, K. E. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology*, 73, 1154-1163.
- Harris, G. T., Rice, M. E., Chaplin, T. C., & Quinsey, V. L. (1999). Dissimulation in phallometric testing of rapists' sexual preferences. *Archives of Sexual Behaviour*, 28, 223-232.

- Harris, G. T., Rice, M. E., Quinsey, V. L., & Chaplin, T. C. (1996). Viewing time as a measure of sexual interest among child molesters and normal heterosexual men. *Behaviour Research and Therapy*, *34*, 389-394.
- Harris, G. T., Rice, M. E., Quinsey, V. L., Chaplin, T. C., & Earls, C. M. (1992). Maximizing the discriminant validity of phallometric assessment data. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *34*, 389-394.
- Hatzenbuehler, M. L., Bellatorre, A., Lee, Y., Finch, B. K., Muennig, P., & Fiscella, K. (2014). Structural stigma and all-cause mortality in sexual minority populations. *Social Science & Medicine*, *103*, 33-41.
- Heitzman, J., Lew-Starowicz, M., Pacholski, M., & Lew-Starowicz, Z. (2014). Children sexual abuse in Poland – study of 257 sexual offenders against minors. *Psychiatria Polska*, *48*, 105-120.
- Henson, D. E., & Rubin, H. B. (1971). Voluntary control of eroticism. *Journal of Applied Behavior Analysis*, *4*, 37-44.
- Howes, R. J. (1995). A survey of plethysmographic assessment in North America. *Sexual Abuse: A Journal of Research and Treatment*, *7*, 9-22.
- Howes, R. J. (2003). Circumferential change scores in phallometric assessment: Normative data. *Sexual Abuse: A Journal of Research and Treatment*, *15*, 365-375.
- Howes, R. J. (2009). Measurement of risk of sexual violence through phallometric testing. *Legal Medicine*, *11*, 368-369.
- Hunter, J. A. (1994). The relationship between phallometrically measured deviant sexual arousal and clinical characteristics in juvenile sexual offenders. *Behavioural Research and Therapy*, *32*, 533-538.
- Imhoff, R. (2012). *Kälter als gedacht - Ein kritischer Blick auf die Annahme affektiver Prozesse in indirekten Messverfahren*. Vortrag auf dem Kongress der deutschen Gesellschaft für Psychologie am 24.09.2012. Abstract abgerufen am 02 Februar 2015 unter http://www.dgpskongress.de/frontend/index.php?page_id=201&v=List&do=15&day=146&ses=122188.
- Imhoff, R. (2015). Punitive attitudes against pedophiles or persons with sexual interest in children: does the label matter? *Archives of Sexual Behavior*, *44*, 35-44.
- Imhoff, R., Schmidt, A. F., Nordsiek, U., Luzar, C., Young, A. W., & Banse, R. (2010). Viewing time effects revisited: Prolonged response latencies for sexually attractive targets under restricted task conditions. *Archives of Sexual Behavior*, *39*, 1275-1288.
- Imhoff, R., Schmidt, A. F., Weiss, S., Young, A. W., & Banse, R. (2012). Vicarious viewing time: Prolonged response latencies for sexually attractive targets as a function of task- or stimulus-specific processing. *Archives of Sexual Behavior*, *41*, 1389-401.
- Israel, E., & Strassberg, D. S. (2009). Viewing time as an objective measure of sexual interest in heterosexual men and women. *Archives of Sexual Behavior*, *38*, 551-558.
- Jahnke, S. (2015). *Understanding and Challenging Stigmatization of People with Pedophilia*. Dortal Thesis, Technische Universität Dresden, retrived September 2015 from: [http://www.qucosa.de/recherche/frontdoor/?tx_slubopus4frontend\[id\]=urn:nbn:de:bsz:14-qucosa-170785](http://www.qucosa.de/recherche/frontdoor/?tx_slubopus4frontend[id]=urn:nbn:de:bsz:14-qucosa-170785).
- Jahnke, S., & Hoyer, J. (2013). Stigmatization of people with pedophilia: A blind spot in stigma research. *International Journal of Sexual Health*, *25*, 169 – 184.
- Jahnke, S., Schmidt, A. F., Geradt, M., & Hoyer, J. (in press). Consequences of Stigma against People with Pedophilia. *Archives of Sexual Behavior*.
- Janssen, E., Vorst, H., Finn, P., & Bancroft, J. (2002). The sexual inhibition (SIS) and sexual excitation (SES) scales: Measuring sexual inhibition and excitation proneness in men. *The Journal of Sex Research*, *39*, 114-126.

- Julien, E., & Over, R. (1981). Male sexual arousal and the law of initial value. *Psychophysiology*, *18*, 709-711.
- Kalmus, E., & Beech, A. R. (2005). Forensic assessment of sexual interest: A review. *Aggression and Violent Behavior*, *10*, 193-217.
- Kingston, D. A., Firestone, P., Moulden, H. M., & Bradford, J. M. (2007). The utility of the diagnosis of pedophilia: A comparison of various classification procedures. *Archives of Sexual Behaviour*, *36*, 423-436.
- Klinitzke, G., Romppel, M., Hauser, W., Braehler, E., & Glaesmer, H. (2012). The German version of the childhood trauma questionnaire (CTQ): Psychometric characteristics in a representative sample of the general population. *Psychotherapie Psychosomatik Medizinische Psychologie*, *62*, 47-51.
- Kloppel, S. et al. (2008). Automatic classification of MR scans in Alzheimer's disease. *Brain*, *131*, 681-689.
- Kolla, N. J., Klassen, P. E., Kuban, M. E., Blak, T., & Blanchard, R. (2010). Double-blind, placebo-controlled trial of sildenafil in phallometric testing. *The Journal of the American Academy of Psychiatry and the Law*, *38*, 502-511.
- Krafft-Ebbing, R. (1924). *Psychopathia sexualis*. Stuttgart: Enke.
- Kramer, R. (2011). *The DSM and the Stigmatization of people who are attracted to minors*. Paper presented at the B4U-ACT meeting (August, 27th 2011) in Baltimore, Maryland. Slides retrieved March 12th 2015 from <http://slideplayer.com/slide/2989235/>.
- Kranz, F., & Ishai, A. (2006). Face perception is modulated by sexual preference. *Current Biology*, *16*, 63-68.
- Kryst, Ł., Kowal, M., Woronkiewicz, A., Sobiecki, J. A. N., & Cichocka, B. A. (2012). Secular changes in height, body weight, body mass index and pubertal development in male children and adolescents in Krakow, Poland. *Journal of biosocial science*, *44*, 495-507.
- Kuban, M., Barbaree, H. E., & Blanchard, R. (1999). A comparison of volume and circumference phallometry: Response magnitude and method agreement. *Archives of Sexual Behaviour*, *28*, 345-359.
- Lalumiere, M. L. (2015). The lability of pedophilic interests as measured by phallometry. *Archives of Sexual Behaviour*, *44*, 255-258.
- Lalumiere, M. L., & Earls, C. M. (1992). Voluntary control of penile responses as a function of stimulus duration and instructions. *Behavioral Assessment*, *14*, 121-132.
- Lalumière, M. L., & Harris, G. T. (1998). Common questions regarding the use of phallometric testing with sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, *10*, 227-237.
- Lang, P. J., Bradley, M. M., & Cuthbert, B. (1999). *International affective picture system (IAPS): Technical manual and affective ratings*. Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.
- Launay, G. (1999). The phallometric assessment of sex offenders: an update. *Criminal Justice and Behaviour*, *9*, 254-274.
- Lauterbur, P. C. (1973). Image formation by induced local interactions: Examples of employing nuclear magnetic resonance. *Nature* 190-191.
- Laws, D. R. (2009). Penile plethysmography: Strengths, limitations, innovations. In Thornton, D. & Laws, D. R. (Eds.), *Cognitive Approaches to the Assessment of Sexual Interest in Sexual Offenders*. Chichester: John Wiley & Sons.
- Laws, D. R., & Holmen, M. L. (1978). Sexual response faking by pedophiles. *Criminal Justice and Behavior*, *5*, 343-356.
- Laws, D. R., & Rubin, H. B. (1969). Instructional control of an autonomic sexual response. *Journal of Applied Behavior Analysis*, *2*, 93-99.

- Letourneau, E. J. (2002). A comparison of objective measures of sexual arousal and interest: Visual reaction time and penile plethysmography. *Sexual Abuse: A Journal of Research and Treatment, 14*, 207-223.
- LeVay, S., & Baldwin, J. (2009). *Human Sexuality* (Vol. 3). Sunderland, Massachusetts: Sinauer Associates Inc.
- Levenson, J. S. (2004). Reliability of sexually violent predator civil commitment criteria in Florida. *Law and Human Behavior, 28*, 357-368.
- Link, B. G., Struening, E. L., Neese-Todd, S., Asmussen, S., & Phelan, J. C. (2001). Stigma as a barrier to recovery: The consequences of stigma for the self-esteem of people with mental illnesses. *Psychiatric Services, 52*, 1621-1626.
- Lippa, R. A. (2013). Men and women with bisexual identities show bisexual patterns of sexual attraction to male and female "swimsuit models". *Archives of Sexual Behavior, 42*, 187-196.
- Lippa, R. A., Patterson, T. M., & Marelich, W. D. (2010). Looking at and longing for male and female "swimsuit models": Men are much more category specific than women. *Social Psychological and Personality Science, 1*, 238-245.
- Looman, J., Abracen, J., Maillet, G., & DiFazio, R. (1998). Phallometric nonresponding in sexual offenders. *Sexual Abuse: A Journal of Research and Therapy, 10*, 325-336.
- Lykins, A. D., Cantor, J. M., Kuban, M. E., Blak, T., Dickey, R., Klassen, P. E., & Blanchard, R. (2010). The relation between peak response magnitudes and agreement in diagnoses obtained from two different phallometric tests for pedophilia. *Sexual Abuse: A Journal of Research and Treatment, 22*, 42-57.
- Lykins, A. D., Robinson, J. J., LeBlanc, S., & Cantor, J. M. (2014). The effects of common medications on volumetric phallometry. *Journal of Sexual Aggression, 21*, 385-393.
- Mackaronis, J. E., Byrne, P. M., & Strassberg, D. S. (in press). Assessing sexual interest in adolescents who have sexually offended. *Sexual Abuse: A Journal of Research and Treatment*.
- Mackaronis, J. E., Strassberg, D. S., & Marcus, D. K. (2011). The latent structure of Multiphasic Sex Inventory-assessed pedophilic interest. *Psychological Assessment, 23*, 1017-1022.
- Mahoney, J. M., & Strassberg, D. S. (1991). Voluntary control of male sexual arousal. *Archives of Sexual Behavior, 20*, 1-16.
- Maier, W., Buller, R., Philipp, M., & Heuser, I. (1988). The hamilton anxiety scale: Reliability, validity and sensitivity to change in anxiety and depressive disorders. *Journal of Affective Disorders, 14*, 61-68.
- Maj, M., Pirozzi, R., Formicola, A. M., Bartoli, L., & Bucci, P. (2000). Reliability and validity of the DSM-IV diagnostic category of schizoaffective disorder: Preliminary data. *Journal of Affective Disorders, 57*, 95-98.
- Malcolm, P. B., Andrews, D. A., & Quinsey, V. L. (1993). Discriminant and predictive validity of phallometrically measured sexual age and gender preference. *Journal of Interpersonal Violence, 8*, 486-501.
- Marshall, W. L., Barbaree, H. E., & Christophe, D. (1986). Sexual offenders against female children: Sexual preferences for age of victims and type of behaviour. *Canadian Journal of Behavioural Science, 18*, 424-439.
- Marshall, W. L., & Eccles, A. (1991). Issues in clinical practice with sex offenders. *Journal of Interpersonal Violence, 6*, 68-93.
- Masters, W. H., & Johnson, V. E. (1963). The sexual response of the human male. I. Gross anatomic considerations. *Western journal of surgery, obstetrics, and gynecology, 71*, 85-95.
- McAnulty, R. D., & Adams, E. (1991). Voluntary control of penile tumescence: Effects of an incentive and a signal detection task. *Journal of Sex Research, 28*, 557-577.

- McConaghy, N. (1974). Measurements of change in penile dimension. *Archives of Sexual Behaviour*, 3, 331-338.
- McGorry, P. D. et al. (1995). Spurious precision: Procedural validity of diagnostic assessment in psychotic disorders. *American Journal of Psychiatry*, 152, 220-223.
- Merdian, H. L., & Jones, D. T. (2011). Phallometric assessment of sexual arousal. In Boer, D. P., Eher, R., Miner, M. H., Pfäfflin, F. & Craig, L. A. (Eds.), *International Perspectives on the Assessment and Treatment of Sexual Offenders: Theory, Practice and Research*. Chichester: Wiley-Blackwell.
- Moen, O. M. (2015). The ethics of pedophilia. *Nordic Journal of Applied Ethics*, 9, 111–124.
- Mokros, A., Butz, M., Dombert, B., Santtila, P., Baeuml, K.-H., & Osterheider, M. (2011). Judgment of age and attractiveness in a paired comparison task: Testing a picture set developed for diagnosing paedophilia. *Legal and Criminological Psychology*, 16, 323-334.
- Mokros, A., Dombert, B., Osterheider, M., Zappala, A., & Santtila, P. (2010). Assessment of pedophilic sexual interest with an attentional choice reaction time task. *Archives of Sexual Behaviour*, 39, 1081-1090.
- Mokros, A. et al. (2013). Computerized assessment of pedophilic sexual interest through self-report and viewing time: Reliability, validity, and classification accuracy of the affinity program. *Sexual Abuse: A Journal of Research and Treatment*, 25, 230-258.
- Molz, G., Schulze, R., Schroeders, U., & Wilhelm, O. (2010). TBS-TK Rezension: »Wechsler Intelligenztest für Erwachsene WIE. Deutschsprachige Bearbeitung und Adaptation des WAIS-III von David Wechsler«. *ReportPpsychologie*, 36, 369-371.
- Mularczyk, M., Zietek-Czeszak, A., & Zietek, Z. (2014). Assessment of sexual dimorphism of finger length ratio (2D:4D). *Annales Academiae Medicae Stetinensis*, 60, 47-51.
- Muller, K., Curry, S., Ranger, R., Briken, P., Bradford, J., & Fedoroff, J. P. (2014). Changes in sexual arousal as measured by penile plethysmography in men with pedophilic sexual interest. *The Journal of Sexual Medicine*, 11, 1221-9.
- Murphy, L., Ranger, R., Stewart, H., Dwyer, G., & Fedoroff, J. P. (2015). Assessment of problematic sexual interests with the penile plethysmograph: an overview of assessment laboratories. *Current Psychiatry Reports*, 17, 567.
- Murphy, W. D., & Barbaree, H. E. (1994). *Assessments of sex offenders by measures of erectile response: Psychometric properties and decision making*. Brandon, VT: Safer Society Press.
- Nichols, H. R., & Molinder, I. (1984). *Multiphasic Sex Inventory manual: A test to assess the psychosocial characteristics of the sexual offender*. Tacoma, Washington.
- O'Donohue, W., & Letourneau, E. (1992). The psychometric properties of the penile tumescence assessment of child molesters. *Journal of Psychopathology and Behavioural Assessment*, 14, 123-174.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9, 97-113.
- Paulus, C. (2009). Der Saarbrückener Persönlichkeitsfragebogen SPF (IRI) zur Messung von Empathie. Psychometrische Evaluation der deutschen Version des Interpersonal Reactivity Index. Saarbrücken: Universität des Saarlandes. Abgerufen am 14.3.2011 unter: <http://psydok.sulb.uni-saarland.de/volltexte/2009/2363/>
- Pohl, A. (2009). *Validierung der Kieler Stimulussets zur Messung der sexuellen Orientierung mittels Zirkumferenzphallometrie Ein Vergleich von visuellen und audiovisuellen Stimuli*. Unveröffentlichte Diplomarbeit, Christian Albrechts Universität zu Kiel.
- Pohl, A., Wolters, A., Ponseti, J. (2015). Investigating the Task Dependency of Viewing Time Effects. *Journal of Sex Research*, 23, 1-9.

- Polak, F., & Knobloch, F. (1957). Ist das Ausgangswertgesetz von Wilder ein Naturgesetz? *Journal of Neural Transmission*, *15*, 473-448.
- Ponseti, J. et al. (2006). A functional endophenotype for sexual orientation in humans. *Neuroimage*, *33*, 825-833.
- Ponseti, J. et al. (2012). Assessment of Pedophilia Using Hemodynamic Brain Response to Sexual Stimuli. *Archives of General Psychiatry*, *69*, 187-194.
- Ponseti, J., Granert, O., Jansen, O., Wolff, S., Mehdorn, H. M., Bosinski, H. A. G., & Siebner, H. (2009). Assessment of sexual orientation using the hemodynamic brain response to visual sexual stimuli. *Journal of Sexual Medicine*, *6*, 1628-1634.
- Ponseti, J. et al. (2014). Human face processing is tuned to sexual age preferences. *Biology Letters*, *10*, 20140200.
- Ponseti, J. et al. (2015). Assessing paedophilia based on the haemodynamic brain response to face images. *The World Journal of Biological Psychiatry*, *9*, 1-8.
- Proulx, J., Côte, M., & Achille, P. A. (1993). Prevention of voluntary control of penile response in homosexual pedophiles during phallometric testing. *The Journal of Sex Research*, *30*, 140-147.
- Pujol, J., Soriano-Mas, C., Alonso, P., Cardoner, N., Menchon, J. M., Deus, J., & Vallejo, J. (2004). Mapping structural brain alterations in obsessive-compulsive disorder. *Archives of General Psychiatry*, *61*, 720-730.
- Qiu, M. G., Zhang, J. N., Zhang, Y., Li, Q. Y., Xie, B., & Wang, J. (2012). Diffusion tensor imaging-based research on human white matter anatomy. *ScientificWorldJournal*, *2012*, 530432.
- Quinsey, V. L., & Chaplin, T. C. (1988). Preventing faking in phallometric assessments of sexual preference. *Annals of the New York Academy of Science*, *528*, 49-58.
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). Sex Offender Risk Appraisal Guide (SORAG). In Quinsey, V. L., Harris, G. T., Rice, M. E., Cormier, C. A. & Catherine, A. (Eds.), *Violent offenders: Appraising and managing risk* (Vol. 2). Washington, DC, US: American Psychological Association.
- Rosenzweig, S. (1942). The photoscope as an objective device for evaluating sexual interest. *Psychometric Medicine*, *4*, 150-158.
- Rösler, M., Retz-Junginger, P., Retz, W., & Stieglitz, R.-D. (2008). Homburger ADHS-Skalen für Erwachsene (HASE). Manual. Göttingen: Hogrefe.
- Rushton, J. P., Brainerd, C. J., & Pressley, M. (1983). Behavioral-development and construct-validity - the principle of aggregation. *Psychological Bulletin*, *94*, 18-38.
- Sachsenmaier, S. J., & Greess, C. L. Z. (2009). The Abel sessment for sexual interests - 2: A critical review. In Thornton, D. & Laws, D. R. (Eds.), *Cognitive Approaches to the Assessment of Sexual Intere in Sexual Offenders*. Chichester: John Wiley & Sons Ltd.
- Santtila, P., Antfolk, J., Rafsa, A., Hartwig, M., Sariola, H., Sandnabba, N. K., & Mokros, A. (2015). Men's sexual interest in children: One-year incidence and correlates in a population-based sample of Finnish male twins. *Journal of Child Sexual Abuse*, *24*, 115-134.
- Santtila, P. et al. (2010). Childhood sexual interactions with other children are associated with lower preferred age of sexual partners including sexual interest in children in adulthood. *Psychiatry Research*, *175*, 154-159.
- Schiffer, B. et al. (2008). Brain response to visual sexual stimuli in homosexual pedophiles. *Journal of Psychiatry and Neuroscience*, *33*, 23-33.
- Schiffer, B., Paul, T., Gizewski, E., Forsting, M., Leygraf, N., Schedlowski, M., & Kruger, T. H. (2008). Functional brain correlates of heterosexual paedophilia. *Neuroimage*, *41*, 80-91.

- Schmidt, A., Gykiere, K., Vanhoeck, K., Mann, R. E., & Banse, R. (2014). Direct and indirect measures of sexual maturity preferences differentiate subtypes of child sexual abusers. *Sexual Abuse: A Journal of Research and Treatment*, *26*, 107-28.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*, 350-353.
- Seto, M. C. (2007). *Pedophilia and sexual offending against children: Theory, assessment, and intervention* (Vol. 1). Washington, DC: American Psychological Association.
- Seto, M. C. (2012). Is pedophilia a sexual orientation? *Archives of Sexual Behaviour*, *41*, 231-236.
- Seto, M. C., Harris, G. T., Rice, M. E., & Barbaree, H. E. (2004). The screening scale for pedophilic interests predicts recidivism among adult sex offenders with child victims. *Archives of Sexual Behaviour*, *33*, 455-466.
- Seto, M. C., Hermann, C. A., Kjellgren, C., Priebe, G., Svedin, C. G., & Langstrom, N. (2015). Viewing child pornography: Prevalence and correlates in a representative community sample of young Swedish men. *Archives of Sexual Behaviour*, *44*, 67-79.
- Seto, M. C., & Lalumiere, M. L. (2001). A brief screening scale to identify pedophilic interests among child molesters. *Sexual Abuse: A Journal of Research and Treatment*, *13*, 15-25.
- Seto, M. C., Lalumière, M. L., & Blanchard, R. (2000). The discriminative validity of a phallometric test for pedophilic interests among adolescent sex offenders against children. *Psychological Assessment*, *12*, 319-327.
- Silva, S. A. (2011). *Viewing time and choice reaction time: Exploring its utility with child sex offenders in New Zealand*. Master Thesis, Massey University, Massey. Retrieved March 2014 from: <http://mro.massey.ac.nz/handle/10179/3169>.
- Simkins, L., Ward, W., Bowman, S., & Rinck, C. M. (1989). The multiphasic sex inventory: Diagnosis and prediction of treatment response in child sexual abusers. *Annals of Sex Research*, *2*, 205-226.
- Singer, B. (1984). Conceptualizing sexual arousal and attraction. *Journal of Sex Research*, *20*, 230-240.
- Soriano-Mas, C. et al. (2007). Identifying patients with obsessive-compulsive disorder using whole-brain anatomy. *Neuroimage*, *35*, 1028-1037.
- Steven, J. (2002). *Applied multivariate statistics for the social sciences*. (Vol. 4). Mahwah, NJ: Lawrence Erlbaum Associates.
- Stoleru, S., Fonteille, V., Cornelis, C., Joyal, C., & Moulrier, V. (2012). Functional neuroimaging studies of sexual arousal and orgasm in healthy men and women: A review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *36*, 1481-1509.
- Strobel, A., Beauducel, A., Debener, S., & Brocke, B. (2001). Eine deutschsprachige Version des BIS/BAS-Fragebogens von Carver und White. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *22*, 216-227.
- Tanner, J. M. (1978). *Foetus into man: Physical growth from conception to maturity*. Cambridge: Harvard University Press.
- Thornton, D., & Laws, D. R. (2009). *Cognitive approaches to the assessment of sexual interest in sexual offenders*. Chichester: John Wiley & Sons.
- Tong, D. (2007). The penile plethysmograph, abel assessment for sexual interest, and MSI-II: Are they speaking the same language? *The American Journal of Family Therapy*, *35*, 187-202.
- Travin, S., Cullen, K., & Melella, J. T. (1988). The use and abuse of erection measurements: A forensic perspective. *The Bulletin of the American Academy of Psychiatry and the Law*, *16*, 235-250.
- Vogt, H. (2006). *Pädophilie: Leipziger Studie zur gesellschaftlichen und psychischen Situation pädophiler Männer*. Pabst Science Publishers.

- Walter, M. et al. (2007). Pedophilia is linked to reduced activation in hypothalamus and lateral prefrontal cortex during visual erotic stimulation. *Biological Psychiatry*, 62, 698-701.
- Williams, J. B. W. (1988). A structured interview guide for the Hamilton Depression Rating Scale. *Archives of General Psychiatry*, 45, 742-747.
- Williams, K. M., Cooper, B. S., Howell, T. M., Yuille, J. C., & Paulhus, D. L. (2008). Inferring sexually deviant behavior from corresponding fantasies: The role of personality and pornography consumption. *Criminal Justice and Behavior*, 36, 198-222.
- Wilson, R. J. (1998). Psychophysiological signs of faking in the phallometric test. *Sexual Abuse: A Journal of Research and Therapy*, 10, 113-126.
- Wittchen, H.-U., Zaudig, M., & Fydrich, T. (1997). *Strukturiertes Klinisches Interview für DSM-IV*. Göttingen: Hogrefe.
- Worling, J. R. (2006). Assessing sexual arousal with adolescent males who have offended sexually: Self-report and unobtrusively measured viewing time. *Sexual Abuse: A Journal of Research and Treatment*, 18, 383-400.
- Wormith, J. S. (1986). Assessing deviant sexual arousal: Physiological and cognitive aspects. *Advances in behaviour research and therapy*, 8, 101-137.
- Wurtele, S. K., Simons, D. A., & Moreno, T. (2014). Sexual interest in children among an online sample of men and women: Prevalence and correlates. *Sexual Abuse: A Journal of Research and Treatment*, 26, 546-568.
- Zamansky, H. S. (1956). A technique for assessing homosexual tendencies. *Journal of Personality*, 24, 436-448.
- Zou, K. H., O'Malley, A. J., & Mauri, L. (2007). Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation*, 115, 654-657.

7 Anhang

7.1 Anhang 1: Probandeninformation für NeMUP



Sektion für Sexualmedizin, UNIVERSITÄTSKLINIKUM Schleswig-Holstein,
Campus Kiel, Arnold-Heller-Str. 3, Haus 28, 24105 Kiel

UNIVERSITÄTSKLINIKUM
Schleswig-Holstein

→ **Campus Kiel**
Sektion für Sexualmedizin
Leiter: Prof. Dr. med. Hartmut A. G. Bosinski
Ansprechpartner: Dr. Dipl.-Psych. Jorge Ponzel
Tel: 0431 / 597-3652
Fax: 0431 / 597-3984
E-Mail: ponzel@sexmed.uni-kiel.de
Internet: www.uni-kiel.de/sexmed

Probanden-Information zur Studie:

Neuronale Mechanismen von Pädophilie und sexuellem Kindesmissbrauch: Ursachen, Diagnose und Behandlung

Sehr geehrter Studienteilnehmer,
wir danken Ihnen für das Interesse an dieser Studie. Bitte lesen Sie dieses Informationsschreiben aufmerksam durch und fragen Sie die Studienleiter, wenn Sie etwas nicht verstehen. Wenn Sie der Teilnahme an der Studie zustimmen, möchten wir Sie bitten, die Einverständniserklärung zu unterschreiben. Ein Exemplar dieser Erklärung zusammen mit dieser Probanden-Information wird Ihnen dann ausgehändigt.

1. Ziel der Studie

Das Ziel dieser Studie ist es, die neurobiologischen Grundlagen von Pädophilie und sexuellem Kindesmissbrauch besser zu verstehen und eventuelle Therapieeffekte auf Gehirnstrukturen und -funktionen zu identifizieren. Diese Erkenntnisse sollen die Diagnostik und Behandlung von Pädophilien und sexuellen Kindesmissbräuchen verbessern.

2. Allgemeine Informationen zum Studieninhalt und -ablauf

Personen, die eine pädophile Neigung aufweisen und/oder wegen sexuellen Missbrauchs von Kindern verurteilt wurden, weisen bestimmte kognitive Funktionsstörungen auf, die mit Besonderheiten in der Verarbeitung von Reizen im Gehirn einhergehen können und möglicherweise auch strukturelle Veränderungen aufweisen.

Die Untersuchung setzt sich aus drei Teilen zusammen, die an drei unterschiedlichen Tagen erfolgen. Der erste Teil umfasst eine psychodiagnostische und eine phallographische Untersuchung die etwa 3-4 Stunden dauert. Die phallographische Messung dient der objektiven Bestimmung der sexuellen Orientierung. Hierzu platziert der Proband eine kleine elastische Schlaufe um seinen Penis und beobachtet anschließend eine Reihe von Nacktbildern. Während dessen werden Umfangsänderungen des Penis im Nachbarraum durch einen Computer registriert. Der Proband befindet sich während der Messung alleine im Raum.

Der zweite Teil umfasst eine magnetresonanztomographische Untersuchung (MRT) und findet zu einem späteren Termin im Neurozentrum des UK-SH statt. Die Untersuchung des Gehirns beim Menschen ist mit der Magnetresonanztomographie (MRT) möglich. Der Magnetresonanztomograph ist eine Röhre, die Schichtaufnahmen des Körpers ohne Röntgenstrahlen erzeugt. Wasser als Hauptbestandteil des menschlichen Körpers wird durch ein Magnetfeld und Radioimpulse angeregt, anschließend werden aus der Kernspin-Antwort die Schichtbilder berechnet. Diese Methode eröffnet die Möglichkeit in die Funktionsweise des Gehirns einen

Universitätsklinikum
Schleswig-Holstein
Ansatz des
öffentlichen Rechts

Vorstandsmitglieder:
Prof. Dr. Jens Scholz
Christa Meyer
Peter Pansegrau

Bankverbindungen:
Sparkasse Kiel
Kto.-Nr. 100 206, BLZ 210 501 70
Dresdner Bank Lübeck
Kto.-Nr. 300 041 200, BLZ 230 800 40



Einblick zu erhalten. Diese Untersuchungsmethode wird hier angewendet, da sie bei Einhaltung der Kontraindikationen risikolos ist und ohne invasive Maßnahmen einen Einblick in die Gehirnfunktion erlaubt. Es ist nicht nötig, Kontrastmittel zu geben und es ist keine Belastung durch radioaktive Strahlung oder Röntgenstrahlung vorhanden. Die MRT-Untersuchung umfasst zwei Teile. Zunächst werden zwei strukturelle Aufnahmen des Gehirns gemacht. Anschließend folgen mehrere Durchläufe bei denen unterschiedliche Testaufgaben zu bewältigen sind. Dabei handelt es sich um vergleichbare Aufgaben wie bereits im ersten Untersuchungsteil der Studie. Außerdem werden während der MRT-Untersuchung auch Nacktbilder von Kindern und von Erwachsenen gezeigt. Die Dauer der gesamten MRT Untersuchung beträgt etwa 1 Stunde. Die Untersuchung kann durch eine Pause in zwei Teile à ca. 30 Minuten aufgeteilt werden.

An diesem Untersuchungstermin erfolgt auch eine Blutentnahme für hormonelle, genetische und epigenetische Analysen (etwa vier Röhrchen). In den Blutproben sollen Hormone wie Prolaktin, Testosteron und Cortisol bestimmt werden, die im Zusammenhang mit Verhalten, Denken und Sexualpräferenz stehen können. Die Epigenetik ist ein Spezialgebiet der Biologie. Sie befasst sich mit Zelleigenschaften, die auf Tochterzellen vererbt werden und nicht in der DNA-Sequenz – dem eigentlichen Ort der Gene – festgelegt sind. Hierbei erfolgen Veränderungen an den Chromosomen, wodurch Abschnitte oder ganze Chromosomen in ihrer Aktivität beeinflusst werden. Man spricht infolgedessen auch von sog. epigenetischer Veränderung bzw. epigenetischer Prägung. Die DNA-Sequenz wird dabei jedoch nicht verändert. Zusätzlich geplante genetische Untersuchungen hingegen beschäftigen sich mit den Erbanlagen. Die genetischen Analysen umfassen im Wesentlichen die Gene für Transporter, Rezeptoren und Enzyme bestimmter Hormone und Nervenbotenstoffe (z.B. Androgene, Serotonin, Oxytocin). Sollten Sie zu diesem Teilaspekt ihre Einwilligung nicht geben wollen, so ist dies selbstverständlich möglich. Genetische und epigenetische Untersuchungen sollen dem besseren Verständnis von Erbanlagen, der Modulation von Erbanlagen durch verschiedene Faktoren wie Lebensereignisse und die dadurch bedingte Beeinflussung von Verhalten, Denken und Sexualpräferenz dienen.

Am dritten Untersuchungstermin wird wieder eine MRT-Untersuchung durchgeführt, die aber dann nur etwa 40 Minuten dauert.

Die Patienten, die an einer psychotherapeutischen Behandlung teilnehmen, sollen nach Abschluss der Behandlung ein zweites Mal an dieser Untersuchung teilnehmen, um etwaige Veränderungen im zentralen Nervensystem abbilden zu können.

3. Freiwilligkeit der Untersuchung

Eine Teilnahme an dieser Studie erfolgt auf freiwilliger Basis. Die Untersuchung kann von Ihrer Seite jederzeit ohne Angabe von Gründen und ohne Nachteile abgebrochen werden. Die zu dem Zeitpunkt gewonnenen Daten oder Blutproben würden vernichtet werden. Die Teilnahme ist unabhängig von der weiteren Unterbringung oder Behandlung (gilt nur für Personen, die zum Zeitpunkt der Untersuchung inhafiert sind).

4. Nutzen für die Teilnehmer

Zum gegenwärtigen Zeitpunkt ergibt sich aus dem Ergebnis noch keine unmittelbare therapeutische oder andere Konsequenz. Wir hoffen aber auf einen Erkenntnisgewinn, der langfristig dazu beiträgt, dass Sie bzw. Patienten mit diesem Krankheitsbild besser behandelt werden bzw. Risikofaktoren für sexuellen Missbrauch von Kindern frühzeitig erkannt und entsprechend behandelt werden können.

Zufallsbefunde

Es besteht die wenig wahrscheinliche Möglichkeit, dass sich in den kernspintomographischen Aufnahmen des Gehirns als Zufallsbefund krankhafte Veränderungen finden. In diesem Fall wird Ihnen ein Vorschlag für eine fachärztliche Abklärung gemacht. Die Kosten für eine weiterführende Abklärung sind von Ihnen bzw. Ihrer Krankenkasse zu tragen. Möglicher Vorteil ist, dass Veränderungen früh erkannt und behandelt werden können. Auf der anderen Seite können Befunde aufgedeckt werden, die nicht oder nur bedingt behandelbar sind. Es können sich mögliche nachteilige Konsequenzen in der Arbeitswelt und bei zukünftigen Versicherungen ergeben. Eine Probandenversicherung, die mögliche nachteilige Folgen für Sie abdeckt, besteht nicht.

5. Risiken und Unannehmlichkeiten

a. Magnetresonanztomographie

Bei Beachtung der Ausschlusskriterien für die MRT bestehen keine bekannten Risiken. Ausschlusskriterien sind insbesondere Herzschrittmacher, Herzklappen, Medikamentenpumpen, andere Metallteile wie Defibrillatoren, Gefäß- oder chirurgische Clips, Gelenkprothesen, Gefäßprothesen (Stent), Metallstifte, Zahnspannen, Granat- oder sonstige Metallsplitter, Spirale (Intrauterin Pessar), Akupunktur-Nadeln, Insulinpumpe, Gehörimplantat, Intraport, Tätowierungen, Permanent Make-up etc. oder psychische Gründe wie Platzangst.

Aufgrund des starken Magnetfeldes dürfen keinerlei Gegenstände oder Geräte, die aus Metall sind oder Metall enthalten könnten, mit in den Untersuchungsraum genommen werden. Legen Sie solche Gegenstände und Geräte (z.B. Mobiltelefon, Münzen, Kugelschreiber, Schlüssel, Uhren, Schmuck, Brillen, Hörgeräte, Funkrufempfänger, Haarnadeln) unbedingt vorher ab, da sie in der Nähe des Magneten zu gefährlichen Geschossen werden können!

b. Neuropsychologische Testaufgaben

Die Testaufgaben können als schwierig oder anstrengend empfunden werden und es kann durch mögliche Anspannung oder Nervosität zu Herzklopfen (Anstieg der Herzrate, des Blutdrucks) kommen. Dies ist in aller Regel vorübergehend und beim Abschluss der Messungen abgeklungen.

c. Blutentnahme für hormonelle, genetische und epigenetische Untersuchungen

Die Blutentnahme kann zu einem kurzen Schmerzerleben führen. An der Einstichstelle kann ein blauer Fleck entstehen. In extrem seltenen Fällen kann es zu einer Verletzung von Hautnerven kommen, was zu sensiblen Ausfalls- oder Reizerscheinungen führen kann. Da die zu entnehmende Blutmenge gering ist (etwa 50 ml), sind andere Beeinträchtigungen des Organismus unwahrscheinlich. Zum Vergleich: Bei einer Blutspende werden etwa 500 ml Blut abgenommen.

6. Vertraulichkeit der Daten

In dieser Studie werden persönliche Daten von Ihnen erfasst. Diese Daten werden pseudonymisiert. Das heißt, dass Identifikationsmerkmale wie Name und Anschrift durch ein Kennzeichen (Codenummer) ersetzt sind, so dass eine Zuordnung zu einer bestimmten Person nur über weitere Hilfsmittel (Referenzliste) möglich ist. Die Daten sind nur Fachleuten zur wissenschaftlichen Auswertung zugänglich. Ihr Name wird in keiner Weise in Berichten oder Publikationen veröffentlicht, die aus der Studie hervorgehen.

7. Entgelt für die Teilnahme an der klinischen Studie

Jeder Patient/Proband erhält eine Aufwandsentschädigung von 150,-€ für die Teilnahme an der gesamten Untersuchung.

8. Versicherungsschutz

Aufgrund der nachgewiesenen Ungefährlichkeit der Untersuchungsmethoden, besteht keine Notwendigkeit für eine spezielle Probandenversicherung, die Leistungen gewähren würde, wenn kein anderer für den Schaden haftet. Des Weiteren wird ausdrücklich darauf hingewiesen, dass kein Versicherungsschutz besteht für Unfälle, die sich auf dem Weg zu und von durch die Studie veranlassten Untersuchungen ereignen.

9. Ethische Voraussetzungen

Diese Untersuchung wurde der für Ihren Studienleiter zuständigen, unabhängigen Ethikkommission zur Beurteilung vorgelegt und wird entsprechend den ethischen Normen durchgeführt.

10. Kontaktperson

Bei Unklarheiten, Notfällen, unerwarteten oder unerwünschten Ereignissen, die während der Studie oder nach deren Abschluss auftreten, können Sie sich jederzeit an folgende Kontaktperson wenden:

PD Dr. Dipl.-Psych. Jorge Ponseti
Sektion für Sexualmedizin
Universitätsklinikum Schleswig-Holstein, Campus Kiel
0431-5973652

Abschließend möchten wir Ihnen für Ihre Mitarbeit sehr herzlich danken.

Kiel, 23.03.2012

7.2 Anhang 2: Einverständniserklärung NeMUP



Sektion für Sexualmedizin, UNIVERSITÄTSKLINIKUM Schleswig-Holstein,
Campus Kiel, Arnold-Heller-Str. 3, Haus 28, 24105 Kiel

UNIVERSITÄTSKLINIKUM
Schleswig-Holstein

→ Campus Kiel
Sektion für Sexualmedizin
Leiter: Prof. Dr. med. Hartmut A. G. Bosinski
Ansprechpartner: Dr. Dipl.-Psych. Jorge Ponseti
Tel: 0431 / 597-3652
Fax: 0431 / 597-3984
E-Mail: ponseti@sexmed.uni-kiel.de
Internet: www.uni-kiel.de/sexmed

Schriftliche Einverständniserklärung des Probanden zur Teilnahme an einer Studie

- Bitte lesen Sie dieses Formular sorgfältig durch!
- Bitte fragen Sie, wenn Sie etwas nicht verstehen oder wissen möchten!

Titel der Studie: Neuronale Mechanismen von Pädophilie und sexuellem Kindesmissbrauch:
Ursachen, Diagnose und Behandlung

Ort der Studie: Sektion für Sexualmedizin & Neurozentrum Universitätsklinikum Schleswig-Holstein
/Campus Kiel

Studienleiter: PD Dr. Dipl.-Psych. Jorge Ponseti

Versuchsperson
Name und Vorname:

Geburtsdatum:

Ich wurde vom Versuchsleiter mündlich und schriftlich über die Ziele, den Ablauf der o. g. Studie und über mögliche Vor- und Nachteile sowie über eventuelle Risiken informiert. Ich habe die zur oben genannten Studie abgegebene schriftliche Patienteninformation gelesen und verstanden. Meine Fragen im Zusammenhang mit der Teilnahme an dieser Studie sind mir zufriedenstellend beantwortet worden. Ich kann die schriftliche Patienteninformation behalten und erhalte eine Kopie meiner schriftlichen Einverständniserklärung. Ich hatte genügend Zeit, um meine Entscheidung zu treffen.

Ich nehme an dieser Studie freiwillig teil. Ich kann jederzeit und ohne Angabe von Gründen meine Zustimmung zur Teilnahme widerrufen, ohne dass mir deswegen Nachteile bei der weiteren medizinischen Betreuung entstehen.

Ich erkläre, dass ich mit im Rahmen der Studie erfolgenden Aufzeichnungen von Krankheits-/Studiendaten einverstanden bin. Außerdem bin ich einverstanden, dass die Studiendaten ohne Namensnennung auf

Universitätsklinikum
Schleswig-Holstein
Anstalt des
öffentlichen Rechts

Vorstandsmitglieder:
Prof. Dr. Jens Scholz
Christa Meyer
Peter Pansegrau

Bankverbindungen:
Sparkasse Kiel
Kto.-Nr. 100 206, BLZ 210 501 70
Dresdner Bank Lübeck
Kto.-Nr. 300 041 200, BLZ 230 800 40



elektronischen Datenträgern aufgezeichnet, ausgewertet und für wissenschaftliche Darstellungen und Veröffentlichungen verwendet werden dürfen. Ich bin damit einverstanden, dass mir mögliche pathologische Zufallsbefunde mitgeteilt werden.

Ich habe zur Kenntnis genommen, dass keine spezielle Probandenversicherung besteht, die Leistungen gewähren würde, wenn kein anderer für den Schaden haftet und dass kein Versicherungsschutz besteht für Unfälle, die sich auf dem Weg zu und von durch die Studie veranlassten Untersuchungen ereignen.

Einwilligung in die Entnahme und Nutzung von Blutproben

1. Ich bin mit der Entnahme einer Blutprobe für hormonelle, epigenetische und genetische Untersuchungen einverstanden.
2. Ich übereigne die mir entnommene Blutprobe hiermit an die oben genannte Institution bzw. Person.
3. Meine nachfolgenden abgegebenen Erklärungen sollen nach Möglichkeit meine Persönlichkeit und meine Rechtsposition schützen. Ich stimme zu, dass die Blutprobe unter der Verantwortung der oben genannten Institution bzw. Person in pseudonymisierter Form verwendet wird.

Widerruf

Ich weiß, dass ich meine Zustimmung zur Studienteilnahme einschließlich der Verwendung meiner Blutproben jederzeit und ohne Angabe von Gründen gegenüber der oben genannten Institution/Person widerrufen kann und dass dies keinen Einfluss auf meine etwaige weitere ärztliche Behandlung hat. Im Falle des Widerrufs bin ich damit einverstanden, dass meine Blutprobe zu Kontrollzwecken weiter aufbewahrt wird. Ich habe jedoch das Recht, deren Vernichtung zu verlangen, sofern gesetzliche Bestimmungen der Vernichtung nicht entgegenstehen.

Ort, Datum & Unterschrift der Versuchsperson

Bestätigung des Studienleiters:

Hiermit bestätige ich, dass ich diesem Probanden Wesen, Bedeutung und Tragweite der Studie erläutert habe. Ich versichere, alle im Zusammenhang mit dieser Studie stehenden Verpflichtungen zu erfüllen.

Ort, Datum & Unterschrift des Studienleiters

7.3 Anhang 3: Interviewleitfaden NeMUP

NEMUP Erhebungsbogen

Erhebungsbogen zu soziodemografischen, forensischen und sexualmedizinischen Basisdaten (inkl. MRT Screening)

Dunkelfeld Hellfeld Ex-Hellfeld Unauffällig

Erhebungsdatum: _____ Untersucher: _____

Probandencode: _____ Alter: _____

EINSCHLUSSKRITERIEN

Sexuelle Präferenz	homosexuell		heterosexuell
	pädophil	hebephil	teleiophil
(Zusätzliche) Paraphilie(n)	Ja/ Nein	Spezifikation:	
Edinburgh Handedness	Lateralisierungsindex:		Händigkeit:

AUSSCHLUSSKRITERIEN

Alter 18 bis 50 Jahre?	Ja/ Nein		
MRT-Kriterien erfüllt?	Ja/ Nein		
Substanzmissbrauch oder -abhängigkeit in den letzten 6 Monaten?	Ja/ Nein		
Schwere Intelligenzminderung?	Ja/ Nein		
Körperliche und psychische Beschwerden? ¹	Körperliche:		Psychische:
			HAM-D ≥ 15 ?
			Ja/ Nein
		HAM-A ≥ 19 ?	Ja/ Nein
Medikation? ²	Anxiolytika	Ja/ Nein	Spezifikation:
	Antidepressiva	Ja/ Nein	Spezifikation:
	GnRH-Analoga	Ja/ Nein	Spezifikation:
	Antiandrogene	Ja/ Nein	Spezifikation:
	Neuroleptika	Ja/ Nein	Spezifikation:

GRUPPENZUORDNUNG

Pädophil / Hebephil (ohne CSA)	Pädophil / Hebephil (mit CSA)	Teleiophil mit CSA	Teleiophil ohne CSA (Kontrollgruppe)	unsicher zuzuordnen oder Ausschlusskriterium erfüllt
--------------------------------	-------------------------------	--------------------	--------------------------------------	--

¹Schizophrenie immer Ausschluss. Depression Ausschluss anhand HAM-D. Angststörung Ausschluss anhand HAM-A.

²Antidepressiva-Ausschluss: Vor MRT Session bitte halbwertszeit-abhängiges Wash-Out anordnen!!!

Sertralin, Venlafaxin, Citalopram, Paroxetin, Fluoxetin, Imipramin, Phenelzin, Duloxetine, Escitalopram, Fluvoxamin. GnRH-Analoga, Antiandrogene und Neuroleptika immer Ausschluss. Anxiolytika-Ausschluss: ???

Zulässig wären hingegen: *Agomelatin, Amineptin, Bupropion, Moclobemid, Mirtazapin, Nefazodon.*

NEMUP Erhebungsbogen

MRT Screening auf Untersuchungstauglichkeit

Größe: _____ cm Gewicht: _____ kg

	<i>Ja</i>	<i>Nein</i>
Sind Sie fehlsichtig?	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> kurz <input type="checkbox"/> weit		
<input type="checkbox"/> Brille <input type="checkbox"/> Linsen		
Dioptrien? _____ L _____ R		
Haben Sie eine Hornhautverkrümmung?	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> schwach <input type="checkbox"/> stark		
Schielen Sie?	<input type="checkbox"/>	<input type="checkbox"/>
Haben Sie Probleme beim Farbsehen?	<input type="checkbox"/>	<input type="checkbox"/>

Befindet sich eines oder mehrere folgender Metallteile in oder an Ihrem Körper?

	<i>Ja</i>	<i>Nein</i>
Metallprothesen/Schrauben/Platten nach Knochenbruch	<input type="checkbox"/>	<input type="checkbox"/>
Kieferorthopäd. Einsätze (Brackets, Retainer)	<input type="checkbox"/>	<input type="checkbox"/>
Piercings / Intimschmuck / Implants	<input type="checkbox"/>	<input type="checkbox"/>
Zahnimplantat / Stiftzahn	<input type="checkbox"/>	<input type="checkbox"/>
Ohrimplantate (z.B. Hörgerät, CI)	<input type="checkbox"/>	<input type="checkbox"/>
Augenimplantate	<input type="checkbox"/>	<input type="checkbox"/>
Künstliche Gelenke	<input type="checkbox"/>	<input type="checkbox"/>
Gefäßprothesen (Stent)	<input type="checkbox"/>	<input type="checkbox"/>
Künstliche Herzklappen	<input type="checkbox"/>	<input type="checkbox"/>

NEMUP Erhebungsbogen

Herzschrittmacher	<input type="checkbox"/>	<input type="checkbox"/>
Neurostimulator	<input type="checkbox"/>	<input type="checkbox"/>
Gefäßclips	<input type="checkbox"/>	<input type="checkbox"/>
Metallsplitter, Metallreste nach Verletzungen	<input type="checkbox"/>	<input type="checkbox"/>
Implantierte Elektroden	<input type="checkbox"/>	<input type="checkbox"/>
Medikamentenpumpe (Insulinpumpe etc.)	<input type="checkbox"/>	<input type="checkbox"/>
Akupunkturnadeln	<input type="checkbox"/>	<input type="checkbox"/>
Sonstiges: _____	<input type="checkbox"/>	<input type="checkbox"/>

Wurden Sie schon einmal operiert, z.B. am Herzen, Kopf, Augen, HNO-Bereich oder orthopädisch?

Ja Nein Wenn ja – Wo/was? _____

Hatten Sie schon einmal einen schweren Unfall?

Ja Nein Wenn ja: Waren Sie danach bewusstlos? Ja Nein

Haben Sie schon einmal mit Metallen gearbeitet, z.B. geschweißt/gefräst?

Ja Nein Wenn ja: Schutzkleidung/Schutzbrille getragen? Ja Nein

Haben Sie sich dabei verletzt? Ja Nein

Tragen Sie eine Tätowierung?

Ja Nein Wenn ja – Wo/wie groß/von wann/woher? _____

Leiden Sie unter Tinnitus?

Ja Nein

Leiden Sie unter Angst vor engen Räumen (z.B. beim Aufzugfahren)?

Ja Nein

Nehmen Sie zurzeit/haben Sie schon einmal (regelmäßig) Medikamente eingenommen?

Ja Nein Wenn ja – Welche/seit wann/wie lange? _____

Konsumieren Sie aktuell psychoaktive Substanzen?

Ja Nein Wenn ja – Welche/seit wann/wie viel? _____

Händigkeitsfragebogen (Edinburgh Inventory nach Oldfield, 1971)

Instruktion: 'Jetzt geht es darum, mit welcher Hand Sie die meisten alltäglichen Tätigkeiten durchführen. Bitte nennen Sie mir die Hand, mit der Sie die nachfolgend von mir abgefragten Tätigkeiten ausführen. Bitte sagen Sie einfach bei jeder Tätigkeit, rechts, links, gemischt, oder selten beide, je nachdem, was auf Sie zutrifft.'

Tätigkeit	Rechts	Links	Mix	Use other hand
1. etwas schreiben				
2. etwas zeichnen				
3. etwas werfen				
4. eine Schere benutzen				
5. Zähne putzen				
6. ein Messer halten (ohne Gabel, z.B. Brot schneiden)				
7. einen Löffel halten				
8. einen Besen halten (obere Hand)				
9. ein Streichholz anzünden				
10. den Deckel einer Flasche öffnen				
11. die Maus eines Computers halten				
12. eine Tür mit einem Schlüssel öffnen				
13. einen Hammer halten				
14. eine Bürste oder einen Kamm halten				
15. ein Trinkglas halten				
Summe ($\geq 8 \rightarrow$ Handpräferenz)				

Elternhaus

Vater (Alter bei Geburt.....)

Alter, Beruf, psychische Gesundheit, strafrechtliche Vorgeschichte

.....

Mutter (Alter bei Geburt.....)

Alter, Beruf, psychische Gesundheit, strafrechtliche Vorgeschichte

.....

NEMUP Erhebungsbogen

Geschwister? Anzahl? Alter?

(Stellung in der Geschwisterreihe: von)

.....
.....
.....

Entwicklungsbesonderheiten

- Geburtskomplikationen? Ja Nein
- Sauberkeitsentwicklung? Ja Nein
- Laufen und Sprechen lernen etc., Ja Nein
- Kopfverletzungen? Ja Nein
- Schädel-Hirn-Trauma? Ja Nein
- Bewusstseinsverlust? Ja Nein
- Kinderkrankheiten? Ja Nein

Schulabschluss

Keinen Hauptschule Realschule Abitur Hochschule/FH

Beruflicher Werdegang

Beschäftigungsverhältnis > 2 Jahre Ja Nein

.....
.....
.....

Berufsstand bei Anlassstraftat

erwerbslos in Ausbildung berentet erwerbstätig

NEMUP Erhebungsbogen

Familienstand

ledig

verheiratet

geschieden

verwitwet

Kinder

Anzahl

.....
.....

Sohn (Alter)

.....
.....

Tochter (Alter)

.....
.....

Hat eine Sexualstraftat zur Vorstellung in der hiesigen Institution geführt?

Ja

Nein

NEMUP Erhebungsbogen

I. Delinquenzgeschichte

Vorstrafen

Anzahl der Delikte über die Lebensspanne

a) Gewaltdelikte

.....

b) Sexualdelikte

.....

c) Andere Delikte

.....

Sexualdelikte im Einzelnen (Kodierung der letzten 5 Delikte)

Jahr der Tat	Entwicklung Opfer (Tanner)	Handlung	Anklage	Inzesttat	Opfer > 24h gekannt	durch Aktenlage verifiziert
			ja <input type="checkbox"/> nein <input type="checkbox"/>			
			ja <input type="checkbox"/> nein <input type="checkbox"/>			
			ja <input type="checkbox"/> nein <input type="checkbox"/>			
			ja <input type="checkbox"/> nein <input type="checkbox"/>			
			ja <input type="checkbox"/> nein <input type="checkbox"/>			

NEMUP Erhebungsbogen

Allg. Charakteristika der Opfer

Anzahl männlicher Opfer

.....

Anzahl weiblicher Opfer

.....

Alter männlicher Opfer

von bis MW ()

Alter weiblicher Opfer

von bis MW ()

Konsum von Kinderpornografie?

Ja Nein

Wenn ja, Häufigkeit des Konsums:

mehrmals pro Tag mehrmals pro Woche ein Mal pro Woche

Nackt- oder Indikativbilder von Kindern?

Ja Nein

Wenn ja, Häufigkeit des Konsums:

mehrmals pro Tag mehrmals pro Woche ein Mal pro Woche

NEMUP Erhebungsbogen

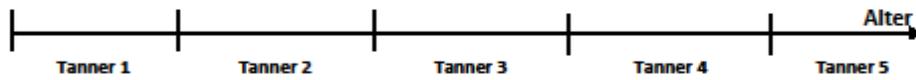
II. Modifizierte Kinsey-Skala

(Kodierung: Range und Peak je Kategorie entsprechend der Selbstauskünfte nach Vorlage der Tannerskalen)

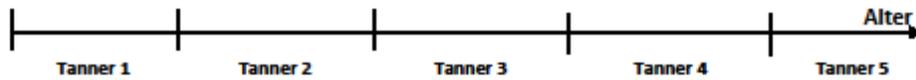
Sexuelles Verhalten/Handlungen

(tatsächliche Handlungen, im Hellfeld --> untere Grenze der Range ggf. mit Aktenlage abgleichen)

heterosexuell:

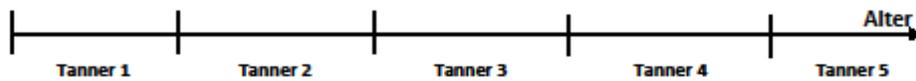


homosexuell:

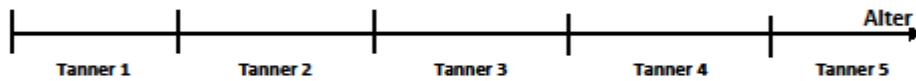


Sexuelle Phantasien (Masturbation)

heterosexuell:



homosexuell:



NEMUP Erhebungsbogen

III. Sexuelle Entwicklung und aktuelle sexuelle Funktion

Erster Samenerguss im Alter von Jahren

Erste Selbstbefriedigung im Alter von Jahren

Erster Geschlechtsverkehr im Alter von Jahren

Anzahl der Koituspartner > 17 Jahre bis gegenwärtig:

Männer

Frauen

Bemerkung

(ggf. mangelnde Motivation bei der Initiierung sexueller Kontakte mit Gleichaltrigen, v. a. im Hellfeld je nach Aktenlage bzw. Auskünften signifikanter Dritte. Befriedigung erlebt?)

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

(Durchschnittliche) Anzahl Orgasmen pro Woche _____

Wie viele davon durch Selbstbefriedigung? _____

Wie viele davon unter Zuhilfenahme pornographischen Materials (bzw. Indikativbilder o.ä.)?

NEMUP Erhebungsbogen

Art des verwendeten Materials *(Erwachsene/Kinder/Praktiken)?*

.....
.....
.....

Leiden Sie unter ein Mangel oder Verlust an sex. Verlangen? Ja Nein

Finden Sie Sex ekelhaft? *(sexuelle Aversion)* Ja Nein

Leiden Sie unter mangelnder sexueller Befriedigung? Ja Nein
(Wenn kein GV-Partner, dann lustlos erlebte Orgasmen bei der Masturbation)

Bestehen Probleme mit der Erektion? Ja Nein

Haben Sie bei der Masturbation oder beim GV Probleme den Orgasmus zu erreichen?
Ja Nein

Leiden Sie unter einem vorzeitigen Samenerguss *(< 1 Min)?* Ja Nein

Leiden Sie unter gesteigertem sexuellem Verlangen? Ja Nein

Leiden Sie unter Schmerzen beim Geschlechtsverkehr? Ja Nein

IV. Sonstige Paraphilien

Haben Sie einen Drang Ihren Penis vor nichtsehenden Fremden zu zeigen?

Ja Nein

Haben Sie das bereits ein- oder mehrmals getan?

Ja Nein

Leiden Sie unter solchen Wünschen bzw. Vorstellungen?

Ja Nein

Besteht dieser Drang wiederkehrend seit mindestens 6 Monaten?

Ja Nein

NEMUP Erhebungsbogen

Haben Sie das Bedürfnis bestimmte Objekte oder Körperteile, wie z. B. Schuhe, Latex, Windeln oder Füße, bei sexuellen Handlungen einzubeziehen (insb. Selbstbefriedigung oder Geschlechtsverkehr)?

Ja Nein

Wenn ja, leiden Sie darunter?

Ja Nein

Besteht dieses Bedürfnis wiederkehrend seit mindestens 6 Monaten?

Ja Nein

Haben Sie wiederkehrend einen sexuellen Drang Personen zu berühren oder sich an ihnen zu reiben, obwohl diese nicht einverstanden sind?

Ja Nein

Haben Sie das bereits ein- oder mehrmals getan?

Ja Nein

Leiden Sie unter solchen Wünschen bzw. Vorstellungen?

Ja Nein

Besteht dieser Drang seit mindestens 6 Monaten?

Ja Nein

Empfinden Sie es als sexuell erregend von jemandem (real) gequält oder gedemütigt zu werden, bzw. sich selbst zu quälen?

Ja Nein

Leiden Sie selbst unter dieser Neigung?

Ja Nein

Besteht dieses Bedürfnis wiederkehrend seit mindestens 6 Monaten?

Ja Nein

Empfinden Sie es als sexuell erregend einen Partner (real) zu quälen oder zu demütigen?

Ja Nein

Haben Sie bereits (ein- oder mehrmals) jemanden gequält oder gedemütigt, um sich sexuell zu stimulieren?

NEMUP Erhebungsbogen

Ja Nein

Leiden Sie selbst unter dieser Neigung?

Ja Nein

Besteht dieses Bedürfnis wiederkehrend seit mindestens 6 Monaten?

Ja Nein

Ist es für Sie sexuell erregend Damenkleidung zu tragen oder sich vorzustellen, eine Frau zu sein?

Ja Nein

Wenn ja, leiden Sie darunter?

Ja Nein

Besteht dieses Bedürfnis wiederkehrend seit mindestens 6 Monaten?

Ja Nein

Haben Sie wiederkehrend den Drang nichtsahnenden Personen, die nackt sind oder sexuelle Handlungen ausführen, heimlich zu beobachten?

Ja Nein

Haben Sie das bereits ein- oder mehrmals getan?

Ja Nein

Leiden Sie unter diesen Wünschen bzw. Vorstellungen?

Ja Nein

Besteht dieser Drang seit mindestens 6 Monaten?

Ja Nein

Empfinden Sie es als sexuell erregend jemanden zu sexuellen Handlungen (z. B. GV) zu zwingen?

Ja Nein

Haben Sie bereits (ein- oder mehrmals) jemanden zu einer sexuellen Handlung gezwungen?

Ja Nein

Leiden Sie selbst unter dieser Neigung?

Ja Nein

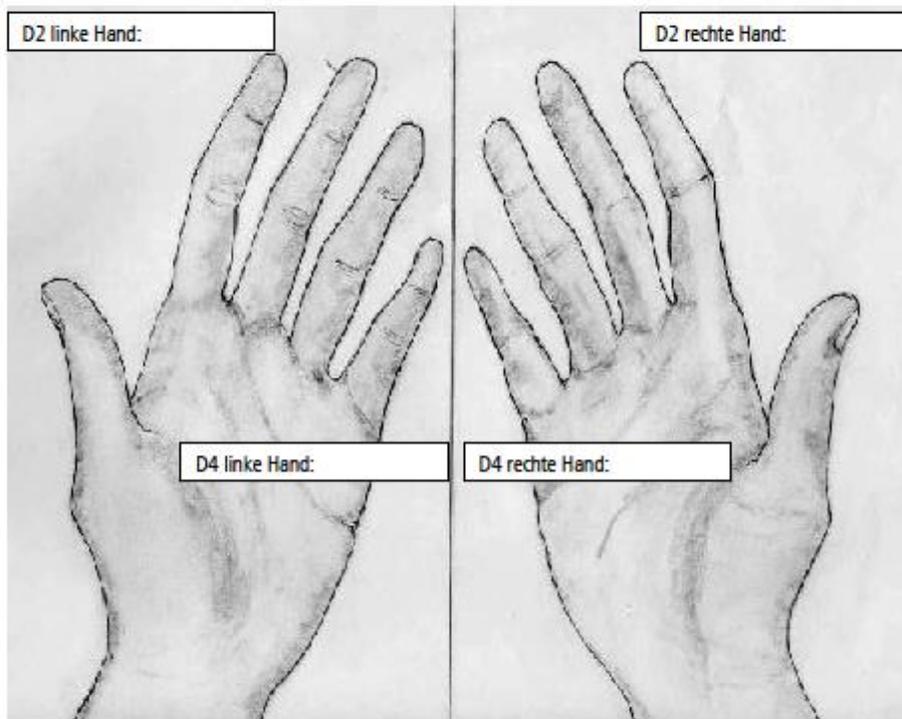
Besteht dieses Bedürfnis wiederkehrend seit mindestens 6 Monaten?

Ja Nein

NEMUP Erhebungsbogen

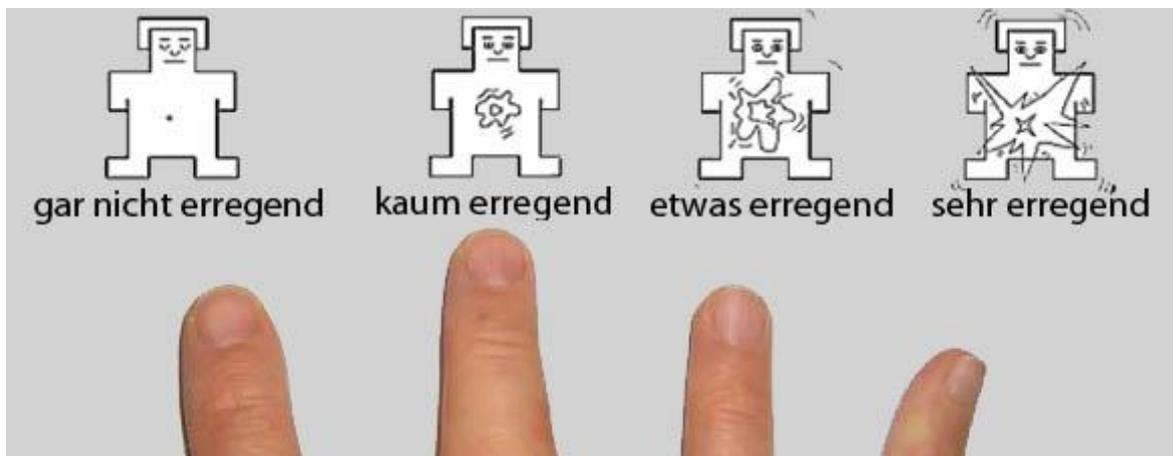
D2:D4 ratio

D2:D4 ratio LINKS
D2:D4 ratio RECHTS



7.4 Anhang 4: Instruktion VT

Sie werden nun nacheinander verschiedene Fotos sehen. Bitte schauen Sie sich jedes Foto an, und bewerten Sie mit den Tasten Ihrer rechten Hand, wie sexuell erregend Sie das Bild empfinden. Die Tasten sind genau wie im Beispiel unten angeordnet. Der Zeigefinger Ihrer rechten Hand steht für gar nicht erregend (auf der Tastatur Zahl 5), der Mittelfinger steht für kaum erregend (auf der Tastatur Zahl 6), der Ringfinger für etwas erregend (auf der Tastatur Zahl 7) und der kleine Finger für sehr erregend (auf der Tastatur Zahl 8). Prägen Sie sich die Skala bitte gut ein. Haben Sie noch Fragen? Wenn nicht drücken Sie einen der Knöpfe und es geht los.



7.5 Anhang 5: Instruktionen PPG – Probedurchgang

Sie sehen gleich ein Bild und werden dazu eine Kurzgeschichte hoeren. Waehrend der Praesentation wird einmalig fuer kurze Zeit ein grauer Punkt zu sehen sein. Druucken Sie dann bitte schnellstmoeglichst die linke Maustaste. Schauen Sie hierfuer am besten entspannt in die Mitte des Bildschirms.

Ausserdem druecken Sie bitte die linke Maustaste, wann immer Sie die Woerter ‘‘Wasser’’ oder ‘‘legen’’ hoeren - auch wenn diese konjugiert oder Teil eines laengeren Wortes sind, wie beispielsweise bei dem Wort ‘‘Wasserspiele’’. Achten Sie hierzu sorgfaeltig auf die Kurzgeschichten.

Druucken Sie bei den geforderten Reizen schnellstmoeglich und versuchen Sie die Maustaste moeglichst nur infolge dieser Reize zu druecken. Druucken Sie nun die Maustaste, sobald Sie sich die Bedingungen eingepraegt haben, auf die Sie reagieren sollen.

Das Experiment startet dann automatisch.

7.6 Anhang 6: Instruktionen PPG – Hauptdurchgang

Sie werden im Folgenden zunächst eine "neutrale" und dann mehrere Kurzgeschichten sexuellen Inhalts hoeren und dazu entsprechende Bilder sehen. Waehrend mancher Bilder wird fuer kurze Zeit ein grauer Punkt zu sehen sein. Druetzen Sie dann bitte schnellstmoeglich die linke Maustaste. Schauen Sie hierfuer waehrend des Experimentes am besten entspannt in die Mitte des Bildschirms.

Ausserdem druecken Sie bitte die linke Maustaste, wann immer Sie die Woerter "Penis" oder "kuessen" hoeren - auch wenn diese konjugiert oder Teil eines laengeren Wortes sind, wie beispielsweise bei den Woertern "kuesst" oder "Penisansatz". Achten Sie hierzu sorgfaeltig auf die Kurzgeschichten.

Druetzen Sie bei den geforderten Reizen schnellstmoeglich und versuchen Sie die Maustaste moeglichst nur infolge dieser Reize zu druecken. Druetzen Sie nun die Maustaste, sobald Sie sich die Bedingungen eingepraegt haben, auf die Sie reagieren sollen.

Das Experiment startet dann automatisch.

7.7 Anhang 7: Verteilung der Probanden in Relation der Altersindizes bei VT, PPG und fMRT

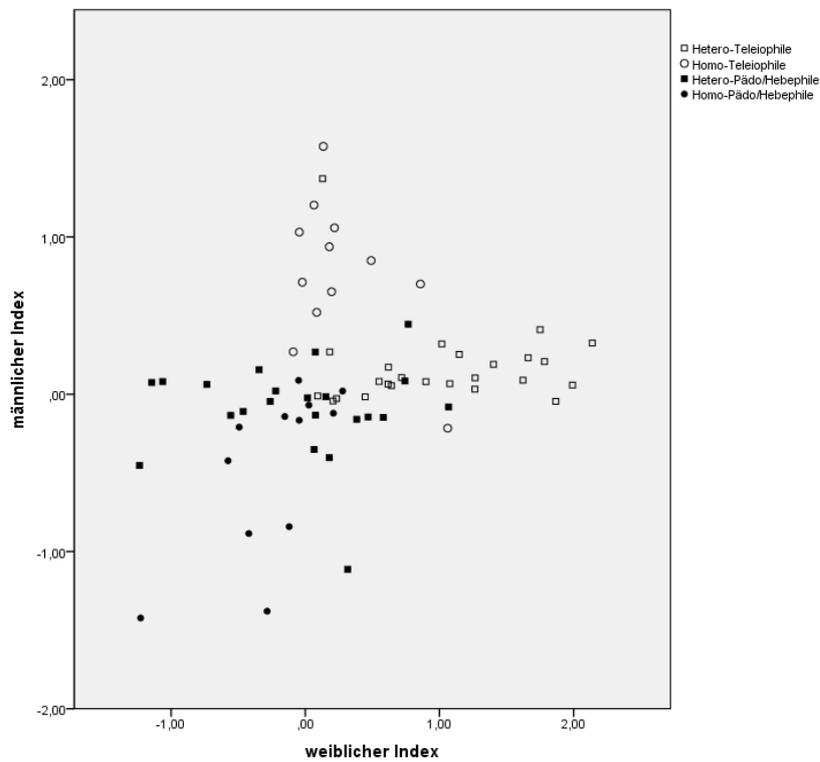


Abbildung 8: Verteilung der Probanden entlang der zwei Altersindizes des VT

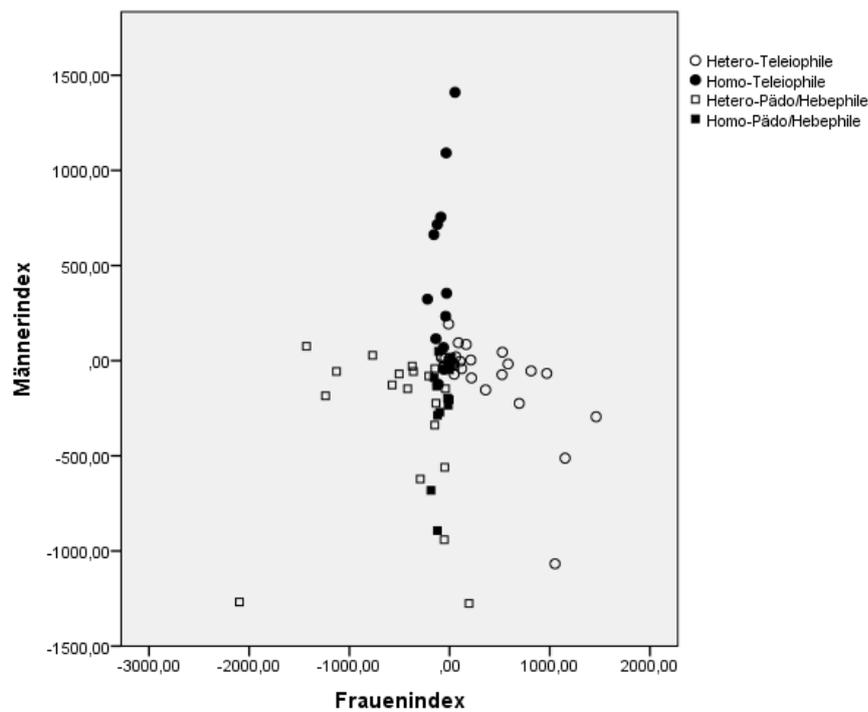


Abbildung 9: Verteilung der Probanden entlang der zwei Altersindizes der PPG (vor Ausschluss vor Low-Respondern)

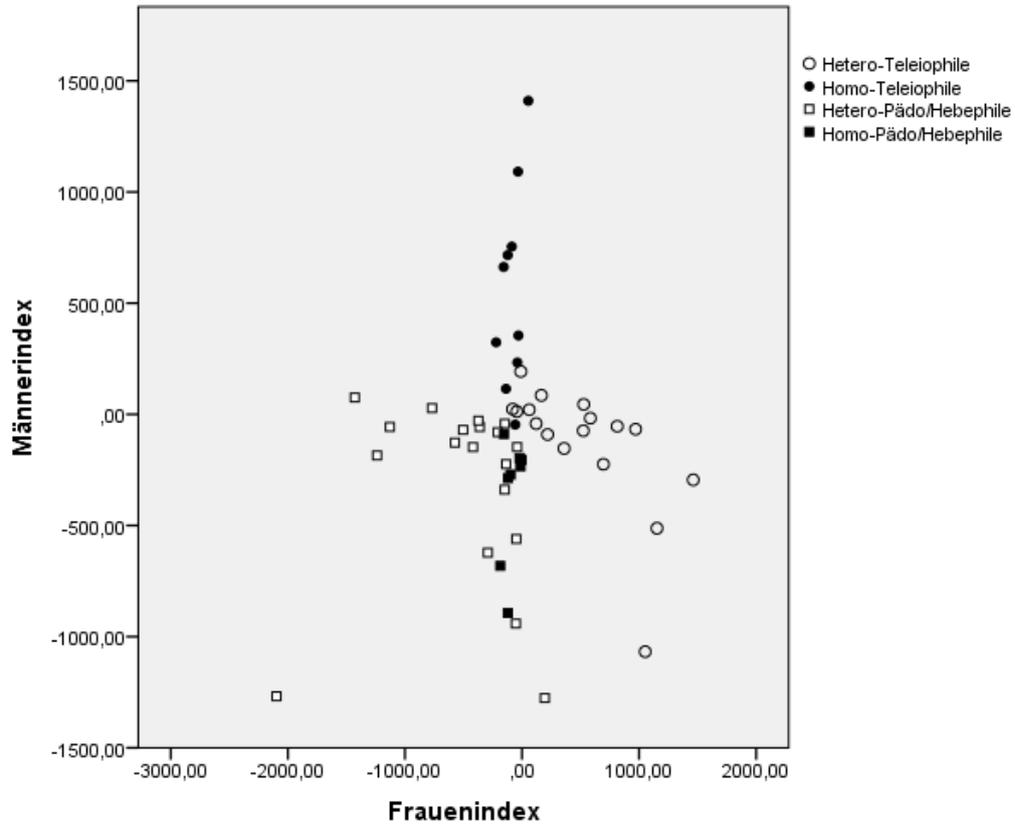


Abbildung 10: Verteilung der Probanden entlang der zwei Altersindizes der PPG (nach dem Ausschluss von Low-Respondern)

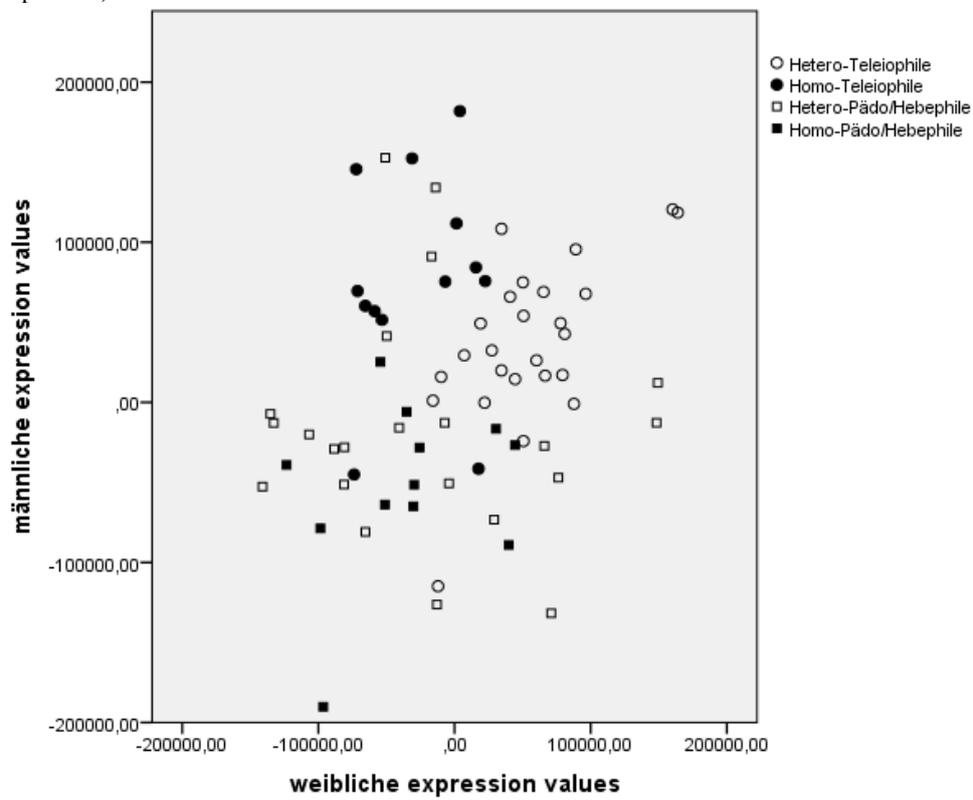
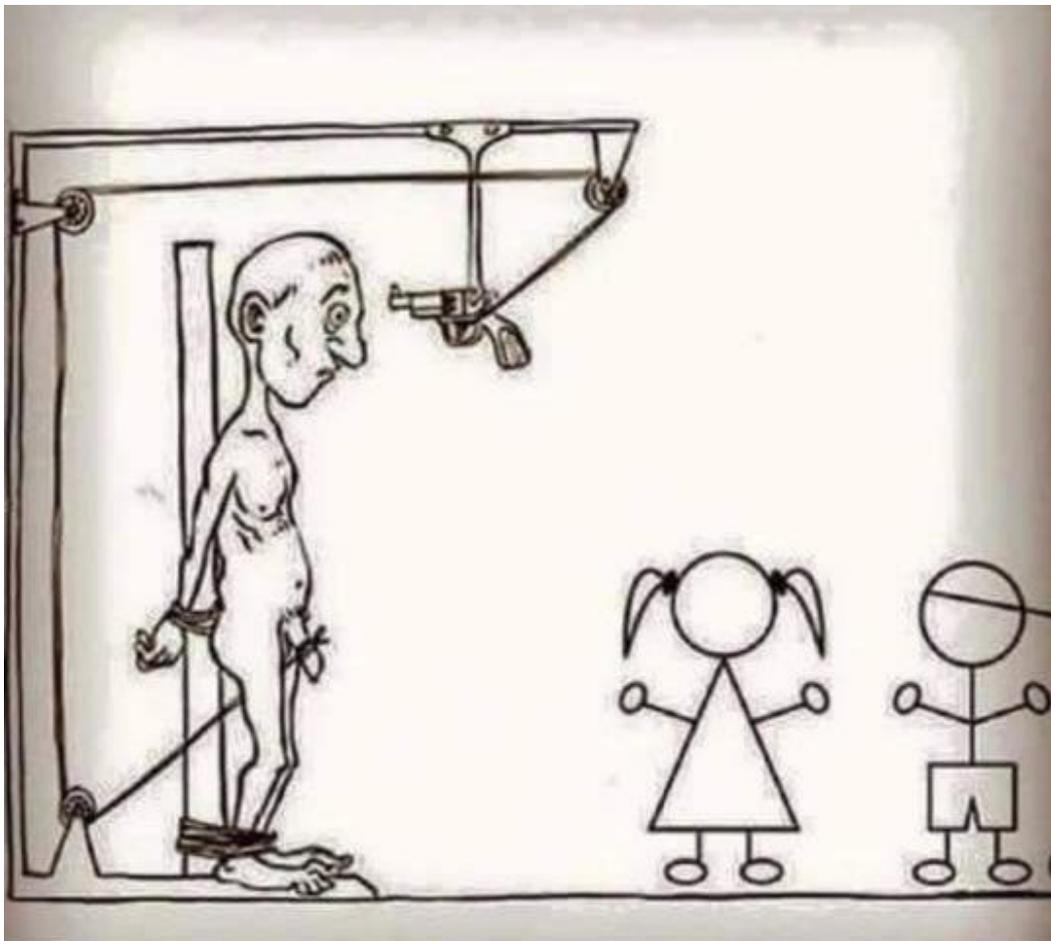


Abbildung 11: Verteilung der Probanden entlang der zwei Altersindizes der fMRT

7.9 Anhang 9: Karikatur zum Thema Pädophilie

Diese Karikatur macht die Vermischung der beiden Konstrukte der sexuellen Orientierung und des sexuellen Verhaltens deutlich und demonstriert gleichzeitig eindrucksvoll die enorme Stigmatisierung, der Pädophile ausgesetzt sind. Wie Moen (2015) beschreibt, sind sexuelle Handlungen mit Kindern als unmoralisch einzustufen, während Pädophilie als sexuelle Orientierung nicht anhand moralischer Maßstäbe zu beurteilen ist. Die Karikatur zeigt allerdings eine Ächtung pädophiler Orientierungen (genitale Reaktion auf kindliche Reize), ohne, dass Informationen zum sexuellen Verhalten vorliegen.



Quelle:

<https://www.facebook.com/bas1991/photos/a.637245986342467.1073741828.637245123009220/912464745487255/?type=1&theater> (abgerufen am 01.06.2015)

8 Lebenslauf

Persönliche Daten

Name Alexander Pohl
Geboren 05.12.1983 in Kiel

Ausbildung

2012 – 2016

Promotion im Rahmen des Forschungsprojektes „NeMUP“

2010 – 2016

Ausbildung zum psychologischen Psychotherapeuten am IFT-Nord

2004 – 2010

Studium der Psychologie an der CAU zu Kiel; Abschluss: Diplom (Note: 1,1)

1994 – 2004

Max-Planck Schule in Kiel; Abschluss: Allgemeine Hochschulreife (Note: 2,0)

Berufliche Laufbahn

Seit 01/2016

Mitarbeiter am Institut für Sexualmedizin und forensische Psychiatrie und Psychotherapie des ZiP in Kiel

07/2015 – 12/2015

Psychotherapeut in Ausbildung am ZiP in Kiel

07/2012 – 06/2015

Wissenschaftlicher Mitarbeiter in der Sektion für Sexualmedizin des UKSH Campus Kiel im Rahmen des Forschungsprojektes „NeMUP“

01/2012 – 06/2012

Psychotherapeut in Ausbildung in der Tagesklinik "Villa Karlstal" des ZiP Kiel

07/2010 – 12/2011

Wissenschaftlicher Mitarbeiter in der Sektion für Sexualmedizin des UKSH Campus Kiel im Rahmen des Drittmittelprojektes „Intramurale Therapie von Sexualstraftätern“